1. Space in the mammalian brain

1.1 Where am I?

You wake up one morning in a hotel room. You have just arrived in Cozumel, Mexico for a short vacation. But right now, coming out of a dream, you do not quite remember this fact. The room is dark, there is only a dim light coming from the shutter on the right side somewhere at a distance. You do not know where you are. Or why you are there. It cannot be your home. Or is it? The bathroom would be right on the left of the door. But where is the door? This is a transient disorientation that many have occasionally experienced. Puzzling and disturbing. But it does not last long. Our normal mode of being includes a persistent awareness of where we are in space and this awareness seems to be tightly connected with our ability to remember facts. For example, to remember that we are on vacation in Cozumel.

Figure 1.1. The hippocampus

A) *Sagittal MRI with medial aspect of human hippocampus. The arrow shows the alveus of the posterior part of the hippocampus. (From Gardner and Hogan, Clinical Anatomy 18: 481-487, 2005)*
B) *Rat Hippocampus (From Teyler and DiSchenna, Brain Res. Bull. 12:711-719, 1984)*
C) *Rodent hippocampal circuitry drawn by S. Ramon Y Cajal. (1911)*

The ability to form certain kinds of new memories as well as the ability to locate ourselves in space are both dependent on the hippocampus, a small yet very important seahorse-shaped structure inside the temporal lobes of the brain. The hippocampus (Fig. 1.1a) has attracted the attention of Neuroanatomists since the beginning of modern neuroscience. The oldest drawings of the hippocampal network (Fig. 1.1b) are credited to the early giant of neuroanatomy, Santiago Ramon y Cajal, who was among the first to propose the "Neuron Doctrine". What is now common knowledge, was then, at the end of the 19th century, the revolutionary idea that nerve cells form a communication network through a dense pattern of interconnections.

Damage to the hippocampus is known to cause anterograde amnesia, the inability to form new memories. This has been studied in the famous clinical case of H.M. (Milner et al., 1968), a man who had the medial temporal lobes on both sides removed to treat a severe form of epilepsy. The intervention cured the epilepsy, but it had a very severe side effect: H.M. was left without the

ability to form new memories of facts. Nevertheless, he maintained the ability to learn new motor skills. Brenda Milner trained HM to execute difficult manual tasks. For example, tracing the shape of a star while looking at it through a mirror (Milner, 1962). After practicing for a few trials on one day he was tested on a subsequent day. On the second day he retained the level of skill that he had acquired in the previous day. But he could not remember the fact that he had practiced that task earlier. In fact, he was quite surprised at his own unexpected dexterity in such a difficult activity. Dramatically more common than the story of HM is Alzheimer disease, which causes a progressive loss of memory capacity and disorientation, associated with a severe damage to the hippocampus and the cerebral cortex.

1.2 Space representations in the Mongolian Gerbil

The amnesia and disorientation following hippocampal damage are cues that the ability to locate ourselves in space and in time is strongly associated with the function of this brain region. This spatial skill is not merely human, as it is present in a multitude of animal species. A well-known example is a species of butterfly, the Monarch, who migrates every year from central Mexico to Canada and then back. On more moderate geographical scales, rodents have notable skills to navigate across various terrains and in various lighting conditions.

Mongolian Gerbils are the small rodents that some people like to keep as pets. Besides being a cute domestic animal, the Mongolian Gerbils is a skilled explorer and scavenger. The name comes from the hostile environment these gerbils come from, the Mongolian steppes. These are large, hostile and arid territories, where food is scarce and long trips are necessary for the gerbils to find seeds and other nutrients and bring them back to their burrows for storage. These small rodents are among our best models of animal exploration.

Figure 1.2. Search-pattern of one gerbil. A: plan-view of landmark (circle) and reward-site (triangle). Calibration is 100 cm. Landmarks are not shown to scale. B: paths followed by the gerbil when released from different points within the arena. C: Cumulative search-distribution resulting from 21 tests of 60 s duration. In this and subsequent figures., the animal's position in relation to the landmark-array is given to within a cell 11 cm across and 13.3 cm high. The blacker the cell, the more time the gerbil spent there. Time spent in each cell is expressed as a percentage of that spent in the most visited cell. The latter is filled by 25 dots arranged in a square. (From Collett et al. 1986)

In a set of now classic experiments, Collett, Cartwright and Smith (1986) trained Mongolian gerbils to seek food inside a circular arena. They hid a sunflower seed under a layer of gravel

scattered on the floor. The environment was carefully designed so that only the objects, or "landmarks" placed by the investigators could serve as spatial references for the gerbils. The arena had a diameter of 3.5 meters and was placed inside a room whose walls were painted black. A light bulb illuminated the floor and left the walls in the shadow. Somewhere in the center of the arena, the investigators placed a white cylinder that was clearly visible to the gerbil. The sunflower seed was hidden always at a distance of 50cm and at fixed angle from the base of the cylinder (Fig. 1.2). Collett and colleagues placed the gerbil in the arena and allowed it to find the seed for several training trials. On each trial, they changed the position of the landmark cylinder, but they maintained the position of the seed relative to the landmark. After a few trials, the gerbils learned to move straight toward the seed and retrieve it.

Once the investigators verified that the gerbils had learned this simple task, they removed the seed and watched as the gerbils searched for their reward. The Fig. 1.2C displays the performance of the gerbils in the test trials. The area of each black square is proportional to the total time spent by a gerbil searching within the corresponding region of space. This simple diagram demonstrates that the gerbil learned to predict the location of the seed in relation to the landmark. Importantly, the gerbil searched in the correct location even when starting different trials from different locations. That is, the gerbils had formed a spatial memory of the location of the goal in terms of its distance and orientation with respect to the landmark. And even if the investigators took care of removing the external visual cues, the ability to identify the correct direction with respect to the cue indicates that the gerbils had a sense of a fixed "north arrow" which could not have been supplied by the landmark alone, since this was uniform in color and cylindrical in shape[i].

To gain a deeper knowledge about space representations in the gerbils, Collett and colleagues changed the pattern of landmarks once the gerbils had learned to retrieve the seed. Their objective was to investigate how the relation between the goal and the landmarks was understood and exploited by the gerbil's brain. In one experiment, the gerbils were first trained to retrieve the seed inside an arena with two identical landmarks (Fig. 1.3). The seed was placed in an equilateral triangle arrangement, at equal distance from the two landmarks. How would the trained gerbil react if the distance between the landmarks were unexpectedly doubled? We make the working hypothesis that the gerbil's brain forms a representation of the environment and that this representation, which we call an "internal model", is updated as the gerbil experiences new spatial relations among itself, the landmarks and the seed. We may think of two possible outcomes. In one case, the gerbil's internal model of the environment preserves the shape of the

seed-landmark pattern and searches at one location which is at the same distance from the two landmarks. Since the distance between the landmarks has doubled, the distance of the seed from the line joining the landmarks must also increase so that the seed is at the vertex of an isosceles triangle. Alternatively, the internal model preserves the expected distance and orientation of the seed with respect to either landmark. This means that we now would have not one but two distinct sites where the seed might be at. This second outcome is what Collett and coworkers observed: the gerbils during the test trials searched for the seed at two distinct positions, each one corresponding to the position of the seed relative to the closest landmark.

Figure 1.3. Euclidean versus scaling transformations. A: Gerbils were trained to retrieve the seed placed at equal distance from two identical landmarks. B: Search pattern after training. C: When the distance between the landmarks is increased, the gerbils do not use a scaling rule. Instead, they look for the seed at two sites. Each searched site is at the same distance from the corresponding landmark in the trained configuration. This means that the Gerbil understood that the environment had changed. The landmarks have been displaced and the expected distance between the seed and each landmark has remained fixed. D: In another experiment, gerbils were trained with two different landmarks: a dark aluminum cylinder on the left and a white cylinder on the right. The seed, again, was placed in an equilateral triangle configuration as in A. The black square indicates the location from which the gerbils start searching. E: search distribution when the landmarks are rotated by 180 degrees. The gerbil searches in the correct location with respect to the triangular arrangement of the landmarks. The black square indicates the point where the gerbil was released for this test F: Search distribution with the landmarks in the same configuration used for training. F: Search distribution with the landmarks rotated by 180 degrees. In this case, the gerbils apply the same transformation to the expected site for the seed.(From Collett et al. 1986)

Consider a new scenario (Fig. 1.3D) in which the gerbils are trained with two different landmarks, an aluminum cylinder and a white cylinder. Again, the seed is placed in a triangular arrangement, at equal distance from the two cylinders. After the gerbils have been trained to find the seed, they are tested with a 180 degree rotation of the landmark arrangement: their positions are now interchanged, but their distances are preserved. In this new environment, the gerbil searches the seed at a location that is consistent with this rigid rotation, that is at a location which is reflected over the segment joining the two landmarks (Fig. 1.3F). These results suggest that not only the gerbils have a clear sense of distance, but they also can distinguish between a rigid transformation that preserves the distances between the landmarks and can be interpreted as a rotation and a transformation in which the distances have been altered. In the latter case, the gerbils' brains deduce that the landmarks have moved. Accordingly, the seed is searched in relation to each landmark separately. Instead, in the former case, the gerbil's brain understands that distances have not changed. Therefore, the scene - landmarks and seed - has undergone either

a rotation or a translation or both. Or, alternatively, the viewpoint has changed by a relative movement of the gerbil with respect to the fixed scene. The gerbil's brain, however, is not willing to accept the possibility that there has been a dilation of the environment. A dilation might occur through development, as the animal grows in size, but certainly not in the time scale of the experiment. As we will discuss in more detail, the two situations – rigid motions and scale changes - are representative of two types of affine transformations of space. These are all linear transformations, and only a subclass of them preserves the distance between points. This is the subclass of isometries or Euclidean transformations. As we move around in our environment, the points and objects around us remain unchanged. Therefore, our representation of these points and objects undergoes transformations of this Euclidean type.

1.3 Some general properties of space maps in psychology and mathematics

There is an ancient debate between two views of animal intelligence. In one view - we call it the "reductionist" view - what we think of as intelligence is nothing but the apparent manifestation of automatic behaviors through which an organism seeks to acquire the largest amount possible of good stuff or, conversely, seeks to avoid bad stuff. Good and bad stuff are the less technical names of positive reward and negative reward. The reductionist viewpoint was once known as the stimulus-response or S-R theory. At its origin is the work of Clark Hull who investigated learning as a consolidation of stimulus-response associations (Hull, 1930). This conceptual framework has had a revival in theories of optimal control (Todorov and Jordan, 2002) and reinforcement learning (Sutton and Barto, 1998), based on solid mathematical principles as well as on empirical observations.

In the other view, that we call the "cognitive" view, organisms do not simply respond to internal and external stimuli, but they create knowledge. As they act in their environment, they acquire and maintain information that may not be immediately conducive to reward but may be later used to this purpose. Edward Tolman was an early champion of the cognitive view (Tolman, 1948). He vehemently opposed Hull-stimulus-response approach. According to Tolman, when a rat moves inside a maze in search of food "something like a field map gets established in the rat's brain." The studies of Tolman as well as of other experimental psychologists of the time were mostly carried out on rats. The rats were placed within more or less complicated mazes, with corridors, curves and dead-end. Tolman described one such experiment as being particularly supportive of the cognitive view. The rats entered a starting chamber at the base of a Y-shaped maze. They

moved forward and were to choose between the right and left arms of the Y. At the end of each arm there was an exit. Near the right exit there was a bowl of water and near the left exit there was a bowl of food. In an initial phase, the rats were satiated with food and water before entering the maze. They did not care about drinking or eating at the end of the maze. However, they wanted to find an exit quickly. Sometimes they took the right exit and sometimes the left with no particular preference. After a few repetitions of these initial trials, the rats were divided into two groups. One group was made hungry and the other was made thirsty before entering the maze. It turned out, Tolman reported, that the thirsty rats went immediately to the right, where there was water, and the hungry rats went immediately to the left, where there was food. He concluded that the rats during the first phase of the experiment learned where the food and the water were despite the fact that they did not receive any water or food reward. In the second phase, as they became hungry or thirsty, they went to the right place. The name for this kind of learning is "latent learning" because it is not associated to the delivery of reward and Tolman saw it as crucial evidence against the S-R theory.

While there is a natural dialectic tension between reductionist and cognitivist views, these views are mutually incompatible only in their most extreme versions. In the initial part of the experiment, the rats entered the maze without interest for water or food. However, they were already endowed with the notion that food and water are important items from past stimulus-response associations. Then, the unexpected presence of food and water was registered as a salient event. One can say that such an event is interpreted by the brain as an error over the expectation that the maze was merely an empty path. This type of unexpected event is known to trigger learning and the formation of new memories. Thus, the pairing between stimuli and responses or, in this case the association of motor commands with their potential to generate reward does not have to be restricted to narrow temporal contiguity. In this text we are not considering the formation of maps in opposition to the mechanisms of reinforcement. It is abundantly evident that both are expressed in our brains. The challenge of reconciling stimulus-response mechanisms with the formation of cognitive maps may well lead to a deeper understanding of biological learning.

Before considering how maps can be formed, we must ask what is a space map. In mathematics a “mapping” establishes a correspondence between the elements of two sets. Take, for example, a typical street map of Paris. In this case we have a correspondence between the points on a sheet of paper, a small planar surface, and points in the French capital. The space of Paris is three-

dimensional, whereas the space of the map is two-dimensional. Therefore the street map involves a projection from 3 to 2-D. The difference in dimensions imposes some restrictions on the way we go between the map and the space it represents. Each point in Paris has an image in each point of the planar surface of the map. But each point of the map corresponds to a whole line - a vertical line – in Paris. One can find different images for different buildings. But all the floors in a building have the same image on the planar map. We see here a first important point about topographical maps. They compress information so that not everything is represented but only what can be of some utility. The Argentinean writer Jorge Luis Borges wrote a short story about a fictional emperor who asked a team of cartographers to create a map so detailed to contain literally everything that was inside his empire: every grain of sand, every speck of wood, etc. The cartographers succeeded in their mission. But at the end, it turned out that their map, being so absolutely complete, was also perfectly useless.

On the street map of Paris, we find an index where we read something like "Eiffel Tower, E4." This means that the object Eiffel Tower is at the intersection of the row labeled E with the column labeled 4. The map is organized in a grid, which divides the territory into a finite number of squares. Each square is identified by two symbols, corresponding to the rows and columns of the grid. In most street maps, one set of symbols are letters and the other are integers. This is a trick to give a distinct identity to rows and columns. The letters, like the natural numbers follow an order: A, B, C, D... They only differ from numbers in that operations such as sums and subtractions are not explicitly defined. This is because the average tourist uses the map only to locate objects. However, if numbers were used in place of generic labels, then one could carry out arithmetic operations and place these operations in correspondence with objects in the real world. For example, one could add the displacements from landmark L1 to landmark L2 and from landmark L2 to landmark L3 to derive the displacement from landmark L1 to landmark L3. This is where a more rigorous mathematical definition of a mapping becomes most valuable.

Figure 1.4. Mapping the circle. Left: The real line is placed in correspondence to the lower half-circle by drawing a line from the center of the circle, through the point P. The intersection with the real line is the number x . This establishes a one-to one correspondence between the points in the lower half- of the circle and the whole real line. Right: Two circular arcs, AC and BD, are mapped by two charts onto two real intervals, (x_A , x_C) and (x_B , x_D). Note that the two charts have different "gains", as the projecting distances between the real lines and the circle are different. They also have overlapping regions in their domains (the arc BC). A collection of charts that cover the whole circle is called an "atlas".

The connection between numbers and "things" -for lack of a better term- is central to the development of topology and geometry. The prime example of this is the concept of the real line. A line is a geometrical entity, made of a continuum of points. The real line, indicated as , is a simple and fundamental correspondence between the real numbers - extending from $-\infty$ to $+\infty$ - and the points of a straight line. The correspondence established by is bijective: each point on the line corresponds to one and only one real number and each real number corresponds to one and only one point on the line. Then, real numbers and points on the line can be used interchangeably. What happens if instead of a real line we consider some other geometrical object? For example, consider a circle (Fig. 1.4). Because the real line is in some ways equivalent to the real numbers, we ask if the points on the real line can be placed in a bijective correspondence to the points on the circle by some type of projection. However, things are now more complicated. We see that, by projection, all the points in the lower half circle can be placed in correspondence with the entire real line, so that the two points on the "equator", Q and R, correspond to $-\infty$ and $+\infty$ respectively. The right panel of Fig. 1.4 illustrates how the entire circle is mapped over multiple segments on the real line. Each segment is placed in correspondence with a portion of the circle, i.e. with an arc. The mapping on each segment is an example of what is called a "chart". By combining multiple charts we may cover the entire circle and obtain what is called an "atlas". This mathematical terminology was borrowed from geography and from the ordinary concept of a world atlas, as a collection of charts that, put together, cover the entire globe. In building an atlas, it is of great importance to insure an precise correspondence in the regions where contiguous charts are overlapping. In a world atlas there are identical regions in different pages covering contiguous areas of the globe. Likewise, to make an atlas of the circle in Fig. 1.4, we need to rescale the two charts on the left of the figure so as to have a consistent mapping across charts.

Figure 1.5. Neural charts. The model of Samsonovich and McNaughton depicts the activities of a collection of 100 hippocampal place neurons, by locating each neuron over a Cartesian x-y plane that represents the extrinsic space in which the rat is moving. The activity map is a snapshot of the activities over the entire set of recorded place cells when the rat was passing by the central location of the chart. (from Samsonovich and McNaughton, 1997, drawing on th ebottom portion is from Eichenbaum et al, 1999)

The mathematical ideas of maps and charts have become increasingly relevant to describe the pattern of neural activities in the hippocampus. This is illustrated in Fig. 1.5, where the activities recorded from about 100 neurons in the hippocampus are represented over a region that

corresponds to a 62x62cm box in which a rat was free to move. These data come from an experiment of Matthew Wilson and Bruce McNaughton (Wilson and McNaughton, 1993) but were arranged in this particular representation in a subsequent article by Alexei Samsonovich and McNaughton (Samsonovich and McNaughton, 1997). Isolated locations on the xy plane correspond to the activity of recorded hippocampal neurons. This correspondence establishes a chart that relates the locations of space explored by the rat to the activity of a family of neurons in the hippocampus that are called "place cells" and were discovered in the early 1970's.

1.4 Place cells

The neurons studied by Samsonovich and McNaughton were discovered twenty years earlier by John O'Keefe and Jonathan Dostrovsky (O'Keefe and Dostrovsky, 1971). They were studying the activity of neurons in the rat's hippocampus and saw that a small fraction of these cells (about 10%) became active when the rat was placed in particular locations and was oriented in particular directions. In these early experiments at the Department of Anatomy of the University College in London, the rats were placed and kept by hand in different sites. Being interested on the formation of space maps in the brain and being aware of the work of Tolmann, O'Keefe and Dostrovsky immediately realized the importance of this relatively small sample of hippocampal neurons. In a later study, O'Keefe and David Conway (O'Keefe and Conway, 1976) had the rats moving within a maze populated with peculiar objects. This is well described by O'Keefe and Lynn Nadel in "The hippocampus as a cognitive map" (O'Keefe and Nadel, 1978) :

> "The environment consisted of a 7 ft square set of black curtains within which was set a T-shaped maze. On the walls formed by the curtains were four stimuli: a low-wattage light bulb on the first wall, a white card on the second, and a buzzer and a fan on the third and fourth, respectively. Throughout the experiment the location of the goal arm of the T-maze and the four stimuli maintained the same spatial relationship to each other, but all other spatial relations were systematically varied." (page 205)

To record the neural activity from hippocampal cells while the rat was exploring the maze, and to place these activities in relation to the place where they occurred, the investigators developed a very ingenious system. Remember, this is 1970, a time when computers were still in their infancy and video recorders were not yet on the scene:

> "Rats were taught a place discrimination in this environment. They were made hungry and taught to go to the goal arm as defined by its relation to the four stimuli within the

> curtains in order to obtain food. After they had learned the task, place units were recorded. In order to relate the firing of these units to the animal's position in the environment, advantage was taken of the fact that these units have low spontaneous activity … outside the place field. Each action potential from the unit was used to trigger a voltage pulse which, when fed back to a light-emitting diode on the animal's head, produced a brief flash of light. A camera on the ceiling of the environment photographed the spots, recording directly the firing of the unit relative to the environment." (page 206)

Figure 1.6. The firing of a place unit when a rat is on the T-shaped maze inside the cue-controlled enclosure. Each dot represents one action potential. Four ground trials are shown in A-D in which the T-maze and the cues on the wall have four different orientations relative to the external world. The unit fires when the rat is in the start arm when it is on the side close to the buzzer, regardless of the orientation relative to the external world. E and F show two ground trials with the start arm rotated 180° so that it is on the side close to the light. There is no unit firing in the start arm. (From O'Keefe and Nadel, 1978)

The activity recorded from one hippocampal "place unit" is illustrated in Fig. 1.6. The dots indicate the location at which the activity was detected while the rats moved into the T-maze. The cell in this case became active only when the rat was near the buzzer and when the relative locations of the 4 landmarks were preserved. The cell was not sensitive to global rotations of the landmark systems, or to the particular physical arm of the maze that happened to be near the buzzer. These and subsequent studies confirmed that place cells became active when the rat was moving at specific locations and with specific heading directions with respect to the set of landmarks that collectively define a spatial frame of reference.

After the seminal work of O'Keefe and colleagues, the last two decades of the past millennium have seen a flourishing of studies on space representations in the hippocampus. While there is still much to understand about this neural code and its relation to episodic memory, there is no doubt that the place neurons are capable to represent single locations in extrapersonal – or "allocentric" – space in a way that is consistent with the presence of a well-organized system of geometrical charts inside the brain of rodents. This has rapidly led to functional imaging studies of the human hippocampus. Of particular significance is the work of Eleanor Maguire and colleagues, who studied the brains of London taxi drivers. One of the studies (Maguire et al., 2000) revealed that licensed taxi drivers have an expanded posterior hippocampus, compared to controls that do not drive taxis. This structural study was complemented by functional imaging

that revealed significant differences in the activities of a collection of brain regions that includes the right hippocampus (Maguire et al., 1997). Functional imaging studies have less spatial and temporal resolution than electrophysiological recordings, which monitor the activities of isolated neurons on a millisecond scale. However, functional imaging offers the opportunity of looking at activity patterns across the whole brain. This has revealed that information processing associated with the representation of space spans a broad network of structures. The hippocampus is only one of them. But it is a very prominent one, as was recently demonstrated in a study by Demis Hassabis, Eleanor Maguire and collaborators (Hassabis et al., 2009). The place cells investigated by O'Keefe and others with recording electrodes are relatively rare and there has not yet been evidence for any topographical organization over the hippocampus. Place cells representing two nearby locations of space may be relatively distant from each other. Vice versa, two cells that are in nearby hippocampal sites may become active at two rather different and distant locations in allocentric space. Does this mean that there is no particular organization of neuronal populations in the hippocampus?

Hassabis and collaborators took a pragmatic approach to this question: If there is any functional organization in the activity of hippocampal neurons representing spatial locations, then it should be possible for a computer program to analyze hippocampal population activities and figure out where one is located in allocentric space. In a way, they took advantage of the poor resolution of functional MRI, where activities can only be discriminated to a limit of about $1mm^3$, the approximate size of a "voxel". While this may appear to be a small size, one cubic millimeter of hippocampal gray matter may actually contain between 10^4 and 10^5 neurons. Therefore a speck of activity detected by fMRI originates from a rather large population of cells. If there were no particular organization in the distribution of place cells, then each voxel would contain more or less the same proportion of the same place cells. The resulting activity snapshot would look like a random blur. In that case, it would be virtually impossible to look at fMRI images of the hippocampus and make a good guess about what space region has "caused" the detected activity. There would be at most a diffuse pattern of signals indicating that the hippocampus is active in a spatial task, but no possibility to extract a space code. To test this possibility, Hassabis and colleagues asked subject to play a virtual reality game while lying in a MR scanner. Subjects were presented with the image of a room, with chairs, tables and other objects that provided a spatial reference. Their task was simply to "move" within the room by pressing on arrow keys. Once at a target location, the view was switched down to the floor mat. This was the same image

at all locations, so as to avoid any visual cue about the place in the room. While in this location and without feedback, a functional image was acquired.

This operation was repeated at four different locations. A standard pattern classification algorithm was required to determine the locations based on the activities observed over a large region of the medial temporal lobe, which included the hippocampus. The algorithm was able to determine the location with high accuracy (> 80%) based only on the activity over the hippocampal region. This provides new evidence – albeit not conclusive - supporting the hypothesis that there is some topographical structure in the population activity over the hippocampus. In this case, activity charts, such as the one of Fig. 1.5, derived from neural recordings, might actually represent the internal model of the extra-personal space. But remember that only a fraction of the hippocampal cells are place cells. The studies on amnesic patients, like HM, had revealed that the hippocampus is critical to the formation of new memories. There is a close connection between the memory of an event and the location in space where the event occurred. We remember where we were when the twin towers were hit on September 11 of 2001. As stated by Hassabis and colleagues, this spatial representation may form "the scaffold upon which episodic memories are built". At the end of the next chapter, we will present a computational argument in support of this statement.

1.5 Grid cells

What is the neural mechanism that leads to the formation of space cells in the hippocampus? While this question remains to be answered, an important clue has come from studies of May-Britt Moser, Edward Moser, Marianne Fyhn and collaborators, who discovered in 2004 an intriguing pattern of activities in the entorhinal cortex of freely moving rats (Fyhn et al., 2004). The entorhinal cortex is part of the parahippocampal complex and is a major source of input to the hippocampus. Part of it can be seen in the top portion of Cajal's drawing, near the letter A (Fig. 1.1C). Some neurons in the entorhinal cortex express a very peculiar and impressive geometrical pattern. Like place cell in the hippocampus, these neurons become active when the animal moves across certain regions of space. But unlike the place cell, the entorhinal neurons display a regular periodic structure (Fig. 1.7): they have distinct peaks of activity placed at the vertices of a grid of equilateral triangles!

Figure 1.7. Firing fields of entorhinal grid cells. a) Nissl-stained section indicating the recording location in layer II of the dorso-medial entorhinal cortex of a rat. b) Firing fields of three simultaneously recorded cells as the rat moved within a large circular arena Cells names refer to

tetrode (t) and cell (c). The left column shows trajectories of the rat with superimposed firing locations (dark spots). The middle column is a gray-scale map of the recorded activity (black: no activity). The peak rates are indicated on the side of each diagram. Note the distribution of activity peaks over the vertices of a grid of equilateral triangles. (from Hafting et al. 2005)

A simple and elegant mathematical analysis by Trygve Solstad, Edward Moser and Gaute Einevoll sheds some light on the significance – if not on the origin - of this pattern (Solstad et al., 2006). Let us begin by asking how a pattern of "peaks", similar to the activities of grid cells may come with a structure of equilateral triangles. Suppose we have a periodic function of space: a standing sine wave with wavelength λ. Three such waves are depicted in Fig. 1.8, with $\lambda = 3$ length units. These could be meters, inches or centimeters --- it does not matter for the present discussion. Instead, it matters that the three sine waves are oriented in three directions, 60 degrees apart from each other. Let us give a mathematical form for these sinusoids. They map each point on the plane, $\mathbf{r} = \begin{bmatrix} x & y \end{bmatrix}^T$ into a number

$$F_i(\mathbf{r}) = \cos\left(\mathbf{k}_i^T \mathbf{r}\right) + 1 \qquad i = 1, 2, \text{ or } 3. \tag{1.1}$$

Figure 1.8. Summing three sinusoidal functions of space ($F_1 + F_2 + F_3$) with orientations that differ by 60 degrees results into a periodic distribution of equispaced peaks (right) in an equilateral triangle configuration. Note the similarity of this simple interference pattern with the activity patterns of the entorhinal grid cells (Figure 7).

The three vectors $\mathbf{k}_i$ represent the wave fronts and are oriented in three directions 60 degrees apart from each other. Their amplitude is the spatial frequency of the wave in radians per unit lengths:

$$\begin{aligned} \mathbf{k}_1 &= \frac{2\pi}{\lambda}\begin{bmatrix} \cos(\theta) & \sin(\theta) \end{bmatrix}^T \\ \mathbf{k}_2 &= \frac{2\pi}{\lambda}\left[\cos(\theta + \frac{\pi}{3}) \quad \sin(\theta + \frac{\pi}{3}) \right]^T \\ \mathbf{k}_3 &= \frac{2\pi}{\lambda}\left[\cos(\theta + 2\frac{\pi}{3}) \quad \sin(\theta + 2\frac{\pi}{3}) \right]^T \end{aligned} \tag{1.2}$$

The orientation of the first wave front $\mathbf{k}_1$ is the angle θ. In Fig. 1.8, $\theta = 0$ and the front is perpendicular to the x-axis. Adding 1 to the cosine functions insures that the range of each wave function remains positive $\left(0 \le F_i \le 2\right)$. When the three functions are added together, they form the interference pattern shown on the right part of Fig. 1.8:

$$\begin{aligned}\phi(\mathbf{r}\,|\,\theta,\lambda) &= F_1(\mathbf{r}) + F_2(\mathbf{r}) + F_3(\mathbf{r}) \\ &= \cos(\mathbf{k}_1^T\mathbf{r}) + \cos(\mathbf{k}_2^T\mathbf{r}) + \cos(\mathbf{k}_3^T\mathbf{r}) + 3\end{aligned} \tag{1.3}$$

Figure 1.9. Interference pattern. The dotted lines represent the wave fronts of Figure 8. These are lines at which each sine waves reach their maximum values. The arrows are the directions of the sine waves, i.e. the directions of the vectors k_1, k_2 and k_3 (see main text). The intersections of three wave fronts are the points at which their sum reaches the maximum value. Simple trigonometry shows that the distance between two such peaks is slightly larger than the wave length of each wave component (the distance between parallel dotted lines)

What we obtain by this simple summation of waves resembles the pattern of firing observed in the entorhinal grid cells (Fig. 1.7). The spacing, d, of the grid depends upon the wave length of the wave fronts, λ (Fig. 1.9):

$$d = \frac{2}{\sqrt{3}}\lambda \tag{1.4}$$

Each grid cell does not inform the rat's brain about the location at which the rat is. It only indicates a set of possible locations at which the rat could be. In that regard, we may see this as a particular coordinate system, like longitude and latitude. If we know our latitude we know a set of places where we may be. To know our position on the sphere we need both the latitude and the longitude. As we shall see, the grid cells can be used as coordinates in a similar way. But many more than two coordinates are needed to specify a position.

1.6 Grid cells to place cells: functional analysis

The French mathematician Jean Baptiste Joseph Fourier discovered two centuries ago the possibility of constructing arbitrary continuous functions by adding trigonometric functions with different frequencies. Fourier series are infinite sums that in the limit converge upon continuous functions. Most remarkably, any continuous function over a finite interval can be obtained as a Fourier series:

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx) \tag{1.5}$$

Daniel Bernoulli first suggested such an infinite series in the late 1740s as he was working on a mathematical analysis of vibrating musical strings. However, Bernoulli was unable to solve for the coefficients of the series. Fourier's great accomplishment was determining the values of the

coefficients. For example, for a function $f(x)$ over the interval $x=[-\pi,+\pi]$, Fourier demonstrated the following:

$$\begin{aligned} a_0 &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)dx \\ a_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)dx \\ b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)dx \end{aligned} \tag{1.6}$$

Briefly, Fourier arrived at his solution by doing the following. First, he integrated the left and right sides of Eq. (1.5) over the range $[-\pi,+\pi]$ and then solved for a_0 by noting that the integrals of the trigonometric functions vanish over this range. Then, to derive the coefficients a_n and b_n he used a more clever observation, which is fundamental to modern functional analysis. He noticed that

$$\int_{-\pi}^{\pi} \cos(nx)\sin(mx)dx = 0 \tag{1.7}$$

for all integer values of m and n. However, the integral

$$\int_{-\pi}^{\pi} \cos(nx)\cos(mx)dx \tag{1.8}$$

and the integrals

$$\int_{-\pi}^{\pi} \sin(nx)\sin(mx)dx \tag{1.9}$$

vanish only when $n \neq m$. Otherwise,

$$\int_{-\pi}^{\pi} \sin^2(nx)dx = \int_{-\pi}^{\pi} \cos^2(nx)dx = \pi\,. \tag{1.10}$$

To derive each coefficient a_n and b_n, Fourier multiplied both sides of Eqn. (1.5) by the corresponding trigonometric function --- i.e. by $\cos(nx)$, for a_n and by $\sin(nx)$ for b_n. For example, to derive a_3 one multiplies both sides of Eqn. (1.5) by $\cos(3x)$ and calculates the integrals[ii] in:

$$\int_{-\pi}^{\pi} f(x)\cos(3x)dx = \frac{1}{2}a_0\int_{-\pi}^{\pi}\cos(3x)dx + \sum_{n=1}^{\infty}a_n\int_{-\pi}^{\pi}\cos(nx)\cos(3x)dx +$$

$$+\sum_{n=1}^{\infty}b_n\int_{-\pi}^{\pi}\sin(nx)\cos(3x)dx = \qquad (1.11)$$

$$= a_3\int_{-\pi}^{\pi}\cos^2(3x)\,dx = a_3\cdot\pi.$$

The Fourier's representation of a function as a sum of other functions has a powerful algebraic and geometric interpretation. The functions that appear in the sum of Eqn. (1.5) are formally equivalent to vectors forming a basis in a vector space. While ordinary geometry is only 3-dimensional, vector spaces can have an unlimited number of dimensions. What matters is that the elements that form a basis be mutually independent - like the three unit vectors pointing along the x, y and z axes. But to be independent, vectors do not need to be mutually orthogonal. In ordinary 3D space, independence means that a vector that lies outside a plane cannot be obtained by adding vectors on that plane. Thus, one cannot obtain a vector sticking out of a plane by adding vectors on that plane. In symbols, if vectors $\boldsymbol{\varphi}_1$, $\boldsymbol{\varphi}_2$, and $\boldsymbol{\varphi}_3$ are linearly independent, then we cannot write $\boldsymbol{\varphi}_3 = a_1\boldsymbol{\varphi}_1 + a_2\boldsymbol{\varphi}_2$. One other way to say this is that the equation:

$$a_1\boldsymbol{\varphi}_1 + a_2\boldsymbol{\varphi}_2 + a_3\boldsymbol{\varphi}_3 = 0 \qquad (1.12)$$

can be true only if all the three coefficients, a_1, a_2 and a_3 are all zero. This is readily extended to an arbitrary number of vectors: N vectors $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_N$ are linearly independent if and only if

$$a_1\boldsymbol{\varphi}_1 + a_2\boldsymbol{\varphi}_2 + \ldots + a_N\boldsymbol{\varphi}_N = \mathbf{0} \qquad (1.13)$$

implies that all the a_i's are zero. Extending this further to an infinite number of independent vectors, we obtain the Fourier series, as in Eqn. (1.5). And, going even further to a *continuum* of vectors, we have the Fourier transform. But let us limit this discussion to a finite number of independent vectors.

Now, suppose that the sum Eq. (1.13) is non-zero, i.e.:

$$a_1\boldsymbol{\varphi}_1 + a_2\boldsymbol{\varphi}_2 + \ldots + a_N\boldsymbol{\varphi}_N = \mathbf{f} \neq \mathbf{0} \qquad (1.14)$$

The vector $\mathbf{f}$ in belongs to the N-dimensional vector space V_N *spanned* by the basis $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_N$. How can we use this basis to represent vectors in a higher-dimensional space? This can be achieved by *approximation*. In this case, the linear combination of the basis vectors cannot generate exactly the higher-dimensional vector. But it can get as close as possible to it.

Consider a vector $\mathbf{g}$, in a higher dimensional space, V_M, which includes V_N as a subspace ($M > N$). To gain an immediate intuition, one may think of N=2 and M=3. V_3 is the ordinary 3D space, with an associated Cartesian reference frame. V_2 is a planar surface, passing by the origin of V_3. The following discussion extends to spaces of higher dimension. We wish now to find the vector in V_N that is as close as possible to $\mathbf{g}$. The intuitive solution to this problem is to look for the projection of $\mathbf{g}$ over V_N. Suppose that we have a basis for V_M which includes the basis in V_N, augmented by M-N vectors, $\boldsymbol{\varphi}_{N+1}, \boldsymbol{\varphi}_{N+2}, \ldots, \boldsymbol{\varphi}_M$, orthogonal to V_N. In this basis, the vector $\mathbf{g}$ has a representation

$$\mathbf{g} = b_1\boldsymbol{\varphi}_1 + b_2\boldsymbol{\varphi}_2 + \ldots + b_N\boldsymbol{\varphi}_N + b_{N+1}\boldsymbol{\varphi}_{N+1} + \ldots + b_M\boldsymbol{\varphi}_M . \tag{1.15}$$

Note that the Fourier expansion of Eq. (1.5)looks much like Eq. (1.15), with infinite terms. The first part of this representation,

$$\hat{\mathbf{g}} = b_1\boldsymbol{\varphi}_1 + b_2\boldsymbol{\varphi}_2 + \ldots + b_N\boldsymbol{\varphi}_N \tag{1.16}$$

is the projection that we are looking for. We know the basis vectors, $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_N$, but we do not know the coefficients $b_1, b_2, \ldots, b_N$. To find them we use the *inner product* operation and we exploit the fact that the inner product of the basis vectors in V_N with the vectors $\boldsymbol{\varphi}_{N+1}, \boldsymbol{\varphi}_{N+2}, \ldots, \boldsymbol{\varphi}_M$ is zero by hypothesis, because these vectors are orthogonal to V_N. Let us step back. We need to remember that the inner product of two vectors produces a number[iii]. Here, we adopt the convention to use angled brackets to denote the inner product, as in $\langle \boldsymbol{\varphi}_1, \mathbf{g} \rangle$. In $\mathbb{R}^N$ we calculate the Euclidean inner product by multiplying component by component and by adding the results. In vector-matrix notation this is $\mathrm{u}^T \cdot \mathrm{v} = u_1 v_1 + u_2 v_2 + \ldots + u_N v_N$.

However, here we use a more general notation that is not restricted to Euclidean spaces. To derive the coefficients $b_1, b_2, \ldots, b_N$, we begin by taking the inner product of both sides of Equation (1.15) with each of the N basis vectors. This produces a system of N linear equations:

$$\begin{cases} \langle \boldsymbol{\varphi}_1, \mathbf{g} \rangle = \hat{g}_1 = \Phi_{1,1} b_1 + \Phi_{1,2} b_2 + \ldots + \Phi_{1,N} b_N \\ \langle \boldsymbol{\varphi}_2, \mathbf{g} \rangle = \hat{g}_2 = \Phi_{2,1} b_1 + \Phi_{2,2} b_2 + \ldots + \Phi_{2,N} b_N \\ \qquad\qquad \ldots \\ \langle \boldsymbol{\varphi}_N, \mathbf{g} \rangle = \hat{g}_N = \Phi_{N,1} b_1 + \Phi_{N,2} b_1 + \ldots + \Phi_{N,N} b_1 \end{cases} \tag{1.17}$$

with

$$\Phi_{i,j} = \left\langle \boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j \right\rangle \tag{1.18}$$

Note that while Equation (1.15) contains vectors - $\mathbf{g}$ and the $\boldsymbol{\varphi}_i$'s - and numbers – the b_i's - Eq. (1.17) contains only numbers. This is a system of N equations in N unknowns. A compact form for it is

$$\hat{\mathbf{g}} = \Phi \mathbf{b} \tag{1.19}$$

with

$$\hat{\mathbf{g}} = \begin{bmatrix} \hat{g}_1 \\ \hat{g}_2 \\ \cdots \\ \hat{g}_N \end{bmatrix} \quad \Phi = \begin{bmatrix} \varphi_{1,1} & \varphi_{1,2} & \cdots & \varphi_{1,N} \\ \varphi_{2,1} & \varphi_{2,2} & \cdots & \varphi_{2,N} \\ \cdots & \cdots & \cdots & \cdots \\ \varphi_{N,1} & \varphi_{N,2} & \cdots & \varphi_{N,N} \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_N \end{bmatrix}$$

In this notation, vectors are represented as matrices with a single column containing all the vector components. The matrix Φ is called the Gramian of the vectors $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_N$, after the Danish mathematician Jorgen P. Gram. Equation (1.19) provides us with a straightforward solution for the coefficients of Eq. (1.15):

$$\mathbf{b} = \Phi^{-1} \hat{\mathbf{g}} \tag{1.20}$$

The only requirement for deriving $\mathbf{b}$ using the above expression is that the inverse of the matrix Φ exist or, equivalently that the determinant of Φ does not vanish. This condition is insured by the fact that the vectors $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_N$ in Eq. (1.15) form a basis for V_N [iv].

To sum, so far we have shown that starting from a set of N independent vectors in V_N it is possible a) to represent all vectors in V_N and b) to find the vector in V_N that lies closest to an arbitrary vector in a higher dimensional space, V_M. But what if the vectors are not all linearly independent? Then, they live in a space V_K of dimension K, lower than N. In this case, it is still possible to use the construct that led to Equation (1.19). Now, however, the Gramian determinant is zero and the matrix cannot be inverted. We can still derive the projection of $\mathbf{g}$ over V_K by using the *pseudoinverse* of the Gramian. This is usually indicated by a * superscript, as in Φ^* and the equation for $\mathbf{b}$ is quite similar to Eq. (1.20):

$$\mathbf{b} = \Phi^* \hat{\mathbf{g}} \tag{1.21}$$

There are many ways to calculate the pseudoinverse of a matrix. Here, we limit ourselves to list its four defining properties[v]:

1. $\Phi\Phi^T\Phi = \Phi$
2. $\Phi^T\Phi\Phi^T = \Phi^T$
3. $(\Phi\Phi^T)^T = \Phi\Phi^T$
4. $(\Phi^T\Phi)^T = \Phi^T\Phi$

Note that Eq. (1.21) is more general than Equation (1.20), since the pseudoinverse of a matrix is equal to the standard inverse, whenever the latter exists. Once we have derived the coefficient vector using Equation (1.21), we see that the vector $\hat{\mathbf{g}}$ of Eq. (1.16) is the projection of $\mathbf{g}$ over the smallest subspace of V_M, which contains the vectors $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_N$. In other words, $\hat{\mathbf{g}}$ is the closest approximation to $\mathbf{g}$ in this reduced subspace.

Next, we wish to see how all the above helps in understanding the function implemented by the grid cells in the entorhinal cortex, and their relation to the function implemented by the place cells in the hippocampus. In the previous discussion, we have assumed that certain quantities are vectors and others are numbers. The method of Fourier led to the idea that continuous functions are a type of vectors, although not of the kind we have learned in our first courses on Geometry. Mathematics seeks abstraction. In the case of vector calculus, the intuitive idea of a vector is extended by considering what are its fundamental properties. In this general sense, vectors are any objects that can be multiplied by a number and can be added to form other vectors. Thus, continuous functions form a *vector space* because the sum of any number of continuous functions generates another continuous function. But, most importantly, any continuous function can be obtained from the weighted sum of other continuous functions, such as sines and cosines. Because a Fourier series – as Eq. (1.15) - has infinite independent terms, the vector space it spans has infinite dimensions. We conclude that the continuous functions are vectors in an infinite dimensional space spanned by an infinite number of "basis functions".

What happens if, instead of the infinite family of basis functions, one only considers a finite number of them? In this case all the previous discussion on vector spaces applies. With the available basis functions we use Eq. (1.20) for deriving the linear combination corresponding to a projection of a desired function over the space spanned by the basis functions. This is, in essence, one of the fundamental mechanisms to carry out function approximation by "Least Squares.

Being a projection under the metric associated with the inner product, this approximation minimizes the square distance from the desired function.

So far, we have only presented the general principles in a rather informal way. Now, we need to clarify what is that we can call an inner product of two functions. All we need is a definition that satisfies the main general requirements for the inner product operation. These are four:

1. The inner product is a real number (but see note iii).
2. The inner product is symmetric. $<\boldsymbol{\varphi},\boldsymbol{\psi}>=<\boldsymbol{\psi},\boldsymbol{\varphi}>$
3. The inner product is bi-linear: $<a_1\boldsymbol{\varphi}_1+a_2\boldsymbol{\varphi}_2,\boldsymbol{\psi}>=a_1<\boldsymbol{\varphi}_1,\boldsymbol{\psi}>+a_2<\boldsymbol{\varphi}_2,\boldsymbol{\psi}>$ and $<\boldsymbol{\varphi},a_1\boldsymbol{\psi}_1+a_2\boldsymbol{\psi}_2>=a_1<\boldsymbol{\varphi},\boldsymbol{\psi}_1>+a_2<\boldsymbol{\varphi},\boldsymbol{\psi}_2>$
4. The square *norm* of a vector is the inner product of the vector with itself: $\|\boldsymbol{\varphi}\|^2 \equiv \langle\boldsymbol{\varphi},\boldsymbol{\varphi}\rangle \geq 0$. The norm is equal to zero if and only if the vector is the null vector.

The last requirement is perhaps the most important: the inner product defines what we mean by "size". Once we have an inner product, we are endowing a space with metric properties and the space becomes a *metric space*. The integral operation[vi] offers a very simple definition of inner product. Of course, the requirement is that a function be integrable, or better, that the product of any two functions (or the square of a function) be integrable. Given two functions, $\varphi(x,y)$ and $\psi(x,y)$, both defined over a domain $D=\{x_{MIN}\leq x\leq x_{MAX},\quad y_{MIN}\leq y\leq y_{MAX}\}$, let us define their inner product as:

$$\langle\boldsymbol{\varphi},\boldsymbol{\psi}\rangle \equiv \iint_D \varphi(x,y)\cdot\psi(x,y)\,dx\,dy \qquad (1.22)$$

We can readily verify that if both functions are integrable over D, then the above definition satisfies all the four requirements. In practical calculations, the integrals are replaced by sums over the indices of the x and y variables. This is a convenient way of "extending" the natural definition of the inner product of two vectors, which is the simply the sum of the products between the corresponding components of each vector.

Let us go back to the physiology. We know that the entorhinal cortex supplies input signals to the hippocampus. Not the other way around. Based on this anatomical fact, Solstad, Moser and Einevoll (Solstad et al., 2006) asked a simple a question: is it possible to obtain the firing pattern of a hippocampal place cell from the activities of multiple entorhinal grid? The answer is affirmative and derives directly from the previous discussion. In a first approximation, following Solstad and colleagues, we model the activity of a place cell as a Gaussian function (Figure 1.10):

$$f = f_{MAX} \exp\left(-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2} \right) \tag{1.23}$$

The place cell attains the maximum firing rate, f_{MAX}, at the location (x_0, y_0) of the arena where the rat is moving and the activity decays monotonically around this point. So, there is a "receptive field" of the place cell, with a "width" of σ^2. In contrast to the place cells, the grid cells in the entorhinal cortex do not specify the single location where the rat is at a given time. Each cell is firing whenever the rat is in one of several locations, as shown in Fig. 1.7. We have already shown that the superposition of three standing sine waves would reproduce this pattern (Fig. 1.8). The function that represents this grid cell has two parameters: the wave-length and the direction. Importantly, the grid cell functions, so reconstructed, contain trigonometric functions, which are known to provide a basis for representing other continuous functions, like the Gaussian of Eq. (1.23). Thus a linear combination of functions corresponding to grid cells can approximate the function corresponding to a place cell. This is shown in Fig. 1.10 where the combination of 49 grid functions generates a pattern that approximates the typical response of a Gaussian place cell. We derived this particular example following the least-squares approach - Equation (1.21) – with the inner product metric afforded by the definition (1.22). The weighted summation of the activities form relatively few grid cells – of the order of 10 to 50 – can account for the responses of individual hippocampal place cells.

Figure 1.10. Place cells from grid cells. By a simple additive mechanism the firing patterns of multiple grid cells contribute to forming the firing pattern of a place cell (right). Each grid cell output is multiplied by a coefficient before being added to the other contributions. In this example, 49 simulated grid cells contributed to generate an activity pattern similar to a place cell. Each grid cell was obtained from the superposition of three standing sine waves, as shown in Figures 8 and 9. The simulated space is a square with 1 meter side. The spacing of the peaks in the grid cells varied between 12 and 80 centimeters and the direction of the central wave front varied within 360 degrees. The multiplicative coefficients were obtained from the approximation of a Gaussian response profile (left) with a variance of 12 centimeters.

How are the hippocampal activities updated as the rat moves around? Fig. 1.11 illustrates a simple approach. We tessellate the space with the contiguous receptive fields of place cells, following the logic of Samsonovich and McNaughton (Fig. 1.5). We are thus building a topographic chart, by associating each place cell with the location where its activity reaches a peak. This arrangement does not correspond to the actual distribution of place cells over the hippocampus. The anatomical distribution could be random (although this is disputed) and it would not matter for what one may call the "functional topography", which is the topography

determined by what is being represented. For each place cell so arranged we derive the coefficient vector b using Equation (1.21). Each element of the coefficient vector represents a "connection weight" that multiplies the input from the corresponding grid cell. All inputs are added, resulting in the net activity of the place cell. Of course, this is an oversimplified neural model to illustrate how a simple summation rule can produce a topographic map similar to that observed in the hippocampus.

Figure 11. Hippocampal GPS. In this model, each simulated place cell receives inputs from 49 grid cells, as shown in Figure 10. The space within which a fictional rat is moving is a 1 square meter region divided in 400 (20x20) small squares. The color of each small square represents the activity level of a simulated place cell, whose maximum of activity falls within that region. Thus, the place cells are distributed topographically to match the locations that they are coding for. Note that this is not intended to reproduce the spatial distribution of the cells within the hippocampus. We simply formed a chart in the style of Samsonovich and McNaughton (see Figure 5). With 400 place cells and 50 grid cells, there is a total of 20,000 connections between the simulated grid and place system. As the rat moves along the dotted line (top-right) the simulated activity on the hippocampal chart follows the pattern shown in the lower panels. The diamonds on the top-right panels correspond to the place cells with maximal activity and tend to match closely the actual position of the rat.

If we partition a region of space in 20x20 place cells and if we have 50 grid cells feeding this system of place cells, we need to form a total of 20x20x50=20,000 connections. While this is a large number of multiplications, they may be carried out simultaneously, in parallel, so that the total computational time of this whole charting operation may be as short as the time needed to carry out a single multiplication. As the fictional rat of our example moves within the environment, we see a wave of activity along a spatial map, as shown in Fig. 1.11. The peak of this wave tracks with good accuracy the instantaneous location of the rat.

This suggests how brain activities evolve between entorhinal cortex and hippocampus, as the rat moves in the environment. However, we have not yet addressed the most fundamental question: How, in the first place, does the rat's brain know where the rat is? How does the brain have an idea of the x and y coordinates that appear in the argument of the simulated grid cells? How can the brain have such basic information starting from sensory and motor data, supplied by the eyes and by the very movement instructions that the nervous system sends to the legs? We must say upfront that the answers to these questions are not yet available. So, we cannot give it here. However, in the next chapter we can outline the computational problems that the brain must solving for creating and maintaining a representation of the extra personal space.

Summary

Earlier studies on Mongolian Gerbils demonstrated the ability of these rodents to form geometrical maps of the space in which they move. These maps represent the locations and the distances of objects in the environment. The way in which the gerbils use past experiences to search for food revealed their ability to represent the Euclidean properties of space: they have a sense of distance that is invariant by rotations and translations, but not by scaling.

The ability to locate ourselves in space is closely connected to our ability to form new memories of events. The relationship between memory and space maps has a physiological substrate in the mammalian hippocampus. Evidence that the hippocampus organizes a map of space came with the first observations of "place cells" that encode single locations of extrapersonal space. A population of place cells in a rat's hippocampus forms a chart, where the instantaneous position of a rat in its environment is revealed as a moving hill of neural activity. The hippocampal place cell system is also studied in humans, where imaging studies suggest the existence of a topographical order.

Upstream from the hippocampus, cells in the entorhinal cortex appear to form a coordinate system, analogous to parallel and meridian lines on the earth's surface. Unlike place cells, the entorhinal "grid cells" become active at multiple places, disposed at the vertices of equilateral triangles over the surrounding environment. Fourier analysis can account for this pattern of activities as a superposition of three sinusoidal spatial waves along three directions, 60 degrees apart from each other. By applying Fourier analysis to a system of grid cells we obtain a family of units with a single localized peak of activity, similar to the activity of the hippocampal place cells. Therefore, a local representation of the body in space, in the form of a topographic map with an isolated peak of activity, emerges from a linear superposition of elements with broad lines of activity implementing the representation of a coordinate system.

References

Burda H, Marhold S, Westenberger T, Wiltschko R, Wiltschko W (1990) Magnetic compass orientation in the subterranean rodent Cryptomys hottentotus (Bathyergidae). Cellular and Molecular Life Sciences 46:528-530.

Dissanayake M, Newman P, Clark S, Durrant-Whyte HF, Csorba M (2001) A solution to the simultaneous localization and map building (SLAM) problem. IEEE Transactions on Robotics and Automation 17:229-241.

Fyhn M, Molden S, Witter MP, Moser EI, Moser MB (2004) Spatial representation in the entorhinal cortex. Science 305:1258.

Hassabis D, Chu C, Rees G, Weiskopf N, Molyneux PD, Maguire EA (2009) Decoding neuronal ensembles in the human hippocampus. Current biology 19:546-554.
Hull CL (1930) Simple trial-and-error learning: a study in psychological theory. Psychol Rev 37:241-256.
Maguire EA, Frackowiak RSJ, Frith CD (1997) Recalling routes around London: activation of the right hippocampus in taxi drivers. Journal of Neuroscience 17:7103.
Maguire EA, Gadian DG, Johnsrude IS, Good CD, Ashburner J, Frackowiak RSJ, Frith CD (2000) Navigation-related structural change in the hippocampi of taxi drivers. Proceedings of the National Academy of Sciences 97:4398.
Milner B (1962) Les troubles de la memoire accompagnant des lesions hippocampiques bilaterales. Physiologie de l'hippocampe:257–272.
Milner B, Corkin S, Teuber HL (1968) Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of HM. Neuropsychologia 6:215-234.
O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain research 34:171-175.
O'Keefe J, Conway D (1976) Sensory inputs to the hippocampal place units. Neuroscience Letters 3:103-104.
O'Keefe J, Nadel L (1978) The Hippocampus as a Cognitive Map. Oxford: Clarendon Press.
Samsonovich A, McNaughton BL (1997) Path integration and cognitive mapping in a continuous attractor neural network model. Journal of Neuroscience 17:5900-5920.
Solstad T, Moser EI, Einevoll GT (2006) From grid cells to place cells: a mathematical model. Hippocampus 16:1026–1031.
Sutton R, Barto A (1998) Reinforcement learning: An introduction. Cambridge, MA: MIT Press.
Thrun S, Fox D, Burgard W, Dellaert F (2001) Robust Monte Carlo localization for mobile robots. Artificial Intelligence 128:99-141.
Todorov E, Jordan MI (2002) Optimal feedback control as a theory of motor coordination. Nature Neuroscience 5:1226-1235.
Tolman EC (1948) Cognitive maps in rats and men. Psychological Review 55:189-208.
Wilson MA, McNaughton BL (1993) Dynamics of the Hippocampal Ensemble Code for Space. Science 261:1055-1058.

Index

Notes

[i] The physiological basis for the "sense of north" is not well known and varies across species. Some are capable of detecting magnetic fields and orient to them. These include migratory birds, who travel to their destination for thousands of miles, cows, who reorient themselves while grazing under electric power lines, certain bacteria, who are endowed with magnetic sensing organelles and also rodents. Some rodents also have a physiological magnetic compass, as it was demonstrated in experiments on the African mole-rats Cryptomys hottentotus (Burda et al., 1990): the presence of an artificial magnetic field deviated systematically the paths followed by the mole-rats when building their nest inside a circular arena. In addition to the earth magnetic field, there are other subtle cues that are hard to suppress in the laboratory, like odors and small variations of colors and shape of the walls. Finally, there are navigation mechanisms by which the nervous system performs what sailors call "dead reckoning", the constant integration of visuomotor information that allows one to maintain a representation of one's position with respect to a fixed frame of reference.

[ii] Fourier first presented this idea in a paper that in 1807 he submitted to Institute de France. The Institute appointed four noted mathematicians, including Laplace and LaGrange, to review the work. Unfortunately, the fourth reviewer, LaGrange, failed to see the importance of the work and objected to the idea that non-periodic functions should be represented as sum of trigonometric functions. The paper was rejected. Discouraged, Fourier turned his attention to writing a series of books titled Description of Egypt, for which he gained fame during his lifetime (remarkably, Fourier was less known as a mathematician during his lifetime and more as an Egyptologist). Only 15 years later could Fourier publish his mathematical result, and then in a book form, The Analytical Theory of Heat. Lord Kelvin, a noted British mathematician, would later refer to Fourier's book as "a mathematical poem."

iii Here, we assume this number to be real. But, in general, vector spaces can be defined over complex numbers or any kind of scalar field . Scalar fields are structures where the fundamental four operations – addition, subtractions, multiplication and division – are defined.

[iv] This statement can be demonstrated in more than one way. One is based on an insightful geometrical view of determinants. The determinant of a matrix is the signed volume of the parallelepiped included between the vectors that constitute the columns of the matrix. To see this, start with the simple case of a diagonal 3x3 matrix. Each column is a vector along the corresponding axis. The product of these vectors is the volume of the rectangular parallelepiped with three edges formed by the three vectors. This argument can be rapidly extended to more complex matrices with more rows and columns.

Each column of the matrix Φ is the representation of each basis vector in the frame established by the vectors themselves. The fact that the vectors of a base are linearly independent implies that span the full volume of their own space. Therefore the determinant of Φ cannot vanish.

[v] We are still considering only real-valued matrices. To obtain the definition for complex-valued matrices, simply replace T (for transposed) with an asterisk (for complex conjugate).

[vi] There are several different types of integrals. The one that is most often used in function spaces is the Lebesgue integral, after Henri Lebesgue, another French mathematician. Another type of integral operation is the Riemann integral, which is the one most commonly introduced in calculus classes. The distinction between Riemann and Lebesgue integrals is important but subtle, and beyond the scope of this text. In most practical cases the two methods give the same result.