Rafael Malach

Perception without a Perceiver

In Conversation with Zoran Josipovic

Rafael Malach is currently a professor in the department of Neurobiology at the Weizmann Institute in Israel. His current research is aimed at understanding how the neuronal circuitry in the human brain translates a stream of sensory stimuli into meaningful perception. Rafael Malach received his PhD in physiological optics from UC Berkeley and did his post-doctorate research at MIT. Originally doing research on the organization of neuronal connections in the primate brain, his focus has recently shifted to the study of the human cerebral cortex using fMRI. Professor Malach has begun this research at Massachusetts General Hospital, exploring a new object-related region called the lateral occipital complex. Since then he expanded this research, studying the human visual cortex using a variety of methods, including adaptation paradigms, backward masking, and more recently naturalistic stimuli — all aimed at deciphering the intriguing link between perceptual experience and brain activity.

Introduction

Conventional models of human perception assume that sensory experience involves a critical interplay between activity in the sensory cortex, representing the stimulus, and the pre-frontal cortex, which serves as an 'observer' system — receiving and interpreting the patterns of activity originating in posterior, sensory cortex.

However, results from a recent fMRI study conducted by Rafael Malach and colleagues (Goldberg *et al.*, 2006), appear to challenge this accepted view.

Correspondence:
Zoran Josipovic *Zoran@cns.nyu.edu*
Rafael Malach *rafi.malach@weizmann.ac.il*

Journal of Consciousness Studies, **13**, No. 9, 2006, pp. 57–66

In a study in which individuals were exposed to a highly engaging popular movie, researchers found a surprisingly robust and widespread activation of most of the posterior part of the brain, which was remarkably 'synchronized' across individuals watching the same movie (Hasson *et al.*, 2004; Mukamel *et al*., 2005). These results attest to the massive engagement of sensory cortex by naturalistic sensory stimuli. However, in contrast to this widespread activation in sensory cortex, they found a remarkably little activation in frontal areas of the brain (Golland *et al.*, 2006).

To examine this issue, and also to start mapping in more detail the functional organization of the various networks of pre-frontal cortex, Malach and his colleagues conducted a series of experiments in which they mapped brain activity during tasks that were explicitly targeted at eliciting 'self-related' brain activity, such as introspection of visual and auditory modalities, or self-judgment evaluations. Although these were high-level cognitive tasks of extreme complexity, results showed a consistent pattern of activity focused primarily on the superior and medial pre-frontal cortex. However, a completely different and highly segregated network of more posterior areas was active during engaging perceptual tasks. No overlap was found between the two networks. Importantly, the prefrontal cortex was actually inhibited during intense perception. Thus, examining the activity of the pre-frontal system during rapid perceptual tasks revealed significant inhibition in these areas below the resting baseline. So not only were the prefrontal areas disengaged from perceptual awareness, they were actually inhibited during conscious perceptual awareness. Goldberg *et al.* interpret these results as arguing against the need for an active role of an 'observer' function during perception.

These findings have led Malach and his team to conclude that

> the role of self-related cortex is not in enabling perceptual awareness, but rather in allowing the individual to reflect upon sensory experiences, to judge their possible significance to the self, and, not less importantly for consciousness research, to allow the individual to report about the occurrence of his or her sensory experience to the outside world. [However] during intense perceptual engagement, all neuronal resources are focused on sensory cortex, and the distracting self-related cortex is inactive (Goldberg *et al*., 2006, p. 337).

The results are actually compatible with the strong intuitive sense we have of 'losing ourselves' in a highly engaging sensory-motor act. They are intriguingly reminiscent of recurrent eastern philosophical themes, found especially in Zen teachings, which emphasize the 'silencing' of the self during intense engagement with the outside

world. D.T. Suzuki, one of the leading Zen interpreters beautifully summarized it thus: 'Life is an art, and like perfect art it should be self-forgetting' (Suzuki, 1964).

Interview

Zoran Josipovic: How did you come to the idea to research whether perceptual consciousness is possible without self-representation?

Rafael Malach: The motivation for the research was an attempt to use fMRI to advance knowledge that may contribute to a neuronal theory of perceptual awareness. My broad perspective on how neuroscientists view theories of consciousness is that there is actually a surprising overlap in the type of processes assumed to be necessary for the emergence of conscious perception. Many of the models share a common theme requiring some kind of integration, such as re-entrant and/or reciprocal activation among groups of neurons (Lamme and Roelfsema, 2000; Tononi, 2005). Of course we are far from fully elucidating the details of such models but they certainly converge in many aspects. What seems more contentious is the issue of what constitutes a sufficient substrate for the emergence of conscious perception. A common theme that is popular in sensory perception research concerns models in which sensory percepts are analysed and represented in the posterior sensory areas and then transferred to higher order frontal cortex. Thus, sensory representations on their own are not regarded as sufficient for awareness. Rather, it is implicitly or explicitly assumed by many models that a higher order 'observer' or evaluation network is necessary, likely in pre-frontal cortex (Baars *et al*., 2003). The criticism that such requirement entails endless regress, i.e., each homunculus should then have its own smaller homunculus, can be reconciled by the possibility that such homunculus function will itself be sub-conscious and only its integration with sensory cortex leads to perceptual awareness.

It occurred to me that brain imaging using fMRI could be an excellent tool to examine such sub-conscious process since fMRI reveals both conscious and non-conscious brain activity. We therefore set out to study to what extent moments of vivid perceptual awareness — when there is no doubt that subjects were consciously aware of the sensory stimuli — were indeed associated with activation in pre-frontal areas. In our work we have focused on the most likely structures that may be involved with such 'observer' functions. These

are the pre-frontal brain areas that are active when subjects are involved in self-related aspects and introspection.

ZJ: Your claim that current theories posit a homunculus could be seen as overreaching, since after Ryle's argument against it, few if any neuroscientists seriously think there is such a construct.

RM: Even though we rule out the concept of a little man sitting in the head, it is still a very powerful concept whose essence is that it is not enough to have a sensory representation, but that you need to have something else in addition to it, some brain structure that is observing, putting in context and interpreting the information from the sensory cortex. It may be some unconscious sub-process that is dialoguing with sensory representations in order to have a conscious perception. This interpreter or 'observer' is another high-level cortical area that is interpreting information. The implicit assumption here is that it is not enough to have a state representation in the sensory cortex but something else is needed. Our entire study was aimed at examining this issue with respect to pre-frontal self representations, and the clear-cut result — which, I emphasize, is not yet conclusive for the entire prefrontal cortex, as we need to explore in detail more areas in the PFC and their potential relationship to awareness and perception — is that we can have a very vivid and engaging sensory perception without the involvement of some high level self-related interpreter in the prefrontal cortex. Furthermore, the more engaging is this sensory perception, the more vivid and absorbing it is, the less activity we get in these pre-frontal areas. So the model we are proposing is a sort of push–pull relationship between pre-frontal, self-related and high level cortical areas, and the sensory representations. When sensory areas are very intensively activated these areas turn off the prefrontal self-related representations through a mechanism we don't yet understand.

Why is our study not conclusive? Because we did not study every area of the high-level processing in PFC. We only focused on the areas that are the most likely candidates for such an 'observer' function, and these are the self-related representations, the areas which are activated when one has to make self-referential or introspective judgments. My use of the term 'self' here is strictly as a construct, a mental abstraction or a representation of the organism generated by the organism's brain. When you ask yourself a questions like: how do I feel about something, or what do I feel at this very moment? The areas in superior-medial pre-frontal cortex become activate, as found in many studies (for review see Northoff *et al*., 2006) including our own.

What we have found is that these prefrontal areas are not necessary for sensory consciousness, and in fact these areas shut off during vivid perceptions. So any attempt to construct a neuronal model of conscious perception should take this observation into account. Our results are clear-cut, at least within the limited 'window' provided by the fMRI methodology, that there can be sensory perception without the activation of self-representations.

Allowing ourselves to go far beyond the given data and derive a speculative hypothesis from it: these results appear to suggest that consciousness might emerge even from anatomically local neuronal interactions. Our long-term project is thus to explore how localized the activity in the cortex can be, and still lead to conscious perception. If I understand S. Zeki's work correctly (Zeki, 2003), it appears to indicate that conscious perception can be a local process. Why do I find this important? Because if perceptual awareness is a localized process, then it is more experimentally tractable. We may be able to identify a minimal structural network and dynamic in the cortex that is sufficient to produce conscious experience, and so you can explore in detail: what is happening in this minimal network activity that is leading to perceptual awareness.

ZJ: What is your speculation about what might be happening there?

RM: Basically if I have to speculate about the direction where our results appear to lead, my bias would be that perceptual awareness likely relates to a very, very tight integration of information among neurons, in which a neuron is not anymore an isolated entity but is an integral part of a network to such an extent that you can almost entirely explain the activity of that neuron by the activity of the whole of the network. This would explain why consciousness is tied to cortex, since cortex is optimally built to allow such fused networks, and why it takes time for conscious percepts to emerge, since it takes time for such networks to 'fuse' together. Although this is a very speculative notion, it does fits well with different experimental findings. Indeed, if you examine various neuronal models of consciousness which have been proposed based on such findings, for example, that proposed by Tononi (Tononi, 2004; 2005) regarding integration, or Lamme (Lamme, 2004; Lamme and Roelfsema, 2000) regarding recurrent processing between cortical areas, the critical element and maybe the only element common to these models is that they require some mechanism eliciting tight integration of information in groups of neurons.

Our working hypothesis is an extreme version of these models proposing an extremely 'local' model of awareness. We hypothesize that perceptual awareness can emerge from any intense integration of activity within a cortical network of neurons, even an extremely small and localized one. This is an exciting possibility to my mind because it implies that conscious awareness may be much more ubiquitous and experimentally tractable than we tend to assume. However, it is important to emphasize that these notions are highly speculative at this stage and a lot more experimental data will be needed before we could attach any validity to them. The important point is that at the very least these notions can lead to testable experiments even with our present day methodology.

ZJ: For many researchers, binocular rivalry has been a starting point for assertions that local processes are not sufficient to create consciousness, and that a large-scale global synchrony among functionally different regions is needed for consciousness. Your theory of consciousness as a local property contradicts the more accepted views that consciousness is globally distributed in the brain, such as Baars' Global Workspace Theory (Baars, 1997; 2005) or Christof Koch's model (Crick and Koch, 2003). In particular, these and other models point to the necessary activation in dorso-lateral PFC and the areas in the parietal cortex, such as the parietal egocentric maps, which are part of the 'body self' system.

RM: Our experiments were aimed at examining some aspects of these models and the results argue against an active role of self-related pre-frontal cortex in perceptual awareness, at least during intense engagement with the outside world — as I pointed above, these regions actually appear to shut-off during vivid perception. We have not examined DLPFC in the way we have medial PFC, but if I have to risk a guess, I would say that DLPFC is also not necessary for conscious perception. This speculation is based on available data from patients who underwent damage to DLPFC. Such patients do not seem to lose consciousness, and they do not lose contents of perceptual consciousness either, even though their memory may be affected by the injury. However, I would like to emphasize again, that this is totally speculative at this stage. Deriving such conclusions from brain damaged patients is very risky because the brain is plastic, so it could be the case that after brain lesions in DLPFC other brain areas may be taking over and compensating for the lost function.

One point I want to make very clear: we are not studying here the issue of the enabling processes of consciousness. We are unable with our type of research, which is based on rapid contrast of cognitive states, to search for the brain structures that allow you to be generally awake or asleep or in coma. For example, a patient with a brain stem stroke can have a completely functioning cortex but be in a coma-like consciousness state. Our fMRI methodology is currently not optimized for revealing such 'global switches' of conscious states.

ZJ: In terms of the parallels with Zen, which you hint at in your paper, it can be argued that the 'pure perception' that occurs in Zen and some other meditations (Izutsu, 1982), where the self-related cognitions and affects are temporarily inactive, is a different state from something like watching a movie. During pure perception, the sense of one's self is suspended and one is fully present in 'just seeing' or 'just hearing', without customary conceptualizations that involve value judgments. However, watching a movie is a different type of experience. While we watch movies a number of different processes are occurring, such as the changing affective states that are elicited, our reflexive judgments, etc., all of which are usually subconscious unless we decide to pay attention to them.

RM: Part of the power of brain imaging and the whole revolution that was initiated in brain research as a result of it, lies in the fact that it allows you to see the unconscious processes as well as conscious processes in action, because any cognitive and affective process whether conscious or unconscious will cause neurons to be active, and therefore can be measured with fMRI (provided it is within the spatio-temporal 'window' of the method). We can even see some completely unconscious processes, which nevertheless elicit clear activations. In our study we saw a process that, we believe, had intriguing parallels with themes common to Zen meditation. This can be thought of as an axis whose origin is an ordinary experience with frequent interference of self-related thoughts; a midpoint would be very engaging movies where one is engrossed in the experience and not thinking of oneself. It could be that Zen meditation experience (of pure perception), where there seems to be no self-reference at all, forms the end point of this axis. Maybe if we could scan the Zen meditators who could get to that intense experiential state, we would see even deeper shutting off of the pre-frontal self-representational networks. As a neuroscientist I see here an intriguing possibility. I would also like to mention, at this point, the earlier research regarding the self and its antagonistic

relationships in terms of brain activity which has been conducted by several groups, particularly M. Raichle and colleagues (Raichle *et al*., 2001; Fox *et al*., 2005).

ZJ: Do you feel that your results match with their findings?

MR: Our findings regarding inhibition of self-related representations are indeed very compatible with Raichle's and Gusnard's findings. In both studies we find inhibitory interactions between the 'perceptual' network and a more internally oriented system that has been termed the 'default' brain network by Raichle and colleagues and the 'intrinsic' network by us (Golland *et al*., 2006). My interpretation of these converging results is that they argue against models that pose co-activation of self-related and sensory-related areas as a pre-condition for perceptual awareness. The general theme that emerges from our results is that instead of thinking about the brain as a global network, we now see evidence for a fundamental, bi-partite, subdivision of the brain into two major networks, an extrinsic one dealing with the outside world and an 'intrinsic' network, whose detailed function is still enigmatic but which appears to be dealing mainly with aspects of the self and internally oriented processes.

A critical point from the viewpoint of consciousness research is that intense sensory and perceptual awareness definitely activate elements of the extrinsic network and do not activate the intrinsic system. Of course there are situations in life when both systems are co-active, for example when you are engaged in introspection about some external object or event. In such cases, one may hypothesize that two states are simultaneously active: you are both engaged internally — being aware of yourself — and you are also engaged with the external sensory stimulus. For example when I am introspecting about the fact that I am now tasting coffee, our hypothesis will suggest that there are simultaneously active representations of my self and of the taste of coffee in the cortex. Our data indicate that in such cases both intrinsic and extrinsic systems are indeed activated.

Trying to think about it more philosophically, this may be precisely the point where the wonder and puzzlement about consciousness starts. Could it be then, that the riddle of consciousness and the intense engagement the mind–body problem poses to us have emerged only with the evolution of humanity, perhaps in part because we human beings have the capacity for a balanced co-activation of the intrinsic, self-related network and the extrinsic, sensory, networks? The intrinsic network may allow us to experience ourselves in the act of

experiencing, and to communicate our sensory experiences to each other, thus giving rise to the deep puzzle of consciousness and the brain.

ZJ: This is very interesting because it brings into focus the difference between meditations as found in the Buddhist traditions such as Zen, and meditation in the Hindu traditions such as Vedanta and Yoga: in Zen the attempt is to focus on the outside (other then one's self) and eradicate the self (Austin, 1998; Loori, 2002), while the Hindu traditions use precisely the opposite approach, shutting out the outside and focusing on the Self (Aranya, 1983; Gambhirananda, 1990).

RM: This is a neat idea because we can think metaphorically about the extrinsic and intrinsic systems as reflected, one in Zen meditation and the other in Hindu meditation!

ZJ: My particular interest is in meditations that balance the internal and the external experience, and thus, possibly balance the activations of the two networks, such as those found in Tibetan Buddhist traditions of Dzogchen and Mahamudra (Chagme, 1998; Namgyal, 2001). To extend your analogy of the axis to this case: there may be a gradation of introspecting with the different degrees of balance between these two systems (and possibly, differential involvement of the parts of these systems), starting with everyday experience of introspecting, and ending in nondual awareness which is a complete balancing of inner and outer experience into one field of experience that is not divided into subject and object.

RM: An analogy in daily life would be the introspection state in which you are aware of both internal and external percepts simultaneously: your intrinsic, self-related processes on the one hand, and the external world on the other, i.e. you perceive yourself in the act of perceiving the outside world. This is where I hypothesize, and we have some initial results, that the two systems might be balanced. So if this framework is not completely wrong, one would expect this type of meditation to have some of this flavour, it should have the intrinsic and extrinsic systems co-active and balanced.

References

Aranya, S.H. (1983), *Yoga Philosophy of Patanjali* (Albany, NY: SUNY Press).
Austin, J. (1998), *Zen and the Brain* (Cambridge, MA: MIT Press).
Baars, B.J. (1997), 'In the theatre of consciousness', *Journal of Consciousness Studies,* **4** (4), pp. 292–309.

Baars, B.J. (2005), 'Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience', *Progress in Brain Research*, **150** pp. 45–53.
Baars, B.J., Ramsoy, T. and Lauries, S. (2003), 'Brain, conscious experience and the observing self', *Trends Neurosci.* **26**, pp. 671–5.
Chagme, K. (1998), *A Spacious Path to Freedom* (Ithaca, NY: Snow Lion).
Crick, F. and Koch, C. (2003), 'A framework for consciousness', *Nature Neuroscience*, **6** (2), pp. 119–27.
Fox, M.D., Snyder, A.Z., Vincent, J.L., Van Essen, D.C., Corbetta, M. & Raichle, M.E. (2005), 'Human brain is intrinsically organized into dynamic, anti-correlated functional networks', *PNAS,* **102** (27), pp. 9673–8.
Gambhirananda, S. (1990), *Mandukya Upanishad with the Commentary of Shankaracharya* (Calcutta, India: Advaita Ashrama).
Golland, Y., Bentin, S., Gelbard, H., Benjamini, Y., Heller, R., Nir, Y., Hasson, U. and Malach, R. (2006), 'Extrinsic and Intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation', *Cerebral Cortex* (electronic publication; Advance Access published on May 12, 2006).
Goldberg, I., Harel, M. and Malach, R. (2006), 'When the brain loses its self: Prefrontal inactivation during sensorimotor processing', *Neuron*, **50**, pp. 329–39.
Hasson, U., Nir, Y., Levy, I., Fuhrman, G. and Malach, R. (2004), 'Intersubject synchronization of cortical activity during natural vision', *Science,* **303**, pp. 1634–40.
Izutsu, T. (1982), *Toward a Philosophy of Zen Buddhism* (Boulder, CO: Prajna Press).
Lamme, V.A.F. (2004), 'Separate neural definitions of visual consciousness and visual attention: a case for phenomenal awareness', *Neural Networks,* **17**, pp. 861–72.
Lamme, V.A.F. and Roelfsema, P.R. (2000), 'The distinct modes of vision offered by the feedforward and recurrent processing', *Trends Neurosci.,* **23**, pp. 571–9.
Loori, J.D. (2002), *The Art of Just Sitting* (Somerville, MA: Wisdom Publications).
Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I. and Malach, R. (2005), 'Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex', *Science*, **309**, pp. 951–4.
Namgyal, D.T. (2001), *Clarifying the Natural State.* Kathmandu, Nepal: Ranjung Yeshe Publications.
Northoff, G., Heinzel, A., de Greck, M., Berpohl, F., Dobrowolny H. & Panksepp, J. (2006), 'Self-referential processing in our brain: A meta-analysis of imaging studies on the self', *NeuroImage,* **31** (1), pp. 440–57.
Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A. & Shulman, G.L. (2001), 'A default mode of brain functioning', *PNAS,* **98** (2), pp. 676–82.
Tononi, G. (2004), 'An information integration theory of consciousness', *BMC Neuroscience,* **5** (1), pp. 5–42.
Tononi, G. (2005), 'Consciousness, information integration and the brain', *Progress in Brain Research,* **150**, pp. 109–26.
Suzuki, D.T. (1964), *An Introduction to Zen Buddhism* (London: Grove Press).
Zeki, S. (2003), 'The disunity of consciousness', *Trends in Cognitive Sciences*, **7** (5), pp. 214–18.