Guest editorial

Some reflections on (or by?) grandmother cells

Origins and variations

It was with relief that I read recently that Jerry Lettvin (1995) admits, with ‘proud shame’, to having thought of the term ‘grandmother cell’ to refer to what we now know as ... well, the grandmother cell. This is not to mean he thought of the concept of grandmother cells, merely that he thought of using that term to refer to the concept. Why is this a relief to me? Because for years I have been worried about the origin of the term, and why this particular term is used. Actually, it turns out that Lettvin thought originally of ‘mother cells’, and the term ‘grandmother’ was added only as an afterthought: having dismissed mother cells as logically untenable, he reckoned grandmother cells were a safer bet, because “grandmothers are notoriously ambiguous and often formless”. (Who was he thinking of?) This was around Christmas 1969, and Lettvin continued to present these thoughts in a lecture series over the next seven years. The earliest written reference I can trace was in a paper delivered to the Perception office on 6th December 1972. In that paper, Horace Barlow (1972) mentions grandmother cells—but only once, and then just to dismiss them. Yet only six months later, in the 14 June 1973 edition of New Scientist, Colin Blakemore was able to write: “Surely animals cannot have individual detector cells for every conceivable object they can recognise? This great debate has become known as the question of the ‘grandmother cell’. Do you really have a certain nerve cell for recognising the concatenation of features representing your grandmother?” (page 675; my italics). So why did the ‘great debate’ adopt this particular term, and why did it catch on so in the early seventies?

The first time I used the term was also during this period, when I was working (?) in Colin Blakemore’s lab. There the hot topic was the development of the primary visual cortex, and whether such properties as orientation selectivity were innate [as suggested by Hubel and Wiesel (1963)] or learned (eg Blakemore and Cooper 1970; Blakemore and Van Sluyters 1975). The reason I thought ‘grandmother cell’ such a good term was that it implied that, if the theory were true, each of us would have different wiring for our grandmother cells; hence this wiring could not be innate (since Colin would not recognise my grandmother and I would not recognise his). Hence, if all areas of cortex operate on the same principles, area 17 responses must be learned through experience too. (Several important points are assumed here: like Lettvin and Blakemore, I thought of grandmother cells as responsible for the recognition of one’s own grandmother, not just grandmothers in general. I was also glossing over such facts as: we each have two grandmothers; siblings may share those grandmothers; and grandmothers may, like mine, die before I was born, hence making me one of those individuals Lettvin tried to imagine who had no grandmother cells. I also assume Lamarckism is not correct. All these pedantries are, however, beside the main point.)

But there is another question about the grandmother-cell concept that has always puzzled me: where do the axons of grandmother cells go? Under Hubel and Wiesel’s hierarchical model, cells in earlier stages of the visual pathway were taken to transmit information to other cells at higher stages in the visual system. The notion of grandmother cells is simply the result of extrapolating that principle. So whither grandmother cell axons? Outside the visual system, I suppose—to ‘higher centres’. Fans of modular theories of the brain might then define grandmother cells as those cells that project out of the visual module—to the memory or arousal systems, for example.
Conversely, one could use grandmother cells to discover or define where the boundaries of the visual system lie. Where the 'centre of consciousness' might be is then, as always, kicked upstairs. On this model, the little man in the head, or whatever other process leads to awareness, 'looks at' the outputs of grandmother cells. Also, it receives the outputs from only those cells, so the activities of earlier stages in the visual pathway are not directly accessible to consciousness—the sites of 'preconscious processing'.

Of course, nowadays we have more sophisticated ways of looking at these problems: recurrent or 're-entrant' processing is quite the vogue (see below), and the idea that memory exists in a separate part of the brain (or, at least, could only be accessed via the hippocampus) has been replaced with the notion of the whole cerebral cortex (and indeed most of the brain) as the site of long-term memories. So, just as a grandmother cell's anatomical wiring has grown through repeated experiences of seeing a particular grandmother facilitating transmission across Hebb synapses, we can even declare that an orientation-selective cell in area 17 is orientation selective because it has a pattern of input wiring that is the engram of having seen oriented edges earlier in life.

However, there is another way of defining grandmother cells: as cells that also respond to the sound of that grandmother's voice, or even to imagining that particular grandmother, ie to the very concept of her, in an amodal, semantic, or person-identity-node sense. Conceivably, such cells might receive converging input from cells responding to the sight of grandmother and from cells responding to the sound of her voice (and so on), or they might be driven directly by lower-level feature-detecting cells without an intervening stage of modality-specific grandmother cells. Lettvin (1995) conceived of grandmother cells as representing the Kantian Ding-an-sich, though I do not see how they are supposed to have transcendental perceptual awareness. However, what is exciting about this variant of grandmother cell theory is that it provides a model for the neural basis of semantic concepts in general: the very notions of truth, justice, even God might have their own cells.

Linguists might call these sememe cells, since they represent units of meaning (although 'grandmother' is two sememes). The theory of sememe cells, however, leads to a contradiction: if grandmother cells and God cells are similar in kind, how did God cells develop through individual ontogeny? Abstract concepts are not acquired in the same way as representations of past visual experiences. So if their ontogenies are different, grandmother cells and God cells are not similar in kind.

Presumably, sociological factors, mediated via linguistic input, 'top–down' must be generating abstract-concept cells, and cells for fictional objects and people—and for grandmother cells! Other mechanisms of thinking could give rise to the general category or stereotype of 'grandmother'. The neural mechanisms that might underlie these processes remain obscure.

Earlier, Konorski (1967) had referred to 'gnostic units', which makes these cells sound as though they are semantic and not just visual entities; however, Konorski divided up the cortex purely into sensory, motor, and emotion 'analyzers' with no regions of multimodal or amodal responsiveness. Grandmother's face cells would therefore be found in the visual analyzer and grandmother's voice cells in the auditory analyzer, with associative links running directly between them.

Interestingly, Damasio and Damasio (1994) turn the theory on its head by suggesting that grandmother cell axons run backwards, to lower levels in the visual system. There they coordinate the reconstruction of complex patterns of activity across many cortical areas, and it is this whole pattern that 'represents' grandmother. This idea is reminiscent of Neisser's (1967) concept of 'analysis-by-synthesis': the brain does not store representations, it stores blueprints from which the representation may be reconstructed.
During ontogeny, the growth of this recurrent system is controlled by subcortical biological drives (Damasio 1995). Bottom-up modularists, however, must postulate that grandmother cell axons run to a special part of the brain that adds emotional 'tone' to stimuli. Putative structures for such an emotion system include the mesolimbic 'reward' system of the behaviourists, or the amygdaloid centres mediating 'fear and avoidance'. For example, faces have aesthetic value attached, and other people evoke affective reactions in us, such as liking, loathing, or loving. Moreover, we change our minds about people as we face to know them and see them behaving in various ways that we approve or disapprove of. Learning such emotional reactions means, under grandmother cell and modularity theory, changing the neural connections between the grandmother cell and the various structures in the emotions system that lead to positive or negative emotional feelings. Now, cognitive psychology pays much service to the idea that the brain contains 'models' of objects (Craik 1943). Can we extend grandmother cell theory to suggest that the firing of such cells be equated with the activation of such models? If so, we could say our emotional responses to certain people are evoked by activity in those models (subsumed by impulses travelling along grandmother cell axons to the emotion centres). To put it bluntly, we do not fall in love with people—we fall in love with the models of them that we form in our heads. This makes a lot of sense when it comes to understanding why we misunderstand people so readily: we complete the gaps in our knowledge in whatever way we can (like filling in a Kanizsa triangle, but of course more unreliably, because the human as 'object' is so much more complicated—we are therefore so often surprised by what people do). This view also makes sense of the bereavement syndrome—it is rather like amputees who experience phantom limbs. On the other hand, we now have a theory of mind that is both Kantian and at the same time Cartesian, in that it implies the emotional 'self' is looking at these cognitive models as though in a theatre.

The content of consciousness
This brings us back therefore to the question of how the firing of grandmother cells could be related to the phenomenal awareness of seeing your grandmother. I do not have space here to go into modern theories of the neural basis of consciousness, which are far more sophisticated than the serial grandmother-cell model. Instead, I would like to start by considering one of the less well studied aspects of visual perception, namely that of the background. We all know about the figure–ground problem, and of Gestalt demonstrations such as the face/vase, where figure and ground periodically reverse. Since then perceptual and cognitive psychology have concentrated on object recognition (the figure), and relatively little attention has been given to the perception of background. It was Michael Arbib (1972) who introduced the analogy between vision and making an animated cartoon: first a eel (a transparent sheet of celluloid) is laid down on which the background is drawn; the sky, ground, clouds, trees and so on. Then other eels are overlaid on which are drawn objects closer and closer to the foreground: particularly characters and other objects that move, in other words that change from frame to frame. This is not only logical but also makes it easier to draw, in that the same background can remain in place throughout the scene, and only the foreground cells have to be redrawn. In neural terms, we know that keeping track of an animal's location within its environment is, perhaps, one function of the hippocampus (O'Keefe and Nadel 1978; O'Keefe 1993). This system keeps a record of what should be present in the environment, at least in a familiar environment such as the animal's home territory. Against this background it is easy to detect any deviation from the expected pattern of incoming stimulation, in other words to detect novelty or change, to generate an arousal response and an orienting reaction that directs attention to the novel stimulus. This system is efficient, in that it reduces the redundancy of
having to analyse all the sensory input continuously. Even the effects of self-movement on the optic flow can be predicted (cf efferent eye-movement signals—why not the same for body movements?). Now imagine moving from one room in your house into another room: there is a change of background cell, slotted in to the hippocampus as the change of environment occurs. [People who know me know that I do a lot of research into sleep: at home, in the office, at conferences ... I am always collecting data. One thing that impresses me is the way that, during a visual dream, I can turn a corner and see a room whose walls are instantaneously visible with all their surface detail. How do such scenes slot in so quickly?]

But what exactly does the visual system contribute to this process of background perception? How do the various mechanisms of object recognition fit into this scheme? Grandmother cells, gnostic units: all are concerned with the foreground objects, not with the background, which may not just be a collection of distant objects (eg trees) but the uniform plane of a wall or a green field, with or without detectable surface texture, or a clear blue sky. Do we have to postulate a special kind of texture or background cell—we may call it the ‘wallpaper cell’ (which responds to the repeating pattern of wallpaper in grandma’s living room)?

But the crucial question is: how do the foreground objects block out the background? Or, how does one object, partially occluded by another, appear only partially visible rather than in its entirety? For, if the occluded object, and the background texture, are sufficiently visible to be identified, their ‘grandmother cells’ must be firing. So why don’t we just see them as unobstructed wholes? Why don’t we see completed circular ‘pacmen’ in the Kanizsa figure? One answer may be that the outputs of those cells are referred back to the precise topographic maps existing at earlier stages in the visual pathway [cf Zeki (1993), who says this can be done without the information having to go as far as forming grandmother cells]. There account can also be taken of information about each object’s location (in absolute or in retinotopic coordinates?), perhaps fed back from the parietal lobe. Some kind of prioritising must be made, based on depth relations, that is capable of blotting out completely and absolutely the occluded parts of the more distant objects and backgrounds. This problem has been relatively little studied. One interesting idea is being developed by Adelson and Wang (1996a, 1996b), who draw an analogy with the alpha channel masks used in graphics software. These carry a parallel channel of information about the transparency of each pixel within a layer (the computer equivalent of the cartoon cell). However, even the separation of figure from ground can be problematic and can vary depending on attentional (ie top–down) influences (Subirana-Vilanova and Richards 1996). Purely bottom–up theories, like the original grandmother-cell concept, are clearly no longer adequate.

Perhaps the most basic question is: what do we really become aware of? Does the firing of a grandmother cell correlate or correspond in some way with our phenomenal, visual perception of her, or only with our semantic amodal identification of her? Do we have two entirely separate mechanisms for sensory awareness and for semantic identification, or are these mechanisms partially or completely the same? This reflects a major argument in the philosophy of mind about awareness (qualia) and understanding or meaning (intentionality; eg Dennett 1978). Some theorists, as far back as Locke, assert we become aware of primitive elements of sensation: colours, movements, brightnesses, sizes, and so on; and the mind uses these to construct percepts and understandings of whole objects. These theorists assert we can see such things as ‘red’ divorced from all other aspects such as location and size [remember that one alternative to the term grandmother cell is the ‘yellow Volkswagen detector’, based on a confluence of inputs from yellow detecting cells and Volkswagen detecting cells: Harris (1980)]. This idea is not compatible with grandmother-cell theory, which asserts that basic features are processed preconsciously.
Alternatively, Konorski (1967) divided experiences into three types. ‘Perceptions’ of external objects occur when gnostic units (his term for modality-specific grandmother cells) are fired bottom-up by cells in lower regions of sensory cortex. Mental ‘images’ occur when gnostic units are fired by associative links from gnostic units in other sensory modalities. Finally, ‘hallucinations’ (including dreams) occur when gnostic units are fired by associative links and in turn activate cells in lower levels via descending connections. It is tempting to map the qualia/intentionality distinction onto Konorski’s theory: gnostic units underlie intentionality by virtue of their cross-modality links (including links to action), while activity in early sensory cortex gives rise to qualia, albeit with particular spatial locations.

Against all the above views are recent denials of a sharp division of sensory processing into two levels: preconscious and conscious (Dennett 1991), or perceptive and associative (Zeki 1993). Thompson (1995) argues we ‘see’ in our phenomenal awarenesses only objects in the real world, complete with colour, size, location, etc. He follows Merleau-Ponty (1945/1962) in denying that we see either features or objects as purely sensory events; instead, all the stimuli we become aware of have meaning. There are no qualia without intentionality. If this position is valid, we can no longer accept a grandmother-cell theory that conceives such cells provide the only route into consciousness. Whether awareness is generated in a separate module, or by a complex interplay between neurons at many different levels in one or more systems in the brain, the role of grandmother cells is not the pivotal one that the theory suggests. Indeed many workers now believe we can do without them at all (eg Zeki 1993).

So what can we conclude about grandmother-cell theory? I have distinguished several varieties and argued that some at least are of dubious validity. Other criticisms have been made by Barlow (1972, 1995), for example that too many cells would be needed and that perceptions are not isolated unique events—and that, while mothers are important, grandmothers are not! On top of all these logical flaws there are also obvious unanswered questions about the theory. For example, what happens when we learn to recognise a new object? Are vast numbers of quiescent unused cells lying around in the inferotemporal lobe waiting to be wired up, as Konorski (1967) postulated, or do we have to unwind or unlearn some old connections—if so, how is this done? Or do we only possess cells for familiar objects rather than all objects [a compromise suggested by Blakemore (1973)]?

Grandmother-cell theory may have misled people by its seductive simplicity and its apparent basis in neurophysiology (extrapolation of Hubel and Wiesel’s hierarchical model, and early reports of face and paw cells), but I think the above problems leave little room for the original form of the theory in modern views of brain function. Despite these arguments, it must be said that research in the tradition does continue, claiming for example that monkeys possess grandmother cells but only for faces (Tanaka 1996) or only for animate objects (Logothetis and Sheinberg 1996), and that learning to recognise new objects can be accompanied by changes in the stimulus preferences of inferotemporal cortical cells. However, even if it can be proved that cells exist that respond only to ‘grandmother’, either in a purely visual sense or amodally, they are at least embedded in a circuitry that must also be fully specified by the theorist before it can be claimed an account has been given of the neural mechanisms that ‘represent’ grandmother and that give rise to conscious awareness of her.

David Rose
Department of Psychology, University of Surrey, Guildford, Surrey GU2 5XH, UK
References
Adelson E H, Wang J Y A, 1996a “Layers, warps, and alpha maps in mid-level vision” Investigative Ophthalmology and Visual Science 37 S467
Adelson E H, Wang J Y A, 1996b “Layers and alpha maps as visual representations” Perception 25 Supplement, 29
Arbib M A, 1972 The Metaphorical Brain (New York: Wiley)
Barlow H B, 1972 “Single units and sensation: A neuron doctrine for perceptual psychology” Perception 1 371 – 394
Barlow H, 1995 “The neuron doctrine in perception”, in The Cognitive Neurosciences Ed. M S Gazzaniga (Cambridge, MA: MIT Press) pp 415 – 434
Blakemore C, 1973 “The language of vision” New Scientist 58 674 – 677
Blakemore C, Cooper G, 1970 “Development of the brain depends on the visual environment” Nature (London) 228 477 – 478
Blakemore C, Van Sluyters R C, 1975 “Innate and environmental factors in the development of the kitten’s visual cortex” Journal of Physiology (London) 248 663 – 716
Craik K J W, 1943 The Nature of Explanation (Cambridge: Cambridge University Press)
Damasio A R, 1995 Descartes’ Error: Emotion, Reason and the Human Brain (London: Picador)
Damasio A R, Damasio H, 1994 “Cortical systems for retrieval of concrete knowledge: the convergence zone framework”, in Large Scale Neuronal Theories of the Brain Eds C Koch, J L Davis (Cambridge, MA: MIT Press) pp 61 – 74
Dennett D C, 1978 Brainstorms (Cambridge, MA: MIT Press)
Dennett D C, 1991 Consciousness Explained (London: Penguin)
Harris C S, 1980 “Insight or out of sight? Two examples of perceptual plasticity in the human adult”, in Visual Coding and Adaptability Ed. C S Harris (Hillsdale, NJ: Erlbaum Associates) pp 95 – 149
Hubel D H, Wiesel T N, 1963 “Receptive fields of cells in striate cortex of very young, visually inexperienced kittens” Journal of Neurophysiology 26 994 – 1002
Konorski J, 1967 Integrative Activity of the Brain (Chicago, IL: University of Chicago Press)
Lettvin J Y, 1995 “J.Y. Lettvin on grandmother cells”, in The Cognitive Neurosciences Ed. M S Gazzaniga (Cambridge, MA: MIT Press) pp 434 – 435
Logothetis N K, Sheinberg D L, 1996 “Visual object recognition” Annual Review of Neuroscience 19 577 – 621
Merleau-Ponty M, 1945/1962 The Phenomenology of Perception translated by C Smith (London: Routledge and Kegan Paul)
Neisser U, 1967 Cognitive Psychology (New York: Appleton-Century-Crofts)
O’Keefe J, 1993 “Kant and the sea-horse: An essay in the neurophilosophy of space”, in Spatial Representation Eds N Elian, R McCarthy, B Brewer (Oxford: Blackwell) pp 43 – 64
O’Keefe J, Nadel L, 1978 The Hippocampus as a Cognitive Map (Oxford: Clarendon)
Subirana-Vilanova J B, Richards W, 1996 “Attention frames, frame curves and figural boundaries: the inside/outside dilemma” Vision Research 36 1493 – 1501
Tanaka K, 1996 “Inferotemporal cortex and object vision” Annual Review of Neuroscience 19 109 – 139
Thompson E, 1995 Colour Vision (London: Routledge)
Zeki S, 1993 A Vision of the Brain (Oxford: Blackwell)