1 Grandmother Cells and Distributed Representations

Simon J. Thorpe

Summary

It is generally accepted that a typical visual stimulus will be represented by the activity of many millions of neurons distributed across many regions of the visual cortex. However, there is still a long-running debate about the extent to which information about individual objects and events can be read out from the responses of individual neurons. Is it conceivable that neurons could respond selectively and in an invariant way to specific stimuli—the idea of "grandmother cells"? Recent single-unit recording studies in the human medial lobe seem to suggest that such neurons do indeed exist, but there is a problem, because the hit rate for finding such cells seems too high. In this chapter, I will look at some of the implications of this work and raise the possibility that the cortical structures that provide the input to these hippocampal neurons could well contain both highly distributed and highly localist coding. I will discuss how a combination of STDP and temporal coding can allow highly selective responses to develop to frequently encountered stimuli. Finally, I will argue that "grandmother cell" coding has some specific advantages not shared by conventional distributed codes. Specifically, I will suggest that when a neuron becomes very selective, its spontaneous firing rate may drop to virtually zero, thus allowing visual memories to be maintained for decades without the need for reactivation.

Introduction

The Distributed vs. Localist Representation Debate

One of the longest-running and thorniest debates in the history of research on the brain concerns the nature of the representations that the brain uses to represent objects, and specifically the question of whether individual neurons may encode specific objects and events (Barlow, 1972; Gross, 2002). The debate has recently received new impetus, with the publication of a significant review paper by Jeff Bowers (Bowers, 2009), which has been followed by a series of commentaries (Plaut and McClelland, 2010; Quiroga and Kreiman, 2010). In addition, there have been a series of fascinating studies on single-unit responses from the human medial temporal lobe that have raised numerous questions about the link between single-unit activity and perception (Quian Quiroga et al., 2005). There is clearly something special that happens when we recognize a familiar visual stimulus. Virtually any such visual stimulus will doubtless activate millions, maybe hundreds of millions of neurons in our visual system. Many of these are presumably involved in generic processing tasks that will take place irrespective of whether the image is recognized or not. For example, simple cells in V1 will presumably signal the presence of an edge with a particular orientation at a particular point in the visual field, irrespective of whether the corresponding object can be recognized or not. But most scientists would probably accept that at some level in the brain, quite possibly relatively high levels in the visual hierarchy, there are neurons that are directly involved in encoding the presence of the object. The debate concerns the way in which those neurons do the encoding. One view, currently quite popular among scientists, is that the representation at the neuronal level is distributed across large numbers of neurons, none of which needs to be specifically tuned to a particular object. At the other extreme, some researchers have proposed that for some highly familiar objects, there may be neurons that respond very selectively to that object—a view often jokingly referred to as "grandmother cell" coding.

This difference between "local" and "distributed" coding models has become a very hot topic in recent years, partly because there have been a number of reports of single neurons that have been recorded from the medial temporal lobe of human patients undergoing presurgical investigations for the treatment of intractable epilepsy. At first glance, some of these cells seem to have many of the properties that one might expect to find if grandmother cell coding was true. But, as some of the authors of the studies have pointed out, there are a number of puzzling features about these cells that do not seem to fit with the simple grandmother cell view. In this chapter, my plan is to look in more detail at some of the issues. I will argue that while the cells reported in the human medial temporal lobe may not be what one would predict for a true localist coding scheme, there may be other explanations of the results. Furthermore, I will argue that there are some good computational arguments for using grandmother cell–like coding in at least some cases. However, I will also argue strongly that there is no requirement to choose in favor of only one type of coding. Rather, as I have argued previously, it is likely that the brain simultaneously uses both highly distributed and localist encoding, quite possibly within the same bit of neocortex.

A Test Case: Recognition in RSVP Streams

To make the nature of the debate clear, consider the following experiment that you can try on yourself by downloading a set of movie files from the following site: <ftp://www.cerco.ups-tlse.fr/Simon/Movies>. Each movie sequence is a string of highly varied photographs of animals drawn from a range of sources that are presented in a Rapid Serial Visual Presentation (RSVP) sequence at 10 frames per second. Ever since the classic studies of Molly Potter in the 1970s, we have known that our visual systems can process images at this rate and that we can spot a particular target image (such as a "boat" or a "baby") very effectively under such conditions, even when only a verbal description of the stimulus is provided (Potter, 1975, 1976). In the case of the demo sequences, all the images are of animals except one, and the task is to report whatever image in the sequence does not fit. Whenever I have shown the first example sequence during lectures, almost everyone will immediately notice that the sequence contains an image of Mona Lisa. In a sense, the fact that Mona Lisa is easy to spot may not be that surprising, since it is effectively one of the most familiar images in Western civilization. Each of us has almost certainly seen it hundreds if not thousands of times. And, given that it is a 2D painting, the visual pattern that it produces on our retina is relatively stable with changes of viewing angle. Size is certainly not prespecified, since we can recognize the Mona Lisa at essentially any scale, but relative to most 3D objects that we encounter, its appearance is nevertheless relatively standardized, making it possible to imagine that recognition could be achieved with a relatively simple "pattern-matching" approach. However, by looking at the other movies, you will be convinced, I hope, that this sort of effortless recognition occurs for many other types of object. For example, one of the sequences contains a photograph of the Statue of Liberty. Again, almost every one will notice its presence in the sequence of images, despite the fact that (unlike Mona Lisa), there are effectively an infinite number of different viewing angles that would work. This certainly makes life harder for the recognition mechanism that is being used by the visual system. In another of the sequences, there is a scene from a restaurant with someone dressed up as Mickey Mouse. Again, a remarkably high number of people who see the sequence will immediately report that they saw Mickey Mouse, even though they had no reason to expect such an image. Indeed, the other "distractor" images in the sequence are all animals, and in a sense, so is Mickey Mouse (albeit a somewhat special one). Why then, do people almost invariably notice the intruder in the sequence? My personal view is that they notice such intruders because some sort of high-level representation is activated in the brain, and that this representation gets noticed because it does not fit with the rest of the context.

There are a number of points that we can make on the basis of this sort of demonstration. One point concerns the question of whether all the images in the sequence need to be processed fully by the visual system, or whether it might be sufficient to only process the one (Mona Lisa, Statue of Liberty, or Mickey Mouse) that we actually notice. This seems to me to be very implausible. It is not because the other twenty or thirty images in the sequence presented at ten images per second cannot be reliably reported that they were not fully processed. Indeed, it seems difficult to imagine how the brain could determine whether any particular image in the sequence was worth noticing without processing them all fully. Rather, it seems more likely that all the images are being processed and that the intelligence of our visual system is demonstrated by the fact that we are automatically able to determine whether or not a particular image is sufficiently important to merit being made the center of attention.

Some Distinctions

A key issue that I want to address here concerns the nature of the neural representations that are activated in such a situation. A first point to make is that there are good reasons to believe that the brain could potentially use a complete range of representational schemes. Consider a classic 3-layer neural network architecture composed of a layer of input units, a layer of output units, and between them a layer of so-called hidden units. To make the problem clear, imagine that the input layer corresponds to a simple 8 × 8 retina, and the output layer corresponds to a set of 128 responses, with one response for each of the 128 members of the ASCII character set (see figure 1.1). The desired input–output function is to generate the appropriate output when a particular character is presented on the "retina." Theoretical studies have shown that with sufficient units in the hidden layer, such a system can implement any arbitrary input function, but there are many different ways in which the function could be implemented at the level of the hidden units. One would be to use just 7 hidden units and the 7-bit ASCII code, illustrated in figure 1.1. This is a perfect illustration of a completely distributed coding strategy, since to know what is being represented, it is necessary to know the state of all 7 hidden units. None of the units actually "means" anything on its own, and an experimenter who was recording the response of any of the neurons to changing inputs would be unable to make sense of why the neuron was active for any given stimulus, because effectively, the assignment of each neuron to the set of stimuli is arbitrary. Each unit will be active for 50 percent of the input patterns, which means that the coding is in fact very efficient. Indeed, that is precisely why the ASCII code was chosen as a way to encode text based information within computers. Note also, that there are a very large number of different ways in which the 7 units could be used to represent all 128 characters—each is effectively as good as any other.

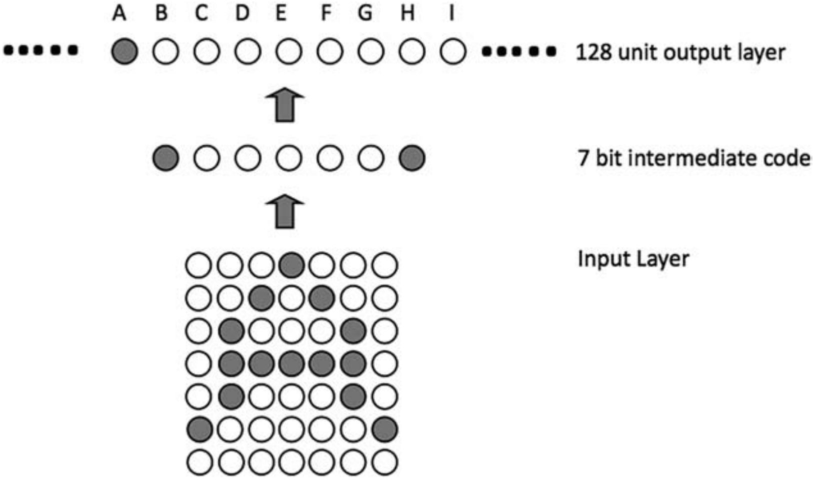

Figure 1.1
Distributed coding in the ASCII code. Each of the seven units in the hidden layer participates to the coding of one of the 128 characters in the ASCII set. However, none of these units has activity that is specifically related to any particular stimulus.

At the other end of the spectrum, one could imagine a hidden layer with 128 units, one for each of the characters in the set. For any one input pattern, only one of the units need be active. This would be an example of extreme localist coding—effectively corresponding to grandmother cell coding. Clearly, both extremes of representation could potentially be used for representation at the neuronal level. Is there any way of deciding which if any of these different schemes is actually used in the brain?

Reading out from Distributed Representations

The distinction between distributed and grandmother cell–based representations has become an increasingly hot issue in recent years for a number of reasons. One is that recent advances in multivoxel-based analysis of fMRI data mean that it has been possible to show how, using the distributed pattern of activity over a large number of voxels, one can make inferences about the stimulus identity and category. Such data provide a very clear demonstration that information can indeed be extracted from the distributed pattern of activity within a cortical region. This has been demonstrated using fMRI voxel activation levels (Kriegeskorte et al., 2007) but more recently, the approach has been extended to recordings from intracerebral electrodes in epileptic patients (Liu et al., 2009), as well as to single-unit recording studies from both monkey inferotemporal cortex (Kiani et al., 2007) and the medial temporal lobe in humans (Quian Quiroga et al., 2007). One particularly significant aspect of this sort of work is that the decoding strategies used, although sophisticated, do not require the existence of mechanisms that could not be implemented with relatively simple neural circuits. Suppose that it can be demonstrated that by applying a classification algorithm to the activation levels of a few hundred voxels in some part of the ventral processing pathway it is possible to make some judgment about the stimulus—for example, whether or not a face is present, and maybe even whether the face is familiar or not. If the classification procedure could be implemented using a neural circuit, then it follows that the brain could also derive the same information simply by forwarding the same pattern of activation to another brain area where neurons could potentially learn to make the same distinction. It might also be that individual neurons within the region would also be in a position to learn to make the same categorical distinctions. In this case, the "readout" neurons would show a more local representation, as is the case for the third layer in figure 1.1.

Does the fact that it is possible to extract information from the activity patterns of large numbers of neurons (Quian Quiroga and Panzeri, 2009) mean that there is no need for the brain to make information explicit at the level of individual cells? In other words, does this recent work actually allow us to say anything about whether or not grandmother cells really exist or not?

A Highly Specialized Neuron in the Orbitofrontal Cortex

One reason why I think one should be careful before assuming that it will be possible to derive all the information needed to interpret the function of a particular brain region from imaging studies comes from my own experience as a doctoral student in Edmund Rolls's lab at Oxford in the late 1970s. We were recording single-unit activity in the monkey orbitofrontal cortex (OFC) using behavioral tasks that we thought were likely to be interesting given the known effects of OFC lesions. Specifically, lesioned animals (and humans) were known to have severe problems in task shifting, for example, when performing visual discrimination tasks with reversals. We therefore explicitly tested this by training monkeys to perform a go/no-go visual discrimination task for fruit juice reward, and then periodically reversing the rule. Thus, initially the monkey might be responding to a green "go" stimulus and withholding responses to a red "no-go" stimulus. However, we then reversed the contingency with the result that the monkey made one "mistake" and received a drop of saline after responding to the green stimulus. After weeks or months of training, he would immediately reverse his strategy and start responding systematically to the other stimulus, until the next reversal occurred. More than three hundred neurons were recorded during the performance of this reversal task, most of which failed to show any significant activity changes related to the task. However, a handful of cells showed some quite remarkable responses during reversal, and one particular cell showed a quite amazing response, illustrated in figure 1.2 (Thorpe

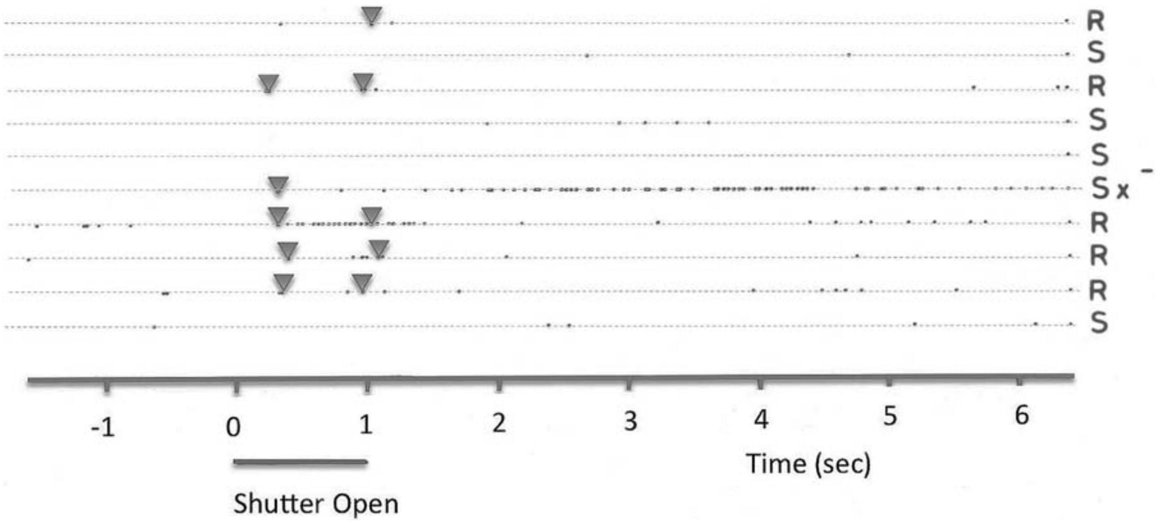

Figure 1.2
Activity of a single neuron in monkey orbitofrontal cortex with activity very specifically related to reversals in a go/no-go visual discrimination task. On each trial, the monkey is shown one of two different stimuli through a shutter that opens for 1 second. One of the stimuli means that he can lick a tube for fruit juice reward, whereas the other stimulus means that he should not lick in order to avoid receiving saline. Licks are indicated by the inverted triangles. Without warning the meanings of the two stimuli are inverted ("Reversal"), at which point the monkey makes one mistake. This was followed by a strong burst of activity from the neuron that lasted several seconds. A second smaller burst occurred on the next correct trial. Adapted from Thorpe et al. 1983.

et al., 1983). Following each reversal of the rule, the neuron showed a very strong increase in firing rate that lasted for several seconds. There was even a second burst of firing following the first correctly performed trial with the new rule. It is therefore difficult to imagine that the neuron is simply responding to the punishment, or to the making of an error. Rather, the neuron appeared to form part of a highly specific circuit that was specifically related to the performance of the reversal task. Given that the monkey was a true expert at performing the task, having spent months performing such reversals, I cannot help thinking that maybe the existence of such a neuron is a direct reflection of the automaticity of the behavior following training. In the current context, it is particularly important to realize that only one such cell was found out of hundreds recorded. It is therefore relatively unlikely that the activity of the neuron could be seen at the level of more global measures of brain activation, such as event-related potential recording or fMRI. Many other examples of highly specific but rare responses in individual neurons can be found in the literature.

The Problem of Inferring Neuronal Selectivity from Global Measures

The existence of this sort of highly specific, yet rare, neuronal response within a cortical area raises an important issue. Global activity measures certainly provide evidence to support the idea that a particular brain region could be involved in performing a certain cognitive task. However, it is probably impossible to make inferences about the degree of specialization of individual neurons on the basis of these global measures. In principle, one can even imagine a situation where the global activation measures provide no evidence for selectivity whatsoever, and yet where there might still be strong selectivity at the single-unit level. Indeed, the reverse can also be true, since there are cases when there is "global" activity in the absence of spiking activity (Sirotin and Das, 2009).

Consider some of the very interesting fMRI-based studies that have shown that it is possible to read out the orientation of a grating from the pattern of activation seen across voxels in V1 (Haynes and Rees, 2005; Kamitani and Tong, 2005). Such techniques rely on the existence of the local variations in preferred orientation within V1. While the size of the orientation columns is small relative to the resolution of the fMRI technique, there are nevertheless sufficient variations in local selectivity to allow the technique to be used. However, it is important to realize that it did not have to be that way. If the neurons selective to different orientations were really mixed up completely at the local level, nothing would be visible at the level of the voxels because each voxel would contain neurons coding all the different orientations. Thus, if information can be extracted from looking at the relatively coarse pattern of activity seen in imaging studies, this is certainly compatible with the hypothesis that the structure is involved in the processing of the stimulus attributes. However, the opposite is not true. The absence of differential fMRI activation does not imply that the structure has no role to play.

Grandmother Cells in the Human Medial Temporal Lobe?

Having made a few general points about the problems of relating results from imaging studies with responses seen at the single-unit level, I would now like to move on to the interpretation of the fascinating series of papers that have described the responses of single units in the human medial temporal lobe. The studies demonstrate that such neurons can have remarkably invariant responses to a wide range of different stimuli that effectively correspond to the same object or concept. One of the earliest such studies was a paper from 2000 describing neurons that would respond to a wide range of different photographs of animals (Kreiman et al., 2000), but over the past few years we have seen reports of neurons that would respond to many different photographs of a particular actress (the famous "Jennifer-Aniston cell"), or even to the name of the person written in text (Quian Quiroga et al., 2005). And it is now clear that the same individual neuron can fire selectively to the same stimulus presented via a number of different sensory modalities—vision, text, voice (Quian Quiroga et al., 2009), implying a truly remarkable degree of invariance. Another fascinating result is the fact that when the stimuli are masked, and the duration of the presentation is so short that the subject can only report the nature of the stimulus on some limited percentage of trials, there is a remarkably high correlation on individual trials between whether the neuron responds and whether the subject can report the nature of the stimulus (Quian Quiroga et al., 2008b).

At first sight, such results might appear to provide strong support for the notion of grandmother cell coding. While it is true that there have not actually been any reports of neurons that fire exclusively to the patient's grandmother, the neurons do tend to respond best to stimuli that are personally relevant to the patient, responding in particular to members of the patient's family or members of the experimental team (Viskontas et al., 2009). However, there is a problem with such a view, which stems from the fact that the hit rate for finding such cells appears to be much higher than one would expect if that part of the brain was really using such an explicitly localist coding scheme.

The critical issue is the number of different objects that the system needs to be able to encode. One widespread source of confusion concerning localist coding is the belief that it would require having one neuron to code every possible stimulus that can be identified. As I have argued previously (Thorpe, 1995; Thorpe and Imbert, 1989b; Thorpe, 2002), this can be easily demonstrated to be erroneous. Consider the output of the retina via the optic nerve, which contains roughly 1 million axons. Even if we only consider the situation where each axon can either be "on" or "off," this means that there are $2^{1,000,000}$ possible patterns that can be presented. If we assume that we need one neuron to encode each one of these patterns, this would need roughly $10^{300,000}$ neurons. Assuming a reasonable size for each neuron, this would require that the brain be larger than the known universe—clearly, not the strategy used by natural selection! The error in the argument comes from assuming that we need to know the total number of possible stimuli. In fact, it is the number of visual categories that is the real number of interest. There are no hard and fast numbers for this, but Irving Biederman suggested about 30,000 distinct visual categories (Biederman, 1987), and I myself have suggested a somewhat higher number based on the number of entries in a large encyclopedia (Thorpe and Imbert, 1989a). If we suppose that the real number is something like 100,000, this would mean that the probability that any given cell could be activated by a given familiar stimulus should on average be 1 in 100,000, assuming all stimuli to be equally represented. However, it is clear that in the human MTL recoding studies, the chances of finding a cell that responds appears to be much higher than this. Indeed, the probability of activation of a given cell to the sorts of stimuli used in these experiments has been estimated to be 0.54 percent (Waydo et al., 2006).

This point is well made in the paper entitled "Sparse but not 'grandmother cell' coding in the medial temporal lobe" (Quian Quiroga et al., 2008a). In a typical experiment, the researchers are able to record from a few dozen cells simultaneously. During a morning session, they show a set of roughly 100 photographs about five times each to the patient in a random order. Often, they will find one or two cells that respond well to one of the images. Let us suppose that one of the effective images was a photograph of Bill Clinton. During the lunch break, the researchers then constitute a new set of test images, including several other images of Bill Clinton that are then used to analyze the neuronal responses during an afternoon session. This is how they are then able to confirm that a single cell is able to respond to a wide range of different images (and even text strings or speech) that correspond to the same object.

The critical point is that the hit rate during the morning session is much higher than one would expect if each neuron in the medial temporal lobe were a grandmother cell in a sort of library containing hundreds of thousands of possible objects.

Distributed Coding and the Totem Pole Cell Hypothesis

In their paper, they conclude that even if individual cells can respond in an invariant way to a highly diverse set of different stimuli that correspond to the same object, it is highly probable that the same cell might also respond to other completely unrelated objects. Rafi Malach has called this idea the totem pole cell hypothesis (Malach, personal communication). According to this idea, each cell has a number of different "faces," and might simultaneously be able to respond invariantly to say "Bill Clinton" but also to some other completely unrelated stimuli—such as the "Taj Mahal" or an episode of *The Simpsons*, for example. Clearly, the probability that the experimenters might hit on two or more totally unrelated stimuli just by chance would be very low. Nevertheless, cells responding to two separate stimuli have been seen occasionally, so the idea is nevertheless a real possibility that deserves to be tested more explicitly. Note that Rafi Malach's totem pole cell hypothesis is a clear case where object identity could only be deduced if one has access to the responses of multiple neurons. Thus, if one such "totem-pole cell" responded to Bill Clinton, the Taj Mahal, and the Simpsons, and another cell responded to another set of stimuli including Bill Clinton, the fact that both cells responded on a given trial could be used to determine that Bill Clinton was present.

An Alternative Hypothesis for the Significance of MTL Responses: Temporal Tagging

While the high-hit-rate issue appears to argue against the idea that the neurons in the human hippocampus are truly instances of grandmother cell coding, the totem pole hypothesis is perhaps not the only option for explaining the phenomenon. Given the well-known implication of medial temporal lobe structures in memory, it may be interesting to think of how the responses of such neurons might fit within an alternative memory related hypothesis. Suppose that one of the key roles of the hippocampus is to keep track of a subset of all the possible objects and events that we are able to recognize, namely, those that have been experienced in the relatively recent past. According to this view, the neurons in the hippocampus are not a dictionary of all the objects that can be recognized, but rather a dictionary of recently experienced events. Although speculative, this hypothesis fits a number of interesting features of the medial temporal lobe.

First, it has been known for decades that synapses in the hippocampus are very plastic and show long-term potentiation (LTP) following strong activation (Bliss and Collingridge, 1993; Bliss and Lømo, 1973). This potentiation can last for days and even weeks, meaning that a sensory input that is repeated is likely to produce a stronger response to the second presentation, even when the interval between presentations is a matter of weeks. Second, in a study of neuronal responses in a region close to the anterior thalamus that could potentially receive information from the medial temporal lobe, Rolls and colleagues described neurons that had the remarkable property of responding strongly to effectively any visual stimulus that had been seen recently (Rolls et al., 1982). A particularly remarkable finding is the fact that such neurons could have visual responses that have latencies as short as 130 ms. This form of invariant response to familiar stimuli is a major challenge for computational models, because it implies that there must be massive convergence from higher-order visual areas to allow such a general response. How might this be achieved? It is just conceivable that somehow all possible visual stimuli converge in one processing stage to produce a generic “familiarity” response, but this seems unlikely. Alternatively, it might be that the brain determines familiarity individually for recently encountered objects before putting them all together. Could this be what is seen at the level of the single-unit responses seen in the human medial temporal lobe?

One of the strategies used by the team performing the human MTL recordings when choosing the initial set of images for testing is to specifically ask the patients for information about their favorite TV programs and movies, together with their preferred actors. This strategy could well be one of the reasons for the high success rates seen in the experiments but leaves open the issue of whether it is the fact that the patients are highly familiar with the stimuli or whether it is the fact that they may have seen them relatively recently that is critical. It appears that the neurons can sometimes respond strongly even on the very first presentation of a particular photograph during the morning recording session (Pedreira et al., 2010), and this might be taken as evidence that recency is not critical for obtaining a response. However, as far as I am aware, it would be difficult to rule out the possibility that the patient has seen the stimulus elsewhere in the relatively recent past.

The hypothesis is therefore that these hippocampal neurons may effectively be keeping track of a relatively limited subset of all possible objects that can be recognized, namely, those that have been experienced within say the past few weeks. The precise duration of this temporal tagging period may not be strictly fixed, but could potentially be related to the maximal duration of LTP mechanisms, that is to say, periods that could extend to several weeks. The critical question would now be to estimate the total number of different objects, scenes, and events that are typically tracked during this period. It may well be thousands or maybe more, but it certainly will be a lot smaller than the total number of objects that we are capable of recognizing, which will probably be orders of magnitude higher. If the numbers really are in the range of thousands, then this might well be able to account for the anomalously high hit rate seen in the hippocampus. Clearly, if during any recording session, it is possible to record from several dozens of neurons, and each is tested with 100 different images, many of which are likely to correspond to stimuli that have been experienced recently, it would be enough to have a system that was only tracking a few thousand stimuli to be relatively confident about finding at least one neuron that could be activated during any given experiment.

One critical implication of this view of the hippocampus is that a neuron that currently shows highly selective and invariant responses to a particular input (for example, "Jennifer Aniston") is not required to remain selective to that particular stimulus indefinitely. Specifically, if one were able to record again from the same cell two years later (clearly a technical impossibility), it might well be found to respond to something completely different. The idea is that a neuron will remain selective for a particular stimulus as long as that stimulus is reexperienced reasonably frequently, that is to say every few weeks or so.

We therefore have at least two quite different views about the highly selective responses reported in human medial temporal lobe, and the puzzling fact that the hit rate for finding these cells seems to be excessively high—too high to be compatible with the idea that the hippocampus contains a complete set of grandmother-type cells. The first is the suggestion by Rafi Malach that the cells may be multifaceted, like totem poles, and each capable of responding to many totally different objects. The second option is that the cells may only be responding to a subset of the large number of recognizable objects, namely the subset corresponding to objects that have been seen or experienced in the relatively recent past. And, of course, these options are not mutually exclusive.

The Origin of Highly Selective Responses

It is important to realize that for both models, we still need to explain how the neurons are able to achieve this high level of selectivity. More specifically, the fundamental question concerns the nature of the coding scheme being used by the neurons that provide the input to such neurons. We know that the neurons in the hippocampus will get their visual inputs from structures in the ventral stream, including areas such as perirhinal and entorhinal cortex. And before that, these structures in turn will be receiving from the hierarchy of processing areas that make up the ventral stream. What sort of coding is being used in these structures?

Here, there are clearly a number of theoretical possibilities. To make things clear, consider a hypothetical neuron that responds selectively to any visual input that corresponds to a particular person—whether it is the patient's brother or a well-known celebrity. What sorts of neurons are providing the inputs to such a cell? One possibility is that the hippocampal cell can somehow generate invariant responses to a wide range of physically different visual inputs directly from a true distributed code at the previous layer. If this is the case, then it is clear that this would involve mechanisms that we currently cannot understand.

In contrast, one way of obtaining an invariant response that we can understand involves pooling together outputs from a population of cells that is each selective to a particular instance of the stimulus. This sort of pooling mechanism has already been suggested for generating selectivity to views of the head that are invariant to the angle of view, based on pooling together different responses, each of which is selective to a particular viewing angle (Perrett et al., 1987). Indeed, this form of view-specific recognition mechanism now seems to be increasingly seen as the most plausible way of generating invariance (Wallis and Rolls, 1997). But if true, it is clear that the neurons providing the input would themselves have to be quite selective, and indeed they might look quite a lot like the hypothetical grandmother cell, although without the remarkable invariance seen in the medial temporal lobe neurons.

In the end, an answer to this sort of question will have to wait until we have single-unit recordings from the cortical regions that provide the inputs to the medial temporal lobe. For the time being, such recordings are simply not available, largely for technical reasons. Few researchers working in humans have been able to record from individual neurons in the ventral stream structures that provide the highly processed information. Obviously, there are far more data available from work on monkeys, but it is difficult to extrapolate from one species to the other. While we can be confident that the human patients can indeed recognize "Jennifer Aniston,"