NEUROSCIENCE of Consciousness

Special Issue: Consciousness science and its theories

Local neuronal relational structures underlying the contents of human conscious experience

Rafael Malach[*,†]

Department of Brain Sciences, Weizmann Institute of Science, 200 Herzl St. POB 76100, Rehovot, Israel;
The School of Psychological Sciences, Tel Aviv University, P.O. Box 39040, Tel Aviv 6997801, Israel

[†]Rafael Malach, http://orcid.org/0000-0002-2869-680X

[*]Correspondence address. Department of Brain Sciences, Weizmann Institute of Science, 200 Herzl Street, Rehovot 76100, Israel. Tel: +972528750225;
E-mail: rafi.malach@gmail.com

Abstract

While most theories of consciousness posit some kind of dependence on global network activities, I consider here an alternative, localist perspective—in which localized cortical regions each underlie the emergence of a unique category of conscious experience. Under this perspective, the large-scale activation often found in the cortex is a consequence of the complexity of typical conscious experiences rather than an obligatory condition for the emergence of conscious awareness—which can flexibly shift, depending on the richness of its contents, from local to more global activation patterns. This perspective fits a massive body of human imaging, recordings, lesions and stimulation data but opens a fundamental problem: how can the information, defining each content, be derived locally in each cortical region. Here, I will discuss a solution echoing pioneering structuralist ideas in which the content of a conscious experience is defined by its relationship to all other contents within an experiential category. In neuronal terms, this relationship structure between contents is embodied by the local geometry of similarity distances between cortical activation patterns generated during each conscious experience, likely mediated via networks of local neuronal connections. Thus, in order for any conscious experience to appear in an individual's mind, two central conditions must be met. First, a specific configural pattern ("bar-code") of neuronal activity must appear within a local relational geometry, i.e. a cortical area. Second, the individual neurons underlying the activated pattern must be bound into a unified functional ensemble through a burst of recurrent neuronal firing: local "ignitions".

Keywords: human cortex; consciousness; structuralism; perception; visual system; cortical areas; face perception; fMRI; iEEG

Overall aim of the review

The issue of the link between brain and consciousness is extremely broad, encompassing a massive body of research. Every subsection of this review could form the basis for one or a number of full reviews on its own. Hence, this review will need to be selective and it is important to clarify what it attempts to achieve. My aim here is not to provide a comprehensive coverage of the literature but to present the local-structuralist perspective in the broadest strokes. I feel that covering a full set of details will defeat this purpose and will mask the broad picture. In a similar vein and in the service of focus, the review will be restricted to human consciousness and will use face perception as the main model system to illustrate the central concepts. However, I propose that principles derived from this model cases can be generalized to a large extent to all other types of conscious content.

A localist perspective on human conscious experience

Surveying the landscape of various theories attempting to link neuronal states and conscious experience, it becomes clear that a dominant theme in most of them is the assumption that any conscious experience necessitates activation of global or at least large-scale networks. This theme is of course most prominent, as its name implies, in the Global Workspace Theory, positing the activation of a global fronto-parietal network as a precondition for conscious experience (Mashour *et al.* 2020) and in a similar vein high-order thought theories, invoking high-order frontal

regions in addition to first-order ones (Lau and Rosenthal 2011). A careful reading of other prominent theories also highlights their reliance on widespread or large-scale connectivity. For example, in Integrated Information Theory, the high level of integrated information underlying human consciousness is assumed to be subserved by widespread causal interactions across the entire cerebral cortex or at a minimum its posterior "hot zone" (Tononi *et al.* 2016). Similarly, a dependence on global network inter-actions is assumed at the level of modality-specific hierarchies-thus—recurrent theories posit dependence of conscious experi-ence on top-down activations across the hierarchy of cortical areas (Lamme and Roelfsema 2000).

Here, I would like to consider an alternative, localist per-spective. In this framework, different categories of conscious content—from face perception to thinking and speech—are linked to localized activity in specific, cortical areas (see also, Zeki 2008). Under the localist perspective, the observed large-scale nature of brain activity associated with conscious experience is the consequence of the fact that conscious experience is often a multi-dimensional and complex occurrence, involving many cate-gories of conscious content. Such complex contents, which engage a diversity of categories, will naturally be linked to multiplic-ity of cortical areas and large-scale network activation patterns. For example, watching a movie involves, at a minimum, visual, auditory and emotional contents, and these will be linked—even under a purely localist framework—to a large-scale co-occurring activity in many cortical areas. However, critically, and in contrast to global theories of consciousness, the awareness of, say, the visual scene in the movie will be associated with local activity in category-selective visual cortical areas and will not be dependent on, e.g., non-visual areas specialized for deci-sion or meta-cognitive processes typically residing in the frontal lobe.

To put it differently, large-scale activations, including fronto-parietal networks, may flexibly emerge as needed but are not "obligatory" for all specific categories of conscious content (Malach 2007, 2011, 2012). Furthermore, these auxiliary regions may even be suppressed during certain conscious states, e.g. being absorbed in an engaging sensory experience (Goldberg *et al.* 2006). The metaphor of an orchestra may be helpful to intuitively clarify the localist perspective. In this metaphor, musical instru-ments stand for local cortical regions, while the music is the overall conscious experience. One can easily envision the orches-tra flexibly switching from a solo performance involving one or few instruments to a large-scale finale, in which the entire orches-tra joins in. The point about the localist perspective is that it does not assume an obligatory minimum number, or a specific set, of instruments that must play in order for the music to be produced.

What is the experimental evidence supporting the localist per-spective? The body of relevant literature relevant to this issue is massive and far beyond the scope of this review. Here, I will very briefly touch on some key lines of evidence. When review-ing this literature, it is important to emphasize the pitfalls and challenges in interpreting the data. An important complication is that the cortex is an extremely connected network, where signals travel fast and wide in a reflexive manner (Bassett and Sporns 2017). Thus, merely detecting any network activation during a conscious experience, or even decoding from such activation category of the experience, cannot be taken as a proof that the network plays a critical role in the experience. For example, it is quite likely, because of a habitual tendency to name visual

objects, that the category of a visual object could be decoded from activity in frontal, language-related regions, but clearly, as many lines of evidence demonstrate, this frontal activation is underlying linguistic rather than visual information in this case.

Three criteria for identifying candidates for conscious content

Functional selectivity

Experimentally, in order to consider a cortical region as a candi-date for a specific category of conscious content, I propose that at least three conditions should converge. As I will detail below, each of these criteria puts some constraints on large-scale theories of conscious experience.

The first criterion is that an area should show functional spe-cialization specifically for the conscious category in question. This means that an area should be preferentially active for the category of conscious content for which this area specializes, e.g. face per-ception, and not others. This is important since the brain is highly interconnected, and as we have shown previously in a large-scale intracranial study, there is an extremely rapid, global, low-level activation spread generated by any new content (a phenomenon we termed "glow"—Noy *et al.* 2015). So a mere widespread activa-tion cannot be interpreted as underlying the diversity of subjective experiences. By contrast, a growing body of Blood Oxygen Level Dependent, Functional Magnetic Resonance Imaging (BOLD-fMRI) and Intra-Cranial Electroencephalography (iEEG) research reveals a mosaic of local functional specializations manifested across the entire human cortical mantle.

A striking illustration of this issue concerns the debate over the role of the frontal lobes. Although frontal regions have been consistently shown to be significantly activated by visual stimuli, Noy *et al.* (2015) demonstrated that these regions are preferentially activated when participants reported about these visual stim-uli compared to when they merely passively viewed them (see bottom panels of Fig. 1). Similar findings were documented dur-ing so-called "report free" paradigms (Pitts *et al.* 2014; Tsuchiya *et al.* 2015). Judging by their preferential activation profile (func-tional specialization) frontal regions should be associated with conscious contents auxiliary to vision perception such as intro-spection, report, decision-making and meta-cognition, rather than sensory experience proper. Accepting this criterion thus puts serious constraints on the global-workspace and high-order thought theories—which propose that fronto-parietal activation are obligatory for every conscious content—to the extent that these theories propose that fronto-parietal activation is obliga-tory for every conscious content, rather than merely the access aspects of conscious experience (Block 2005), or place the high-order functionality elsewhere, e.g. subset of neurons throughout the brain.

The complementary aspect of this specialization criterion is that regions that show high activation despite the absence of a conscious content are unlikely to underlie this content. An illustration of such case can be found in the phenomenon of spontaneous blinks. We are typically unaware of the optical inter-ruptions associated with such blinks (which happen, unnoticed, every 5 seconds or so) perceiving instead an illusory, stable visual image. Despite such perceptual stability, early visual areas show robust activations by the optical interruptions during sponta-neous blinks—in fact, of very similar magnitude to externally induced, perceptually salient, optical simulations of these inter-ruptions (see top panels in Fig. 1). Thus, neuronal activation
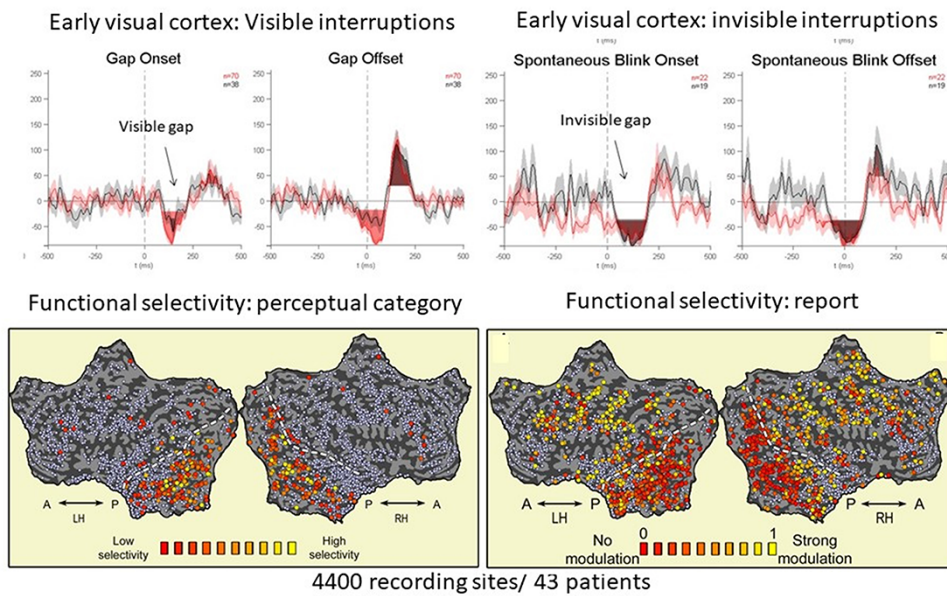
Figure 1. Functional selectivity as a criterion for a central role in conscious content. Top panels: early visual cortex does not differentiate between invisible spontaneous blinks and visible interruptions. Example iEEG recording from early visual cortex showing robust response to visible, externally introduced interruptions in the visual image (left) and a similar magnitude responses to similar, but invisible interruptions occurring during spontaneous blinks (Golan *et al.* 2016). Bottom panels: functional selectivity differences between fronto-parietal and occipital cortex. Large-scale iEEG recordings from wide cortical expanses during 1-back visual paradigm. Left panel: cortical sites showing selectivity to visual categories. Right panel: cortical sites showing selectivity to report. Yellow points indicate highly selective sites. Note the striking differences in functional selectivity across the cortex—with report-related selectivity in fronto-parietal regions and perceptual category selectivity in occipito-temporal regions.

levels in early visual areas do not appear to reflect the difference between consciously perceived vs. identical, but not-perceived, optical stimuli- providing evidence against a central role of early visual cortex in conscious vision. Interestingly, findings from the domain of sleep appear to lead to a similar conclusion. Thus, during sleep, and despite experiential disconnection from the external environment, early auditory cortex appears to be responsive to external auditory stimuli (Sela *et al.* 2020; Wilf *et al.* 2016). These results support the conclusion that early sensory areas, while serving pivotal function in conducting peripheral stimuli to high order areas, do not play, on their own, a major role in conscious perception. These findings counter theories that pose an obligatory role for recurrent activation to early sensory areas during conscious perception (Lamme and Roelfsema 2000).

Lesion-associated loss of function

The second criterion is that a local damage to a cortical area must result in the abolishment of all experiences of the conscious content subserved by this region. Classic examples of a loss of specific conscious content following a cortical lesion include the case of a neurological patient that lost form vision following specific damage to the visual object area LO (Malach *et al.* 1995; James *et al.* 2003) and the well-known loss of face perception (prosopagnosia) associated with damage to face-selective cortical regions (Barton 2008). In some sense, large parts of the neurological literature can be viewed as a detailed compilation of an atlas of the human cortex referencing circumscribed cortical lesions and their consequent loss of specific conscious content. As in the functional specialization criterion, the complementary conclusion is also true: if a cortical region is damaged and a specific conscious experience is not affected, it can be cautiously concluded that the lesioned area does not play an obligatory role in the above experience. A

striking example of such a case is the existence of vivid, realistic, visual hallucinations reported by late-onset blind individuals (Santhouse *et al.* 2000; Hahamy *et al.* 2021) and in particular hallucinating patients whose blindness was due to damage to early visual cortex (Martinelli *et al.* 2020). Such cases complement the functional specialization data cited above in providing evidence against a central role of early sensory areas in conscious perception. However, it remains an open question whether early sensory cortex activation is still necessary for the fine details of visual experience.

In a similar line of argument, numerous examinations of frontal neurological patients consistently reveal deficits in language abilities, sequence planning, context setting, decision and error detections, personality traits, etc. (Stuss and Benson 1984), but rarely such lesions result in the abolishment of sensory perception. Again, these results are compatible with the functional specializations discussed above in countering a central role for frontal regions in conscious perception as proposed by global workspace and high-order thought theories.

It should be noted that a different interpretation of such lesion studies has been proposed, contending that prefrontal lesions do impact perceptual awareness (Odegaard *et al.* 2017). However, importantly, the criterion for Pre-Frontal cortex (PFC) involvement under this perspective is far weaker; the authors actually acknowledge that PFC lesions only partially affect subjective perceptual awareness and have never been shown to completely abolish it. Given the broad inter-connections of PFC with the rest of the brain, such partial effects are actually compatible with the localist view. The criterion under review here aims to identify cortical areas that are "obligatory" for visual perception. The failure to completely abolish visual perception through PFC lesion is further explained by a hypothetically unique, more diffuse architecture. While indeed one cannot rule out such possibility, it

should be noted that non-visual functions, such as speech, are totally abolished by localized PFC lesions.

Conscious modulation by electrical stimulation

Finally, a third, and particularly powerful, criterion concerns conscious experience induced by electrical stimulation of specific cortical sites. This criterion is methodologically more challenging to interpret, due to the rather coarse nature of such stimuli. Furthermore, the possible spread of activation to other cortical areas is not fully established. Thus, in many cases, the precise physiological impact of electrical stimulations on the neuronal populations is difficult to assess. However, one can apply this criterion by stating that whenever stimulation of a cortical site results in a consistent and highly specialized positive or negative modulation of conscious content, it is likely to support a link between the stimulated cortical region and the specific conscious content. Examples of such links have been numerous, starting with the seminal work of Penfield (Penfield 1958; Fox et al. 2020). More recently, a particularly striking illustration has been the finding of perceptual distortions specific to faces following electrical stimulation of face-selective cortical sites (Parvizi et al. 2012; Keller et al. 2017).

While negative results should be treated with caution, it is interesting that a recent review of massive stimulation study failed to find evidence for specific perceptual modulations following frontal lobe stimulations (Raccah et al. 2021) again in agreement with converging evidence from functional specialization and cortical lesion data. A related, and potentially very informative field pointing to a similar cortical specialization is the observed link between specific conscious contents during epileptic aura and their anatomical cortical loci (Wilkinson 2003).

Convergence of functional selectivity, lesion and stimulation data supporting the localist perspective

In contrast to the incompatibility of the three criteria with global models of awareness, it is interesting to note the consistent and natural convergence of these criteria in support of a localist account of conscious awareness. Thus, one can view the large and rapidly growing literature of BOLD-fMRI mapping as repeatedly associating local cortical regions to specific contents of conscious experiences. These mapping links now span the entire gamut of human conscious experiences, to name just few examples: from perceptual categories (Grill-Spector and Malach 2004) to abstract thoughts (Berkovich-Ohana et al. 2020), to introspection (Goldberg et al. 2006) and memory recall (Harmelech et al. 2015). In a number of cases, such non-invasive BOLD localizations were successfully corroborated by more direct, high-resolution, single-unit (Mukamel et al. 2005; Khuvis et al. 2021), and iEEG recordings (Mukamel et al. 2005; Privman et al. 2007) in demonstrating, for example, rapid activation during memory recall and the default mode network (Foster et al. 2015). Similarly, the neurological and direct stimulation literature includes numerous, striking, cases of localized modulations of conscious categories that often complement the recordings data.

Thus, one can conclude that based on a careful consideration of converging experimental data, the evidence supports a localist perspective in which locally specialized cortical regions each underlie a different category out of the diverse variety of conscious contents. It should be emphasized that although much of the cited literature centers on sensory perception, the converging criteria apply to non-sensory aspects as well. Loss of speech associated with localized frontal lesions is a striking example, as well as the surprisingly localized nature of introspective and thought processes revealed in the frontal lobes by fMRI (Goldberg et al. 2006; Berkovich-Ohana et al. 2020).

However, it is important to clarify that this is not an argument against a pivotal role for large-scale networks and information integration across long-range cortical expanses. Surely, these play a crucial role in many large-scale functions—such as introspection, meta-cognition, report, memory, feed-forward flow of information, top-down attentional selections and global modulation effects such as during arousal, novelty detection global changes in awareness states and many more. The point is that these large-scale activations emerge when conscious experience is complex and depends on concurrent categories, rather than playing an obligatory role for any conscious experience per se. Thus, an extensive body of research in recent years has highlighted the prevalence of large-scale ongoing spontaneous (also termed resting state) activity throughout the cortex (Nir et al. 2008, Harmelech and Malach 2013). While it appears that such activity goes on below the consciousness threshold (Moutard et al. 2015; Ramot et al. 2016), it may serve important functions influencing the local activations by providing context or even underlying spontaneous and creative behaviors (Broday-Dvir and Malach 2021).

Another important aspect of large-scale dynamics is in setting the overall state of consciousness, such as the switch from being aware to a dreamless sleep. Under the localist perspective, such large-scale state changes can be considered as content-independent, global enabling factors that provide the necessary conditions for the content-specific local events, such as local ignitions (see below), to take place.

A non-selective role for PFC activation in conscious experience

It should be emphasized that the three criteria outlined above address the issue of neuronal correlates of specific conscious content. However, it has been argued that the PFC may play a more general, non-specific role. Thus, it was suggested that while conscious contents are maintained in specialized areas, the mechanism determining whether those contents are conscious or not is localized in PFC, which acts as a kind of pointer, enabler or index of conscious changes. A detailed elaboration of this argument has been recently introduced by Michel and Morales (2020) who review a number of studies in which PFC activation was higher for "conscious" vs. "non-conscious" perceptual images. For example, in masking paradigms, it is consistently found that when participants perceive the target stimulus, PFC shows a higher activation compared to when they fail to perceive it—even in the so-called no report paradigms—when participants are not required to report what they perceive.

However, it is important to highlight the potential confusion inherent in such conscious vs. non-conscious terminology. A more proper description of the perceptual dynamics during contrast experiments, such as backward masking, should be "consciously perceiving the target" vs. "consciously perceiving the non-target." Clearly, in both target seen and target unseen conditions, the participants are consciously perceiving a visual stimulus. The difference between these two states is not in the existence vs. non-existence of a conscious experience but rather in a difference in conscious content, which in the "conscious" state includes the target stimulus, while in the "non-conscious" state includes only the non-target stimuli. However, critically, apart from the content of

conscious experience, there is no lapse whatsoever in the overall state of conscious visual perception proper.

To put it differently, if we accept the hypothesis that PFC activity is associated with overall changes in conscious perception, then the observed differential activation in PFC associated with perceiving vs. not perceiving the target during a backward masking experiment should be linked, under this hypothesis, to lapses in perceptual awareness during masking (or rivalry) experiments when failing to see the target. However such lapses have never been reported. Thus, one should not interpret the modulations in PFC activity during, e.g. backward masking experiments, or binocular rivalry, as reflecting a switch from conscious perception to a non-conscious state, but rather these PFC activation changes are likely to be associated with changes of specific experiential states, for example, the feeling of change, enhanced novelty and levels of arousal,. Such non-perceptual effects are perfectly in line with a localist account, which, it is important to emphasize, definitely accepts that frontal regions underlie certain categories of conscious experience but attributes visual perception to specialized, high-order visual areas, located posteriorly, in Occipito-temporal cortex (for further discussion, see Noy *et al.* 2015). Finally, one may argue that it is the variety of conscious experiences that preferentially activates the PFC in the target conscious condition during back-ward masking, - since in the "non-conscious" state one perceives only the mask while in the "conscious" state both target and mask are perceived sequentially. A similar argument can be proposed for the fronto-parietal activation found during the perceptual switches of binocular rivalry. However, direct comparison of high vs low rates of image changes contradicts this possibility (see e.g. Figure 4A in Goldberg et al. 2006).

A central conundrum inherent in the localist perspective

However, a central problem that is particularly fundamental in the case of the localist perspective concerns the mechanism that endows different cortical regions with their unique experiential categories. Simply stated, the question is as follows: given the rather uniform nature of the cortical architecture, why do different local parts of this architecture subserve such strikingly different conscious contents? Why, for example, activation of a group of neurons in the lateral occipital complex gives rise to visual object perception (Malach *et al.* 1995) while activation of a similar number of neurons in area MT (middle temporal) merely a short anatomical distance away will elicit the perception of visual motion (Tootell *et al.* 1995)? Where is the information that underlies this difference derived from? A coherent localist perspective must search for this information within the local bounds of the specific cortical regions themselves. Unlike global theories, which may derive their information from large-scale network architecture a strictly localist theory, it cannot ground these differences neither in the unique inputs entering different regions (which will be incompatible with the observed elicitation of conscious contents in these regions in the absence of such inputs, e.g. during electrical stimulation). Similarly, from a localist perspective, such information cannot be attributed to a higher-order "readout" area (which will be incompatible with neurological lesion data as discussed above).

Taking a broad historical perspective, it is intriguing to note that a very similar conundrum faced the pioneers of the structuralist school when confronting this problem within the domain of linguistic meaning. What confers specific semantic meaning to a word? Their ingenious and elegant solution was to propose that the information defining linguistic meaning is embedded in the relational structure of the word to other words in the body of language (Saussure 1959). So, for example, the word "mountain" acquires its semantic meaning not by virtue of its (arbitrary) letter or phoneme combinations but by its close semantic association to the words "hill" and "peak" and enhanced distance from words such as "valley" and "gorge." Below, I will discuss how such relational structures may be implemented in neuronal terms and underlie the content of conscious experience. I will consider examples of experimental results that support it. An informative discussion of some philosophical implications stemming from the structuralist perspective can be found in Rosenthal (2010).

Neuronal implementations of relational structures

How can a relational distance between various contents of awareness be implemented in cortical neuronal activity? The neuronal implementation of distances becomes readily apparent when we examine the nature of neuronal responses associated with the experience of a specific conscious content. Here, I will illustrate this notion using conscious face perception since this is a well-studied aspect of human vision. However, it is important to emphasize that this is merely an experimentally convenient model system whose operational principles should be generalizable to the entire range of conscious categories mediated by human cortical function.

Distributed and unique neuronal responses to faces

Examining the neuronal responses in human high-order visual cortex during the experience of seeing a face (e.g. when presented with a picture of a face), one is struck by three major characteristics of the response. First, there are neurons that are functionally specialized to the category of faces (e.g. Khuvis *et al.* 2021). Second, these neurons are clustered together and can be found with high probability in the occipito-temporal region called the fusiform face area (Kanwisher *et al.* 1997). Finally, and critically, face-selective neurons possess an extremely distributed tuning within the face category (see Fig. 2, top panel). Thus, each human face-selective neuron shows a significant response to essentially every face exemplar presented to it. However, critically, each face-selective neuron has a unique "profile" of activation to different face exemplars. This means that when looking at the pattern of responses of the entire ensemble of human face-selective neurons, perception of each face exemplar is associated with a unique configuration or a pattern of multi-neuron activity. Metaphorically, one can envision this link as an association of each face image with a unique "bar-code" made up of the unique activity pattern of the entire ensemble of face-selective neurons (see Fig. 2, bottom panel; Grossman *et al.* 2019). With these observations in hand, the neuronal implementation of similarity relationships between conscious contents within a category becomes straightforward: it is simply the level of configural similarity between the different patterns or "Bar codes" made up of the activity of the content-selective neurons in a particular cortical region. The structuralist perspective then states that what defines the content or the "feel" of a specific conscious experience will be how similar or dissimilar is the "bar-code" associated with this unique content from all the "Bar codes" potentially elicited by all other contents within the conscious category.
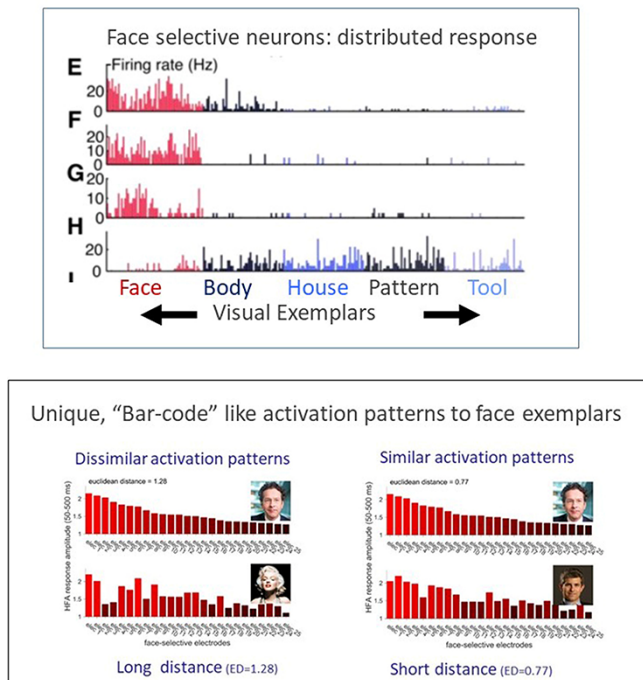
Figure 2. Distributed and unique properties of human face-selective neurons. Top panel: example recordings from three face-selective neurons (E–H). Each vertical line represents the neuron's response to a single exemplar of the relevant category. Note the high face-selectivity, on the one hand, and the broad tuning to face exemplars on the other. Critically different face exemplars elicit different magnitudes of firing in each neuron (see Khuvis *et al.* 2021) for more details. Bottom panel: multi-site patterns of iEEG activations in face-selective cortex to different face images. Note that each face exemplar elicits a unique configural pattern, forming a "bar-code"-like signature. Pairs of faces show a diversity of similarity distances (see Grossman *et al.* 2019 for more details).

Practically, it is easier to conceive how such relational patterns can be formed by depicting the distance relationships between patterns in the format of a multi-dimensional vector space, in which each dimension is formed by the activity of a single neuron. In such a format, each activation pattern occupies one point in this vector space and the similarity distances become the normalized geometric distances between the points (see Fig. 3, top panel). Hereafter, I will use the term "relational geometry" when referring to the set all similarity distances encoded by an ensemble of neurons in a specific cortical region.

Different conscious categories are associated with different relational geometries

A comprehensive structuralist perspective of conscious experience must provide a neuronal account for the different categories of conscious content. For example, when viewing a picture of an object, there are at least two perceptual categories that can be distinguished with regard to the conscious experience associated with this object. One category is the pictorial nature of the percept—how the object looks like, what shape and color it has, etc. The other is the conceptual category—is the object a cup, a cat, a car, etc. Critically, the structuralist perspective necessitates different relational geometries to express these different categories. Again, face perception is a good model to illustrate why this
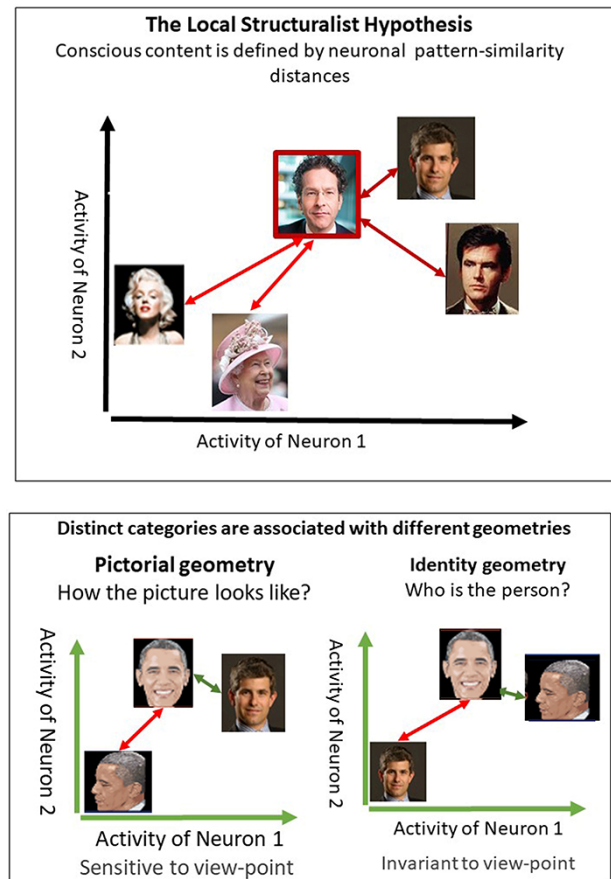

Figure 3. Schematic depiction of the relational geometry concept. Top panel: the neuronal underpinning of a face percept. A schematic illustration of neuronal pattern similarity distances between different face images. For this graphical illustration, the patterns are assumed to consist of only two face-selective neurons forming the horizontal and vertical axes (in reality the vector space is built up of likely millions of neurons Levy *et al.* 2004). The structuralist hypothesis proposes that the appearance of each face, e.g. the framed face in the center, is defined by its location in this vector space relative to all other faces. Bottom panels: different geometries underlying different conscious categories. The left panel depicts a hypothetical relational geometry underlying the pictorial appearance of faces, while the right panel depicts the different relational geometry underlying the personal identity of these faces. Note that different viewpoints of the same face will be distant from each other in the pictorial relational geometry but overlapping in the case of the identity geometry.

is so (see Fig. 3, bottom panel; see also Schwartz and Yovel 2016). Consider looking at a picture of a famous face, say Barak Obama, shown in a front view and then looking at a picture of Obama's profile. A relational geometry that implements the pictorial category of Obama's face will necessarily represent the neuronal activation patterns elicited by these different viewpoints of Obama as distant from each other—since Obama's front and side views appear, pictorially, as very different from each other. However, a relational geometry that represents the category of personal identities of the faces rather than their visual appearance will represent these different viewpoints as very close to each other since the two, very different viewpoints of Barak Obama underlie the same identity recognition experience. From a neuronal point of view, we would expect very different pattern relationships implemented by these two structures.

Intrinsic Lateral Connections
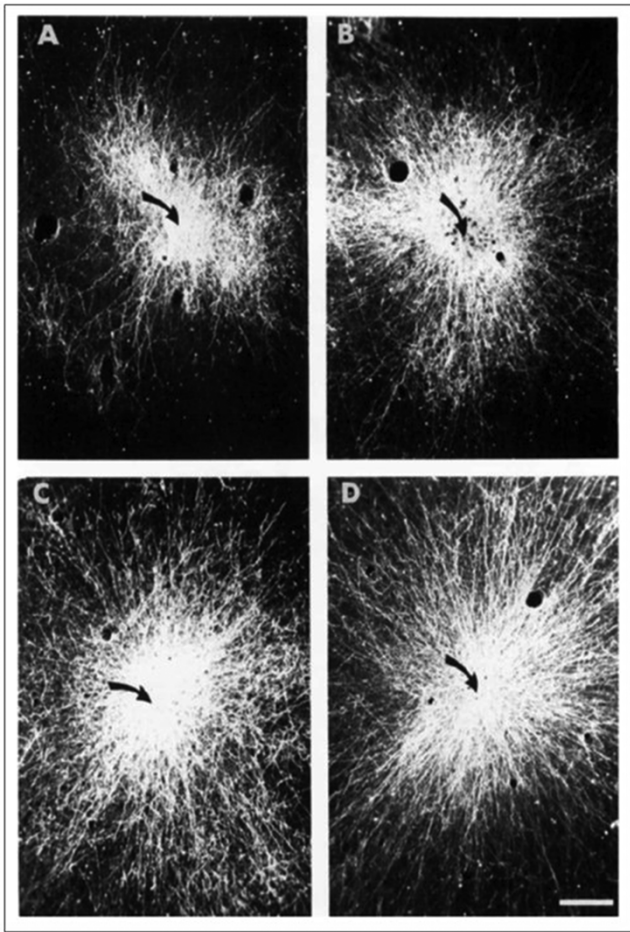### A massive substrate for implementing relational geometries

Figure 4. Lateral cortical connections provide a massive substrate for local relational geometries. Four examples (Panels A-D) of traced lateral connections from early (top) and high-order (bottom) visual areas. White lines depict axonal fibers, mostly excitatory that emanate from the central (marked red) cluster of neurons (scale bar: 0.1 mm). Note that this dense halo of connections exists in every cortical point and allows massive, recurrent, local connectivity between neurons, which could be a powerful substrate for the formation of relational geometries. However, the geometry of the lateral connections differs significantly across areas, in agreement with their hypothetical role in underlying different local relational geometries (for more details, see Amir *et al.* 1993).

While it may be possible to multiplex different relational geometries within the same cortical region, what we find experimentally is that often such geometries are anatomically segregated, i.e. located in different brain regions. For example, in the case of the appearance vs. recognition categories, both fMRI and iEEG studies suggest that the former can be localized to high-order visual areas (Grill-Spector *et al.* 1999; Grossman *et al.* 2019) while the latter are likely represented in medial-temporal lobe structures (Reber *et al.* 2019). Thus, it appears that, from a local structuralist perspective, the most straightforward implementation of different categories of conscious content is by assigning them to different cortical regions—whose relational geometries manifest the rules of similarity characteristic of each specific category.

Intrinsic cortico-cortical connections as a plausible substrate implementing the relational geometries

When considering the very rich and diverse group of conscious categories and their proposed link to relational geometries, an obvious question is how such a wide range of different geometries can be implemented in the cortex. Conceptually, this question is equivalent to asking how is the information that codes the different distance relationships between contents stored locally in each cortical region. Concerning how information is stored in the cortex, there is of course an enormous body of data and models, going back to the seminal work of Hebb *et al.* (1994) concerning how information is stored in cortical circuits (e.g. Harmelech and Malach 2013). For local storage of such information, a plausible anatomical candidate could be the intrinsic lateral cortico-cortical connections that are found within each cortical region (see Figure 4) . This massive and dense network of local cortical connectivity is often overlooked but is certainly one of the most prominent and distinguishing features of the cortical architecture (e.g. Malach *et al.* 1993). Given that cortical information is stored in the unique sets of synaptic connectivity, the local lateral connections offer a straightforward and massive substrate for such information storage.

The proposed build-up of local category-specific geometries through lateral connections offers an attractive explanation for one of the most defining characteristics of cortical function, namely, the clustering of neurons sharing similar functional properties in close proximity forming so-called cortical "columns" (Hubel and Wiesel 1968; Malach 1994). It should be noted that this clustering is what allows neuronal functional selectivities to be revealed through mass recording methods that rely on neuronal group activations—from BOLD-fMRI to intracranial iEEG recordings. Thus, the observation of the high functional selectivity of such neuronal groups attests to their clustered functional similarity. It is an attractive possibility that this cortically ubiquitous functional clustering is not accidental but serves a role in optimizing the formation of local relational geometries across the cortex.

Finally, seminal studies are compatible with a role for lateral connections in conscious integration (e.g. Luck *et al.* 1989). However, as will be shown below, relational geometries can be generated also through the convergence of feed-forward inputs, and clearly, cortical regions differ markedly in the functionality of their inputs (as well as top-down influences)—so the impact of these external sources on the nature of the local geometries should be considered as well. Future studies will be needed to map the detailed synaptic connectivity patterns of the different relational geometries found in different cortical regions to resolve this issue.

Experimental support for the structuralist perspective

There is a rapidly growing body of literature revealing relational geometries across many domains and even neuronal coding of human brain function. These studies use various diagnostic tools to uncover such geometries, most commonly, by plotting the pair-wise distances of activity patterns (termed representational similarity, or dissimilarity, analysis). Such analysis has been applied in numerous human studies and fields, ranging from

visual categories (Edelman *et al.* 1998; Kriegeskorte and Kievit 2013) to memory (Chen *et al.* 2017), including a recent discovery of a hippocampal relational structure embodying the transformation from autobiographic to semantic memory items (Norman *et al.* 2021) and many others. However, a central question remains: are these relational structures of any functional significance? For example, could these geometries merely emerge accidentally or be due to some developmental constraints and actually do not serve any informative role in the moment-by-moment emergence of conscious contents?

Here, I will illustrate how this question can be addressed experimentally by presenting examples of two experimental studies, derived from our own work on face-selective cortical regions. As I will show, these studies provide experimental support for the notion that relational geometries play an important role in conscious perception.

Convergent evolution of human and artificial relational geometries

The first study addressed the issue of relational geometries indirectly, by taking advantage of the recent unprecedented success of deep convolutional networks Deep Convolutional Neural Networks (DCNNs). These artificial networks, inspired by the hierarchy of the visual system, are feed-forward networks that through massive training (involving millions of labeled exemplars) achieve human-level performance in face and object recognition. Considering these artificial networks (Grossman *et al.* 2019) reasoned that they can serve as an informative testing ground for the functional significance of cortical relational geometries. The rationale here can be illustrated by the metaphor of convergent evolution. Imagine an alien that knows nothing about flying or aerodynamics, examining a range of flying creatures of different kinds—both biological and artificial. Despite the alien's complete ignorance, it will be immediately struck by the appearance of a common structure that emerges in many of these flying creatures: their wings. What is particularly convincing about the wings' role in aviation is that they appear in creatures that are built and evolved in completely different manners—say airplanes and flies. Precisely this structural convergence, despite massive divergence in make-up, is what justifies the conclusion that the wings are essential for flying. This "convergent evolution" toward wings can be taken then as a strong indicator that wings play a critical role in flying even without any understanding of flight or aerodynamics.

In a similar vein, Grossman *et al.* (2019) conjectured that if relational geometries are critical for face perception, we should find them across very different networks that are capable of performing face recognition tasks, even though these networks may be fundamentally different in nature. On the contrary, precisely this difference is what will make their convergence a strong case for the functional importance of relational geometries.

To test this conjecture, Grossman *et al.* (2019) compared the relational geometries of neuronal activation patterns in face-selective cortical areas to the geometries of the different layers of an artificial, face-trained DCNN. In this experiment, multi-site patterns of activation "Bar codes" were recorded from intra-cranial contacts placed in face-selective visual areas during the presentation of different face images, and the similarity distances between these patterns induced by pairs of faces were calculated. This results in a representational distance matrix that captures the neural relational geometry of human face-selective cortex. To examine whether this geometry has any relation to DCNN geometries, a similar distance matrix was built, separately, for each layer

of the DCNN, by feeding the artificial network with the same faces viewed by the patients and generating a pair-wise representational distance matrix for each layer of the DCNN. Finally, and crucially, the DCNN and human neural geometries were compared. The results revealed a highly significant correlation between the human and artificial networks (see Fig. 5). Such a remarkable convergence strongly suggests that the relational geometry of the human face-selective cortex is not an accidental, developmental epi-phenomenon but rather is a structure that plays a crucial functional role in face perception.
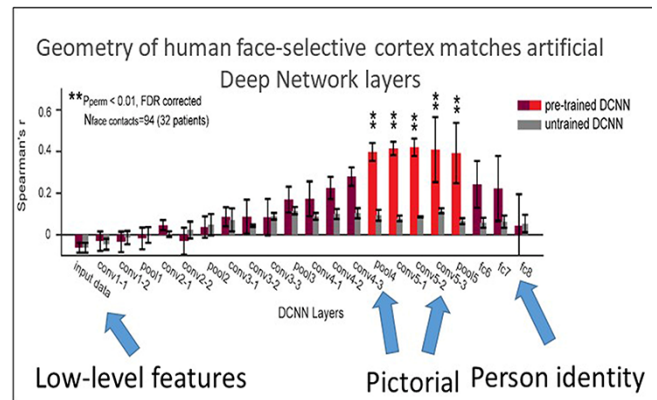

Figure 5. Evidence for "convergent evolution" between human face-selective relational geometries and deep convolutional networks. The correlation between the human and artificial geometries was measured in each layer of a DCNN exhibiting human-level face-recognition performance. This was achieved by generating matrices of pair-wise similarity distances of iEEG-recorded multi-site activity patterns in face-selective cortex and comparing them to the distances of activation patterns in each layer of the DCNN that was fed the same set of face images. Note the highly significant correlation between the human cortex and DCNN geometries. Importantly, the correlation was confined to middle-high DCNN layers—indicating a pictorial rather than identity-related geometry (for more details, see Grossman *et al.* 2019).

As we have argued above, a prediction of the local structuralist hypothesis is that different categories of conscious experience—e.g. the pictorial vs. the person identity categories—should be reflected in separate relational geometries. Interestingly, Grossman *et al.* showed that the human relational geometry of face-selective cortex was found in the middle layers of the DCNN—layers that are related to the pictorial, rather than person—identity categories of the faces (which are localized in the top DCNN layer). To further examine this conclusion, Grossman *et al.* fed the DCNN with the same faces but presented at different viewpoints, a transformation affecting the pictorial but not the identity categories. This resulted in the abolishment of the brain-like DCNN geometry, further demonstrating the pictorial rather than identity nature of human face-selective cortex. Indeed, a converging line of evidence appears to indicate that the identity category is represented downstream in the more conceptual representations of medial temporal lobe structures (Quiroga *et al.* 2005; Gelbard-Sagiv *et al.* 2008; Reber *et al.* 2019). The relational geometry found in the DCNN—a purely feed-forward network—illustrates a possible mechanism of constructing a human-like geometry in a feed-forward rather than intrinsic manner. However, it is equally likely that the human geometry is constructed via a different connectional mechanism. As we noted above, local lateral connections offer a powerful and rich substrate for such constructs. It will be interesting to explore whether future

artificial networks, implementing intrinsic lateral connections, may generate human-like geometries using such connections as well.

Similar neuronal and perceptual geometries in high-order visual cortex

The most straightforward way to test the structuralist hypothesis is to examine to what extent the neuronal relational geometries can explain the psychophysics of conscious contents. This was tested in the case of face-selective cortex by Davidesco *et al.* (2014). In the study, the neuronal distances among pairs of faces were measured from face-selective cortical sites, in a manner similar to that by Grossman *et al.* (2019). In parallel, participants were asked to judge the perceptual similarity between the same pairs of faces by simply placing their printed pictures physically closer or further away from each other. Critically, comparing the neuronal and perceptual distances for each face pair revealed a highly significant correlation (see Fig. 6). Essentially, this result directly demonstrates a correspondence between the relational geometry of the psychophysics of conscious perception and the neuronal relational geometry of face-selective cortex, directly supporting the local structuralist hypothesis. Importantly, the brain-perception correspondence was specific to the geometry of high-order visual cortex. In contrast, the relational geometry in early visual cortex failed to show such correspondence to face similarity judgments, further attesting to the unique, local, nature of the cortical relational geometries.

To emphasize, an essential prediction of the hypothesis is that the neuronal similarity distances should correspond precisely to the perceptual (or more generally cognitive) similarity distances experienced by the individual. This prediction can be extended to non-conscious discriminations as well; here, we would expect that the neuronal distances appearing in non-conscious processes will predict the non-conscious discrimination performance as is evident in, e.g., blindsight. Furthermore, it is predicted that different activation patterns within a specific cortical region (geometry) will always be associated with a subjectively discriminable content of conscious experience.
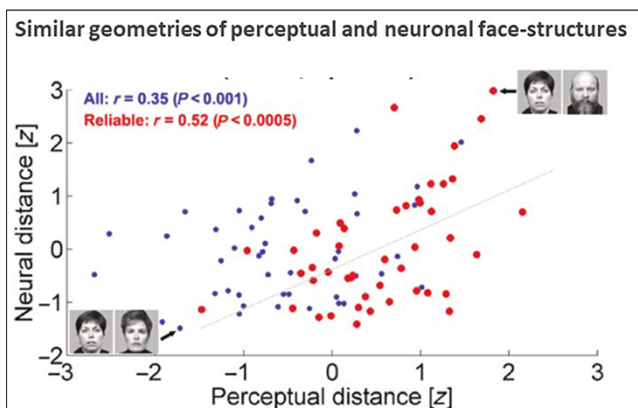
What is the role of local neuronal "ignitions" in the structuralist framework?

One of the defining features of the transition from non-conscious to conscious perception are non-linear bursts of activity (termed "local ignitions") of content-selective neurons that last several hundreds of milliseconds, regardless of stimulus duration (see Fig. 7; Fisch *et al.* 2009; Podvalny *et al.* 2017). Since the structuralist framework proposed here is based on configurations of activation patterns rather than the magnitude of activation, this raises the question of the role that these ignitions may play in conscious experience.

One attractive possibility is that the high bursts of activity are necessary to rapidly bind the individual neurons into a functional and coherent assembly. From the perspective of relational geometries, this will be a mechanism for rapidly constructing an informative relational geometry within which the different contents can be implemented. It is important to note that looking at cortical anatomy, it is difficult to demarcate distinct connectional segregations between neurons located in different cortical regions. On the contrary, the cortex is a massively interconnected network that, with the exception of few instances (e.g. area V1), does not seem to be isolated along sharp anatomical boundaries. This raises the attractive possibility that the relational geometries are assembled flexibly, through recurrent neuronal activations. Thus, the high magnitude bursts of activity, which we and others have speculated may be driven by recurrent cortical connections (Moutard *et al.* 2015), may reflect the operation of a binding mechanism that transforms a set of independently active neurons into an integrated pattern-generating assembly. This recurrent activity forms a relational geometry above and beyond the low-level, ongoing, spontaneous activity that goes on even in the absence of any conscious content (Moutard *et al.* 2015).
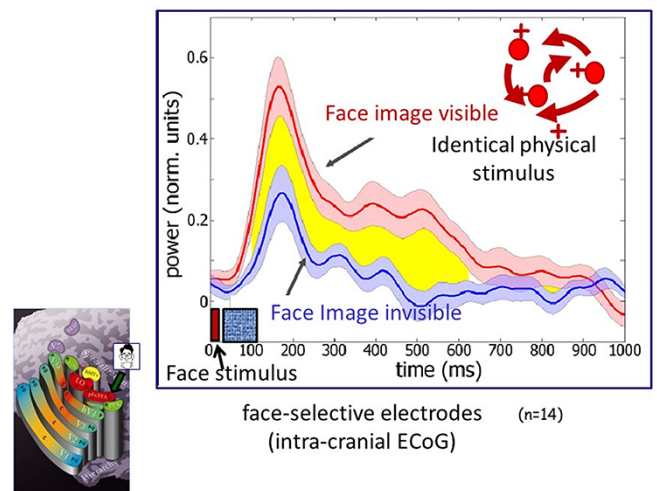

Figure 6. Correspondence of neuronal and perceptual geometries. Pair-wise distances of iEEG recorded patterns of activation, derived from face-selective sites in high-order visual cortex in response to face image pairs, were compared to psychophysical similarity judgments obtained for the same pairs. Note the significant correlation between the neuronally and perceptually derived distances, indicating similar relational geometries. Importantly, the relational geometry of early visual cortex was different and was not matched to the perceptual face similarity judgments (Davidesco *et al.* 2014).

Figure 7. Bursts of high neuronal activity ("ignitions") associated with conscious face perception in high-order face-selective cortex. iEEG recordings from high-order face-selective cortex (left inset) during backward masking of face stimuli. Contours depict the activation dynamics when patients reported seeing (red) vs. not seeing (blue) the target face stimuli. Note the non-linear activation for the seen target images that outlasted the brief stimuli (red vertical bar, followed by the mask-grey square) by hundreds of milliseconds (see Fisch *et al.* 2009 for details). An attractive role of ignitions within the structuralist perspective may be to recurrently bind independent neurons into a functional relational geometry.

Thus, to become experientially explicit, each relational geometry may require its constituent neurons to activate each other in an ignition-like dynamics in order for the geometry to become functional. The time it takes for such a geometry to be established will naturally depend on the rate of information transfer—i.e. firing rate—and this could explain why the ignitions typically manifest as high neuronal firing rates. However, it is important to note that these ignition bursts must produce meaningful patterns that can fit within a relational geometry in order to underlie conscious content. For example, a global, undifferentiated, burst of activation associated with epileptic discharge is not expected to lead to any conscious content.

Another possible function of ignitions may be in broadcasting the conscious content forward to high-order decision-related regions—a function that has been extensively discussed in global work-space theories (Baars 2005; Mashour *et al.* 2020) or to memory-related regions (Quiroga *et al.* 2008). Alternatively, such ignitions may be fed backwards for controlling input signals as suggested by recurrent theories (Lamme and Roelfsema 2000; Hochstein and Ahissar 2002) or for top-down attentional modulations (e.g. Davidesco *et al.* 2013). Finally, the ignitions may serve to induce plasticity changes in the neuronal connectivity. All these various functions are not necessarily mutually exclusive and may all depend on the high-activity local ignitions of neuronal assemblies forming the relevant structural geometries.

A structuralist resolution of the adaptation puzzle?

While high and sustained neuronal ignitions characterize the transition from non-conscious to conscious content, their termination poses a major challenge to localist models. The problem stems from the fact that, as mentioned above, the ignited bursts typically last 200 ms and then rapidly decline by at least 2-fold even when the visual stimuli, and the accompanying perceptual experience, remain unchanged (Gerber *et al.* 2017; Podvalny *et al.* 2017). This rapid-adaptation effect constitutes one of the most ubiquitous and striking dissociations between neuronal firing rates and perceptual content. However, in line with the discussion above, the observed dissociation between activation magnitude and conscious perception can actually be considered as an indirect, supporting evidence for the structuralist perspective. Note that under the structuralist hypothesis, conscious content depends on the configural shape of the activation patterns and its relationship to the other activation patterns, rather than their overall magnitude of activation. Thus, according to the structuralist hypothesis, perceptual content should be invariant to the magnitude (i.e. firing rate) of the response, which is fully compatible with the observed perceptual stability of stationary visual stimuli despite the rapid adaptation. Support for this structuralist prediction has been obtained recently in our lab by examining the dynamics of neuronal activity during relatively long (1.5 s) image presentation times. Our study reveals that the activation magnitude of high-order visual neurons selective to the presented images strikingly declines (80% reduction) during the presentation time. In contrast, the configural activation pattern and its embedding distances within the relational geometry remain remarkably stable across the entire image presentation duration. Thus, compatible with the structuralist hypothesis,

perception of visual images is maintained by virtue of the stability of the relevant relational geometry despite the highly transient overall activation magnitude (Broday-Dvir and Malach, in preparation).

On the other hand, the perceptual stability challenges the functional role suggested for the ignition phenomena. If, as proposed above, the ignition is needed to bind the separate neurons into a coherent assembly, should not this binding be terminated when activity declines during adaptation? A possible solution to this conundrum is that once the binding has been formed by the ignition process, its maintenance may require only low levels of activity. Similar suggestions have been made in modeling working memory processes (Mongillo *et al.* 2008). Clearly, additional studies are needed to clarify these issues.

Overall summary and open questions

To summarize, converging lines of evidence from human BOLD-fMRI, iEEG recordings, neurological cases and stimulation data are all consistent with a local theory of conscious content—in which local cortical regions uniquely underlie specific categories of conscious content. Real-life conscious experiences typically consist of complex, multi-dimensional, contents, necessitating integrating activity in larger cortical clusters and networks consisting of many cortical regions, but such integration is flexible and does not depend on any specific obligatory global network. Along these lines, it is proposed that the information needed to generate specific conscious contents is derived locally, within the relevant cortical regions.

Following ideas of the structuralist school, it is proposed that this information is encoded in the relational structure or geometry of each cortical region. The relational geometry can be operationally revealed through the pair-wise similarity matrix relating the activity patterns associated with different conscious contents within each category. Neuronally, such relational geometries are generated by the unique architecture of synaptic connections within each cortical region, likely involving the dense network of local lateral connections. The specific conscious content, e.g. the unique appearance of a specific face image, is formed by embedding the activation pattern associated with this unique content within the synaptic relational architecture of the cortical region specializing for the content's category. Finally, under the structuralist hypothesis, high bursts of activity (ignitions) serve as a binding mechanism whose function is to recurrently assemble the separate neurons into a coherent, functional geometry.

Supporting evidence of the local structuralist hypothesis is beginning to emerge, but many questions still remain and will require an extensive set of experiments. Among the most glaring ones are questions related to the boundaries of relational geometries—are they fixed or can be flexibly integrated into larger clusters of geometries? Can neighboring cortical areas flexibly integrate into a larger functional geometry that may thus underlie more broad aspects of conscious experience? Can different geometries be multiplexed within a single cortical region? Another important question concerns the role of non-local inputs—i.e. how do the top-down and feed-forward inputs contribute and modulate the relational architecture of each cortical region? How are the geometries modulated by plasticity and experience as well as by brain states such as sleep and arousal? Finally, direct experimental evidence for the role of lateral connections and recurrent processing to conscious awareness is still lacking. This very partial

list of open issues is daunting on the one hand but demonstrates the rich experimental landscape opened by the local structuralist perspective.

An interesting open question is whether relational geometries emerge strictly for conscious content. It is important to emphasize that in the context of the localist theory proposed here, the role of such geometries is merely a local code, i.e. a source of information, used to define the contents of conscious experience. In principle, there is no reason to assume that at least some of this information may not be used also by subliminal, non-conscious processes. The finding of similar relational geometries in artificial DCNN (Grossman *et al.* 2019) suggests that indeed such geometries play a role even in a strictly feed-forward, automatic processes. However, given the massive changes in neuronal activity associated with conscious vs. non-conscious processes, it remains an open question whether the relational geometries of, say, face representations remain the same under conscious and non-conscious conditions.

On the other hand, the fact that content-defining information may be available does not imply that this information is sufficient to underlie the emergence of a conscious experience. Here, it is proposed that an additional dynamics—that of binding through local ignitions—is necessary.

Conclusion

To conclude, I propose here a neuronal mechanism that underlies the emergence of specific conscious contents from subconscious processes in the human mind. A central pillar of the proposal is the realization that individual patterns of neuronal activation—when isolated from their similarity relationships to other patterns—are meaningless and cannot underlie conscious content. More specifically, I propose that in order for a specific content, e.g. a visual image of a face, to become conscious, a convergence of two central aspects must occur. First, there must be a burst of recurrent neuronal activation— ignition—that binds a group of neurons into a specific functional pattern. Second, the ignited pattern must reside within a local relational geometry— e.g. a face-selective area—embodied in a synaptic structure that encodes the experiential similarities and differences between the specific conscious content, e.g. a specific face and its similarities and differences all other potential face images. Large-scale activations as well as global arousal states and their physiological and pharmacological counterparts are all factors that enable and modulate the local ignitions but are not, on their own, determinants of the conscious content which is linked to the local patterns of activations.

Data availability

Data is available upon request.

Conflict of interest statement

None declared.

References

Amir Y, Harel M, Malach R. Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *J Comp Neurol* 1993;**334**:19–46.

Baars BJ. Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Prog Brain Res* 2005;**150**:45–53.

Barton JJS. Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. *J Neuropsychol* 2008;**2**:197–225.

Bassett DS, Sporns O. Network neuroscience. *Nat Neurosci* 2017;**20**:353–64.

Berkovich-Ohana A, Noy N, Harel M *et al.* Inter-participant consistency of language-processing networks during abstract thoughts. *Neuroimage* 2020;**211**:116626.

Block N. Two neural correlates of consciousness. *Trends Cogn Sci* 2005;**9**:46–52.

Broday-Dvir R, Malach R. Resting-state fluctuations underlie free and creative verbal behaviors in the human brain. *Cereb Cortex* 2021;**31**:213–32.

Chen J, Leong YC, Honey CJ *et al.* Shared memories reveal shared structure in neural activity across individuals. *Nat Neurosci* 2017;**20**:115–25.

Davidesco I, Harel M, Ramot M *et al.* Spatial and object-based attention modulates broadband high-frequency responses across the human visual cortical hierarchy. *J Neurosci* 2013;**33**: 1228–40.

Davidesco I, Zion-Golumbic E, Bickel S *et al.* Exemplar selectivity reflects perceptual similarities in the human fusiform cortex. *Cereb Cortex* 2014;**24**:1879–93.

Edelman S, Grill-Spector K, Kushnir T *et al.* Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology* 1998;**26**:309–21.

Fisch L, Privman E, Ramot M *et al.* Neural "ignition": enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron* 2009;**64**:562–74.

Foster BL, Rangarajan V, Shirer WR *et al.* Intrinsic and task-dependent coupling of neuronal population activity in human parietal cortex. *Neuron* 2015;**86**:578–90.

Fox KCR, Shi L, Baek S *et al.* Intrinsic network architecture predicts the effects elicited by intracranial electrical stimulation of the human brain. *Nat Human Behav* 2020;**4**:1039–52.

Gelbard-Sagiv H, Mukamel R, Harel M *et al.* Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* 2008;**322**:96–101.

Gerber EM, Golan T, Knight RT *et al.* Cortical representation of persistent visual stimuli. *Neuroimage* 2017;**161**:67–79.

Golan T, Davidesco I, Meshulam M *et al.* Human intracranial recordings link suppressed transients rather than 'filling-in' to perceptual continuity across blinks. *eLife* 2016;**5**.

Goldberg II, Harel M, Malach R. When the brain loses its self: prefrontal inactivation during sensorimotor processing. *Neuron* 2006;**50**:329–39.

Grill-Spector K, Kushnir T, Edelman S *et al.* Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 1999;**24**:187–203.

Grill-Spector K, Malach R. The human visual cortex. *Annu Rev Neurosci* 2004;**27**:649–77.

Grossman S, Gaziv G, Yeagle EM *et al.* Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. *Nat Commun* 2019;**10**: 4934.

Hahamy A, Wilf M, Rosin B *et al.* How do the blind 'see'? The role of spontaneous brain activity in self-generated perception. *Brain* 2021;**144**:340–53.

Harmelech T, Friedman D, Malach R. Differential magnetic resonance neurofeedback modulations across extrinsic (visual) and intrinsic (default-mode) nodes of the human cortex. *J Neurosci* 2015;**35**:2588–95.

Harmelech T, Malach R. Neurocognitive biases and the patterns of spontaneous correlations in the human cortex. *Trends Cogn Sci* 2013;**17**:606–15.

Hebb DO, Martinez JL, Glickman SE. The organization of behavior - a neuropsychological theory - Hebb, DO. *Contemp Psychol* 1994;**39**:1018–20.

Hochstein S, Ahissar M. View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 2002;**36**: 791–804.

Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 1968;**195**:215–43.

James TW, Culham J, Humphrey GK *et al.* Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. *Brain* 2003;**126**:2463–75.

Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 1997;**17**:4302–11.

Keller CJ, Davidesco I, Megevand P *et al.* Tuning face perception with electrical stimulation of the fusiform gyrus. *Hum Brain Mapp* 2017;**38**:2830–42.

Khuvis S, Yeagle EM, Norman Y *et al.* Face-selective units in human ventral temporal cortex reactivate during free recall. *J Neurosci* 2021;**41**:3386–99.

Kriegeskorte N, Kievit RA. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn Sci* 2013;**17**:401–12.

Lamme VAF, Roelfsema PR. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci* 2000;**23**:571–9.

Lau H, Rosenthal D. Empirical support for higher-order theories of conscious awareness. *Trends Cogn Sci* 2011;**15**:365–73.

Levy I, Hasson U, Malach R. One picture is worth at least a million neurons. *Curr Biol* 2004;**14**:996–1001.

Luck SJ, Hillyard SA, Mangun GR *et al.* Independent hemispheric attentional systems mediate visual search in split-brain patients. *Nature* 1989;**342**:543–5.

Malach R. Cortical columns as devices for maximizing neuronal diversity. *Trends Neurosci* 1994;**17**:101–4.

Malach R. The measurement problem in consciousness research. *Behav Brain Sci* 2007;**30**:516.

Malach R. Conscious perception and the frontal lobes: comment on Lau and Rosenthal. *Trends Cogn Sci* 2011;**15**:507.

Malach R. Neuronal reflections and subjective awareness. *Adv Consc Res* 2012;**88**:21–36.

Malach R, Amir Y, Harel M *et al.* Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proc Natl Acad Sci USA* 1993;**90**:10469–73.

Malach R, Reppas JB, Benson RR *et al.* Object-related activity revealed by functional magnetic-resonance-imaging in human occipital cortex. *Proc Natl Acad Sci USA* 1995;**92**: 8135–9.

Martinelli F, Perez C, Caetta F *et al.* Neuroanatomic correlates of visual hallucinations in poststroke hemianopic patients. *Neurology* 2020;**94**:e1885–91.

Mashour GA, Roelfsema P, Changeux JP *et al.* Conscious processing and the global neuronal workspace hypothesis. *Neuron* 2020;**105**:776–98.

Michel M, Morales J. Minority reports: consciousness and the prefrontal cortex. *Mind Lang* 2020;**35**:493–513.

Mongillo G, Barak O, Tsodyks M. Synaptic theory of working memory. *Science* 2008;**319**:1543–6.

Moutard C, Dehaene S, Malach R. Spontaneous fluctuations and nonlinear ignitions: two dynamic faces of cortical recurrent loops. *Neuron* 2015;**88**:194–206.

Mukamel R, Gelbard H, Arieli A *et al.* Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science* 2005;**309**:951–4.

Nir Y, Mukamel R, Dinstein I *et al.* Interhemispheric correlations of slow spontaneous neuronal fluctuations revealed in human sensory cortex. *Nat Neurosci* 2008;**11**:1100–8.

Norman Y, Raccah O, Liu S *et al.* Hippocampal ripples and their coordinated dialogue with the default mode network during recent and remote recollection. *Neuron* 2021.

Noy N, Bickel S, Zion-Golumbic E *et al.* Ignition's glow: ultra-fast spread of global cortical activity accompanying local "ignitions" in visual cortex during conscious visual perception. *Conscious Cogn* 2015;**35**:206–24.

Odegaard B, Knight RT, Lau H. Should a few null findings falsify prefrontal theories of conscious perception? *J Neurosci* 2017;**37**:9593–602.

Parvizi J, Jacques C, Foster BL *et al.* Electrical stimulation of human fusiform face-selective regions distorts face perception. *J Neurosci* 2012;**32**:14915–20.

Penfield W. Some mechanisms of consciousness discovered during electrical stimulation of the brain. *Proc Natl Acad Sci USA* 1958;**44**:51–66.

Pitts MA, Metzler S, Hillyard SA. Isolating neural correlates of conscious perception from neural correlates of reporting one's perception. *Front Psychol* 2014;**5**:1078.

Podvalny E, Yeagle E, Megevand P *et al.* Invariant temporal dynamics underlie perceptual stability in human visual cortex. *Curr Biol* 2017;**27**:155–65.

Privman E, Nir Y, Kramer U *et al.* Enhanced category tuning revealed by intracranial electroencephalograms in high-order human visual areas. *J Neurosci* 2007;**27**:6234–42.

Quiroga RQ, Mukamel R, Isham EA *et al.* Human single-neuron responses at the threshold of conscious recognition. *Proc Natl Acad Sci USA* 2008;**105**:3599–604.

Quiroga RQ, Reddy L, Kreiman G *et al.* Invariant visual representation by single neurons in the human brain. *Nature* 2005;**435**: 1102–7.

Raccah O, Block N, Fox KCR. Does the prefrontal cortex play an essential role in consciousness? Insights from intracranial electrical stimulation of the human brain. *J Neurosci* 2021;**41**: 2076–87.

Ramot M, Grossman S, Friedman D *et al.* Covert neurofeedback without awareness shapes cortical network spontaneous connectivity. *Proc Natl Acad Sci USA* 2016;**113**:E2413–20.

Reber TP, Bausch M, Mackay S *et al.* Representation of abstract semantic knowledge in populations of human single neurons in the medial temporal lobe. *PLoS Biol* 2019;**17**:e3000290.

Rosenthal D. How to think about mental qualities. *Philos Mind* 2010:368–93.

Santhouse AM, Howard RJ, Ffytche DH. Visual hallucinatory syndromes and the anatomy of the visual brain. *Brain* 2000;**123**:2055–64.

Saussure F. *Course in General Linguistics*, Vol. xvi. New York, NY: Philosophical Library, 1959, 240.

Schwartz L, Yovel G. The roles of perceptual and conceptual information in face recognition. *J Exp Psychol Gen* 2016;**145**: 1493–511.

Sela G, Krom AJ, Bergman L *et al.* Sleep differentially affects early and late neuronal responses to sounds in auditory and perirhinal cortices. *J Neurosci* 2020;**40**: 2895–905.

Stuss DT, Benson DF. Neuropsychological studies of the frontal lobes. *Psychol Bull* 1984;**95**:3–28.

Tononi G, Boly M, Massimini M *et al.* Integrated information theory: from consciousness to its physical substrate. *Nat Rev Neurosci* 2016;**17**:450–61.

Tootell RB, Reppas JB, Kwong KK *et al.* Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J Neurosci* 1995;**15**:3215–30.

Tsuchiya N, Wilke M, Frassle S *et al.* No-report paradigms: extracting the true neural correlates of consciousness. *Trends Cogn Sci* 2015;**19**:757–70.

Wilf M, Ramot M, Furman-Haran E *et al.* Diminished auditory responses during NREM sleep correlate with the hierarchy of language processing. *PLoS One* 2016;**11**:e0157143.

Wilkinson F. Auras and other hallucinations: windows on the visual brain. *Prog Brain Res* 2004; 144:305–20.

Zeki S. The disunity of consciousness. In: Progress in Brain Research, Banerjee R, Chakrabarti BK (ed.), *Models of Brain and Mind: Physical, Computational and Psychological Approaches*, Vol. 168. 2008, 11–268.