SISSA
Scuola
Internazionale
Superiore di
Studi Avanzati

Neuroscience Area - PhD course in

Cognitive Neuroscience

Fragmentary Understanding of Memory

Candidate:

Oleksandra Soldatkina

Advisor:

Alessandro Treves

Academic Year 2020-21

MA PER SEGUIR VIRTUTE E CANOSCENZA • SISSA •

Declaration

Some ideas and figures have appeared previously in the following publications:

Articles

K. Ryom*, V. Boboeva*, *O.Soldatkina*, A. Treves, “Latching dynamics as a basis for short-term recall”. In: PLOS Computational Biology, 2021.

S. Andreetta, *O. Soldatkina*, V. Boboeva, A. Treves, ”In Poetry, if Meter has to Help Memory, it Takes its Time”, in Open Research Europe, 2021.

Book chapter:

*O. Soldatkina**, F. Schönsberg* and A. Treves, “Challenges for place and grid cell models”, in “Computational Neuroscience Approaches to Cells and Circuits”, M. Giugliano et al, eds. Springer, 2021 (in press).

Acknowledgements

I want to thank the many people who supported me on this rocky path. First of all, I thank Alessandro Treves for careful supervision, for patient guidance, for all the wisdom shared with humor.

I would like to thank Mikhail Katkov and Misha Tsodyks for generous help, sage advice and fulfilling discussions during my stay at Weizmann Institute.

I am greatly thankful to all the limboers. Thanks to Sara Andreetta, Elisa Ciaramelli and Kwang Il Ryom for the enriching collaborations and all the ideas that have blossomed within. Thanks to Yifan Luo, Davide Spalla, Vizhe Boboeva, Massimilliano Trippa, Zeynep Kaya, Serena Di Santo, Judit Fiedler, Aline Viol and others who I have deeply enjoyed sharing the Limbo experience with.

I separately thank Francesca Schönsberg for lifting my spirit when most needed, for her open mind and humorous advice. I'd like to express my gratitude to Anna Tonon Appiani, for her smart counsel, for, in her words, relaxing my relax and leisuring my leisure, for all her kind care and for showing me the best of Italy. I'd like to thank Filip Agatic for the witty comments and sunny hikes that kept me going. Many thanks to former and present CNSers, an ensemble of beautiful people who taught me so much, for all the stimulating discussions and the really fun times.

Finally, I would like to thank my family and friends. My parents, for patient support and sharp advice. My brother, Ivan, for listening and for ingenious suggestions that led to at least two breakthroughs in this work. Eugenia, for her encouragement and unbeatable jokes. Ksenia, for sisterhood. Denys, for keeping my spirit cozy warm while pushing me to cycle this rocky road uphill for the better views.

Abstract

My thesis argues that memory resembles navigation by fragments, as proposed in [1]. To relate to it, imagine you are in Paris and you already know quite well the neighbourhood of where you are staying. Now, it is easy to get to the Eiffel Tower: you may need some help with an overall direction, but once you follow it, it does not really matter how exactly, very soon you will see the Tower and you will find your way to it. When you get closer you will remember the immediate surroundings of the Tower and how to find the nearest coffee place. Similarly from there you can get to Notre Dame: now you may just follow the river, it is one-way and you may only remember just a couple of spots along the way, but near the Cathedral you recognize every pigeon. Here the remembered fragments are the two landmarks, their neighbourhood and the river, to some extent. Keep this in mind.

I prepared a guided tour for you, where the fragments I learned to navigate through are memory phenomena that can be studied from a point of view of navigation by fragments. Pick a sustainable vehicle, I suggest a sailing boat as it must have the right speed, or you may imagine this trip as cycling uphill (as I certainly felt it), and follow my thoughts.

We start with the hippocampus and its relation to memory and navigation. We discuss how the discovery of spatially selective cells there led to study memory systems in our brain as attractor neural networks, what fragmentary knowledge about the representation of space could be learnt from the hippocampus and the open questions that remain.

Just across a bridge, in Chapter 2, we will attempt to answer some of those

questions, studying a mathematical model of an attractor neural network in CA3. We will expand our knowledge about the quasi-continuous maps our model forms for multiple sample environments, their storage and their usage. We will argue that CA3 network storage can in fact be thought of as a fragment assembly.

We must take a series of one-way turns to reach some understanding of how human recall relates to virtual rat navigation, but I will be your guide. In chapters 3 and 4 we will discuss, using free recall as a model example, how human memory, too, can be thought of as navigation by fragments. We will present a series of experiments and simulations of a Potts network that together point at the semi-random nature of human recall. We suggest that when given a plain environment to learn – a hexagonal grid on a screen, as an empty box for a rat in a lab, human participants tend to memorize locations on the screen by 'seeing' there various familiar fragments. And the more restrictive the memorization task is, the least they can reach these attractive patterns. We will discuss how common biases and unanimity across participants in fragment activation can be predictive of human recall capacity.

Finally, we will briefly visit Milan of Mind Wandering and Rome of Remembering Poetry. We will argue that, despite being seemingly (and luckily) far from each other, both of these places-processes have in common their functional reliance on fragmentary schemata. First, we will propose an experiment that aims at quantifying the effect of recently acquired episodic schemata on mind wandering in participants with a lesion to vmPFC and in their healthy controls. Separately, we will suggest a mechanism of selective involvement of poetic meter variables as schemata helping remember non-words in non-poems.

In the end we will gather to review the pictures from the journey and discuss the takeaways. Let's go!

Contents

Chapter 1

Introduction

1.1 Spatially selective cells

First to have been discovered of spatially selective cells, place cells are operationally defined by the characteristic firing behavior these neurons show when the animal explores a typical laboratory environment, usually a one-dimensional path of various shapes or a two-dimensional, flat, empty recording box. They can be found in different areas of the hippocampal division, a region of the brain identified to be crucial not only for spatial cognition but also for episodic memory. Their discovery, in the early 70's for place cells [2] and in 2005 for grid cells [3] led to the Nobel Prize in 2014. Let us take first a brief look at the anatomy and at the firing properties of these cells.

1.1.1 Main anatomical traits

Place and grid cells have been discovered in the hippocampal system, a brain region situated in the medial temporal lobe. The hippocampal system can be subdivided in several areas, and first in two main regions, the hippocampal formation and the parahippocampal region [4], which can be differentiated by their gross cytoarchitectonic organization. The hippocampal system is highly similar in different mammalian species; here we give a short overview focusing on rodents.

The hippocampal formation – with place cells: The hippocampus proper, or cornu ammonis (CA) has pyramidal principal cells in one layer -– a cortical structure called allocortex – and is further subdivided in a sequence of three areas, CA1, CA2 and CA3, with remarkably distinct connectivity between them. It is flanked on the input end by the dentate gyrus, or DG, which evolves out of the same type of cortex but with small granule cells instead of pyramidal cells, and on the output end by the subicular complex, which, in as many as 5 internal subdivisions [5], links the hippocampus to the adjacent multi–layer cortex. Place fields have been found throughout the hippocampal formation and have been studied especially in CA1 and CA3. For a long time, in fact, it was puzzling how place cells in the two subfields looked so similar, apart from minor statistical differences, when, instead the circuitry is so different: CA3 is dominated by recurrent connections (RC), unlike CA1, and the main afferent connections to CA3 are from the DG granule cells and from Entorhinal Cortex layer II (these connections are referred to as perforant path or PP), unlike those to CA1 which are from EC layer III and from CA3 itself (see Fig. 1.1).

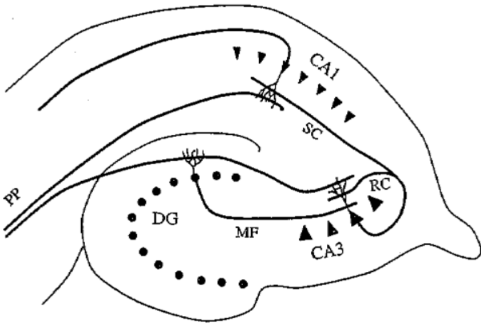


Figure 1.1: Schematic representation of the connectivity between three main regions of the hippocampus: DG, CA3 and CA1.

The parahippocampal region – also with grid cells: The parahippocampal region is characterized in part as periallocortex, to emphasize its transitional nature to fully neocortical structure with multiple layers of principal cells. It is formed by the entorhinal, perirhinal, postrhinal cortices and by the components of the subicular complex, that some prefer to view separately from the subiculum proper. The medial subdivision of the entorhinal cortex (mEC) has risen to particular prominence after the discovery of grid cells, somewhat obscuring the fact that most of its principal cells do not conform to the grid cell stereotype even in standard laboratory settings, nor do those of the other parahippocampal areas. At the system level, perirhinal cortex makes afferent connections to lateral EC that do not appear to convey fine spatial information, unlike the connections from postrhinal cortex to mEC. Grid cells emerge, in this perspective, as one form of refinement of spatial information before it is merged with nonspatial information in the hippocampus, where both lEC and mEC project, and largely transformed into a place cell code, at least in rodents.

The entorhino-hippocampal circuitry: Principal cells from EC layer II reach DG and CA3, while principal cells from EC layer III reach CA1. Internally in the hippocampus, activation propagates in a sort of one-directional loop, with recurrence (in CA3) and shortcuts. DG granule cells project their so-called mossy fibers to CA3, where they make sparse but powerful synapses on the apical dendrites close to the cell body of CA3 pyramidal cells. Since the same CA3 cells receive many more (but weaker) synapses on their distal apical dendrites from the same fibers originating in EC layer II that, *en passant*, connect to the granule cells, a major riddle has been to understand this apparent duplication of the information arriving to CA3, directly and, as it were, translated by the DG. A more recent question involves CA2, which had long been regarded merely as a small transition region between CA3 and CA1; recent evidence on a potentially important role in social cognition [6] has been accompanied by the observation of CA3-like anatomical features in CA2, such as prominent recurrent collaterals [7] and the formation, perhaps in pathological conditions, of mossy synapses [8]. Feedforward connections from CA3 to CA1 (the Schaffer collaterals) and from CA1 to subiculum are also intriguingly combined, in what may be called a heteroassociative architecture, with EC layer III inputs to these two regions. Fibers then project back from CA1 and subiculum to EC layers

V and VI.

1.1.2 Single cell selectivity

When considering spatial cognition, cells in the hippocampal division have been first characterized, mainly in rodents, through their individual selectivity, by looking at the *firing rate map* of each cell. In such a map, the spike events are plotted in a drawing representing the environment in which the animal is moving, at the position of the head of the animal when each spike occurred. Spikes clustered in a specific region form a *field*. The number of spikes occurring in each spatial bin is typically divided by the time spent in that bin, and the map is then regularized to look smoother. A common trait to the various types of cells listed below is that the localization of their fields appears unrelated to the position of each cell in the tissue, and neighboring cells do not necessarily show overlapping firing fields in the environment. The main types are:

- **Place cells**: Originally discovered half a century ago [2], their activity is peaked at one or a few positions in space in the typical environments in which rodents are made to run in the laboratory. In one-dimensional environments, such as circular paths, n-arm mazes or linear tracks, place fields seem to be directional, i.e. there occurs remapping of place cell activity when the animal is running in one direction with respect to the other direction [9] (although on a ring there seems to be only rate remapping [10]). On two-dimensional environments they tend to be, or to become, non-directional. Place cells have been most extensively described in CA3 and in CA1, where it is estimated that between a quarter and a half of all pyramidal cells show at least one place field in a typical 1 m^2 box. Place activity, like other types of selective spiking, is typically modulated by the speed of the animal.

- **Head Direction cells**: First reported in 1984 [11], HD cell activity depends on the direction of the head of the animal, which on average tends to coincide with, but is quite distinct from, its direction of motion. They are found in a variety of areas, especially in the parasubiculum and in the EC.

- **Grid cells**: A startling discovery [3], their activity is peaked, ideally, at the vertices of a hexagonal lattice, spacing from a few tens of centimiters upwards, giving rise, in a typical two-dimensional box, to several grid fields per cell. The spacing and orientation of the lattice appear to be shared by neighboring cells, but not the position of their fields. Whereas in EC layer II grid activity is characterized as a-directional (but see [12]), in deeper layers of EC the activity of most grid cells is modulated by head direction, and they are called conjunctive (grid) cells [13]. The spacing of the grid lattice increases towards the ventral portion of mEC [14] in what appear to be discrete steps, or *modules*. Grid cells are predominantly found in mEC but are also present in the pre/para-subiculum.

- **Border cells**: Described by [15], the activity of border cells is intense at one or several borders of the environment the animal is exploring. They are found in the EC and pre/para-subiculum.

- **Speed cells**: Originally found in [16], their firing rates linearly depend on the velocity at which the animal is navigating. They have been found in the EC, but variants sensitive to angular velocity have been recently reported, also in the pre/para-subiculum.

- **Object, object-trace, object-vector, social cells**: A still burgeoning variety of selectivity types is observed when objects (or other animals[17]) are introduced in the same environment, starting with those observed by [18], which fire selectively at positions related to an object and which were found in the lateral enthorinal cortex.

One should note that the selectivity of cells in the parahippocampal region tends to be stable, presumably due to the mixture of inputs they receive and the network they are embedded in. Instead, cell selectivity in the hippocampal formation is thought to be determined by the context: the same pyramidal cells may show two place fields in one box, none in another, and be selective for an odor in an olfactory discrimination task [19].

1.2 Place cell memory models

By 1971, at the time place cells in the rat hippocampus were discovered, two other milestones had been reached: half a century of investigations by many laboratories on synaptic plasticity in the mammalian brain had just begun, with the discovery of long-term potentiation (LTP) in the rabbit hippocampus (preliminary findings from 1966, reported in [20], see [21]); and the solitary daring enterprise of a young student, David Marr, had been concluded with the publication of his *theory of archicortex*, i.e., of the hippocampus [22]. While the work on LTP had the potential, later expressed, to bridge the other two, the theoretical model developed by David Marr seemed at first to have nothing to do with place cells, and viceversa.

1.2.1 Integrating place cells within memory representations

Marr's vision is of a memory system, a *simple* memory, as he calls it in contrast with the theory he had developed earlier for neocortex [23]; in his memory system, representations are disembodied, abstract entities, to which neurons, or simple binary units, are recruited as required by the contingency. The initial description of place cells, instead, appeared to reveal that what they encode is very much concrete, a specific position of the animal, with the same dedication and reliability with which primary visual cortex cells would encode the presence of light in certain regions of their receptive field. Integrating the two approaches has required gradually broadening both perspectives, so as to ground Marr's and to lift up O'Keefe's.

A theoretical perspective on how the hippocampus forms and stores memory

The human hippocampus had already been associated with episodic memory. Most of the observations on hippocampal *intellectual* function had come in fact from studies in brain-lesioned patients, the most famous ones with patient HM [24]. Following a bilateral hippocampal lobectomy in adult age, he had lost most memories about his life, extending several years before the operation, and he was not able to form new ones, but he had preserved cognitive capacities, relatively spared remote memories, and remembered who he was – though not his present circumstances, where he was and why he was there [25]. It was from the thorough study of HM that Brenda Milner proposed that the fundamental role of hippocampus is in the formation of episodic memories.

With his 1971 paper [22], David Marr developed a detailed neural network theory for this function, bridging with a mechanistic model the observations in patients and the neuroanatomy of the mammalian hippocampus.

In the very middle of the brain, as it were, the hippocampus gets inputs, direct or indirect, from all the sensory areas, and "binds" them in a way that later, when cued with partial information, say a visual signal, the hippocampus integrates all the elements related to that memory – and we are able to "relive" a whole episode, for example a birthday dinner two months ago. So, at least, the mainstream narrative goes.

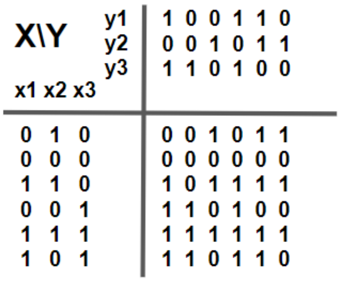
Marr's theory was a grandiose attempt to structure such narrative into a well-defined mathematical model, aiming to understand the anatomical structure of the hippocampus based on the memory impairment described in patients with hippocampal damage. This general logic is clear, and it has been profoundly inspirational for later work by many researchers. The implementation, however, is rather complicated, often becoming obscure, perhaps to Marr himself, and definitely hampered by the lack of adequate mathematics – it will be contributed by physicists over 10 years later – and of adequate numerics, which forced Marr to continuously zig zag between logic and quantitation, relying solely on his powerful intuition.

Marr's work did not consider the place cells, that were being discovered at the same time by O'Keefe and Dostrovsky in rodents. The discovery would stimulate a computational hypothesis in a different direction: that the location of the animal in space is computed within the hippocampus, and therefore its internal circuitry has to be understood as functional to self-localization, and hence in general to navigation, rather than to memory.

First computational theory uniting space and memory

16 years after Marr and O'Keefe with Dostrovsky, McNaughton and Morris in a review paper [26] set out to recombine the two hippocampal narratives – the memory function and the spatial function: they suggested that the hippocampal circuitry *stores* spatial representations within its synapses. Although they obviously cite the discovery of place cells and the book that framed it into a conceptual theory [27], the emphasis of the review is on the mechanics of learning. For that, McNaughton and Morris suggest a set of simple network models, all based on the Hebb [28] idea that "neurons that fire together, wire together" – associative memory at the synaptic level, which envisages that a pair of neurons with conjunctive activity develop a stronger synapse between them. The spatial character of the information presumed to be the bread and butter of the hippocampus does not really inform the network models, all constructed with binary units and binary synaptic weights, that are difficult to relate to continuous space; but the different networks are brought into tantalizing correspondence with different parts of the hippocampus.

At the core of each network there is a matrix of associatively modifiable weights, which is taken to capture, in binary form, and through cumulative learning, the occurrence of conjunctive activity between input patterns on two streams X and Y, as represented in Fig.1.2: if two patterns on X and Y are paired together in which neurons j and k are both active, the associative matrix is taken to *learn* the pairing by activating the corresponding weight (to a standard "1" value, which cannot be raised further).



X\Y		y1	1	0	0	1	1	0
		y2	0	0	1	0	1	1
		y3	1	1	0	1	0	0
x1	x2	x3						
0	1	0	0	0	1	0	1	1
0	0	0	0	0	0	0	0	0
1	1	0	1	0	1	1	1	1
0	0	1	1	1	0	1	0	0
1	1	1	1	1	1	1	1	1
1	0	1	1	1	0	1	1	0

Figure 1.2: Example of a correlation matrix of converging input patterns X and Y

The memory mechanism initially proposed by Marr in these terms looks like schema B from Fig. 1.3: the two converging streams of information are *paired* in an asymmetric fashion: one (above in the schema) determines the original activation pattern on the receiving units, and needs not modify its synaptic weights onto them; the other (below, in the schema) modifies them when paired with the first, and as a result comes to reactivate a very similar pattern, alone, acting as a learned *cue*, to retrieve what may hence be called a *memorized* representation – in the simplest model, a binary one. But McNaughton and Morris point out that usually, particularly for the spatial context, any part of it may act as the cue, that is we may recall the whole scene starting from any arbitrary element it contains, as long as it identifies the scene among those concurrently in storage – the stable division between primary and modifiable inputs of schema B has no meaning, in this case, and one reverts to the undifferentiated form of schema A, called *auto-association.*

Finally, a mechanism like that exemplified in schema C associates the input X to the system's own output – this type of auto-association could serve to store sequences of scenes, as in episodic memory, that is, cued with the first memorized pattern (or a fraction of it serving as a suitable cue) it can recall the whole sequence of consecutively stored patterns. In addition, with its recurrence it can keep the output units activated longer than the afferent inputs, thereby realizing a simple form of short-term memory.

These simple schemata appear intriguingly related to elements of hippocampal circuitry, although McNaughton and Morris are quick to point out that the correspondence should not be interpreted too rigidly. Thus, the cortical inputs to DG include the medial perforant path, conveying spatial information through supposedly stronger synapses, and the lateral perforant path, which can be paired with it, enabling its object-related information to act as a cue to elicit the same downstream pattern that had original been activated by the whole spatial scene, a bit like in the schema of Fig. 1.3B. At the same time, they argue that 3% of the synapses on the medial pathway are 10-20 times stronger than the others, thus acting as *detonators* that impress a representation on the receiving units, which can be later reactivated by any arbitrary subset, as in the schema of Fig. 1.3A. The highly recurrent collateral network in CA3, already noted by David Marr, resembles the scheme in Fig. 1.3C, and there the detonator synapses could be those on the mossy fibers, the axons of the granule cells which, McNaughton and Morris note, provide a transform of the same cortical inputs arriving also directly onto the apical dendrites of CA3 pyramidal cells. In this way, not only CA3 can serve for pattern completion, it could also store sequences of patterns. A cue coming from the dentate gyrus would help recall the whole sequence.

The simple schemata may help interpret also components of hippocampal anatomy not explicitly highlighted in [26]: the convergence of distinct input pathways onto the same cells, for example, is particularly prominent in CA1 (Fig. 1.1), where it is the Schaffer collaterals from CA3 and the cortical layer III inputs that could entertain the asymmetric relationship depicted in Fig. 1.3B.

Attractor neural networks help handle spatial information

The recurrent connectivity of the auto-associative model in Fig. 1.3C implies that the neurons, serving as inputs and output to the same synaptic matrix, will tend to reach a stable configuration, or *pattern*, if they can find one in which the activation of each neuron is consistent with that of the neurons that feed its inputs. This consistency is of the same nature as that describing the relaxation dynamics of

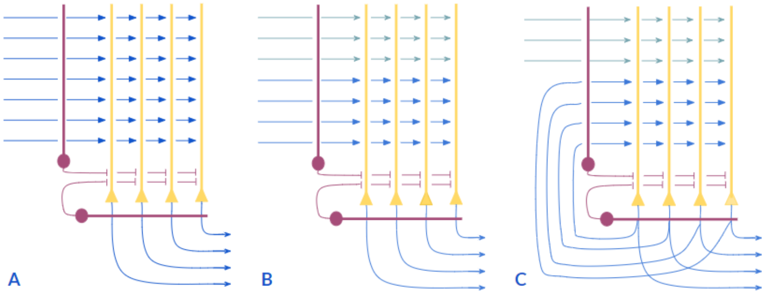


Figure 1.3: Adapted from [26]: simple schemata of neuronal networks embodying variations of associative memory mechanisms. The red circles represent feedforward and feedback inhibitory cells, whose control of the yellow pyramidal cells is discussed both by [22] and by [26].

dissipative physical systems of interacting variables, as envisaged by John Hopfield in his seminal paper on content addressable memories [29]. Relaxation to a steady state, though subject to additional constraints in physical systems, such as symmetric interactions, can be regarded as a mechanistic paradigm for the cue-driven reactivation of a memory pattern, in which the memory is selected on the basis of the partial content represented by the cue. Distinct steady states can be accessed by the different ensembles of cues that they *attract*, and typically it takes a very short time for the relaxing activity pattern to become very similar to its attract, as the first few steps, as it were, are much larger than the later ones. Amit, Gutfreund and Sompolinsky showed how the attractors of such dynamics can be studied with a beautiful nontrivial mathematical formalism derived from the statistical physics of disordered systems [30]. Applying the formalism, however, and even simply conceptualizing attractor dynamics, is less straightforward when dealing with the representation of spatial, continuous variables.

Let us take therefore a step back, and first look at attractor dynamics in another special type of spatially selective cells, to understand basic aspects of attractor dynamics in the representation of space.

Cells sensitive to the absolute head direction of the animal were discovered in different parts of the brain: in the subiculum, thalamus, retrosplenial neocortex, dorsal striatum, etc. [31]. In these studies, head direction (HD) is calculated, typically from two diodes attached to the head of the animal, independently of its spatial location or the relative position of the head to the body. With HD cells in each of these regions, the striking finding is that the direction that most activates a cell remains the same in every environment, familiar or new. In fact, this is striking because often the information on the basis of which the animal can calculate its head direction is partially misleading, e.g., when an object has been moved. Therefore, although all the information might be concurrently available, it has to be interpreted, and perhaps in part discarded. Further, when most of it is not coming through the senses, for example because lights are turned off, and olfactory cues have been washed, HD can be reconstructed from memory, if a system exists that keeps it in memory.

This system can be an attractor network, and in fact such an observation has motivated the development of a simplified version of the theory of continuous attractor neural networks. In 1995, Skaggs et al [32] proposed that a *ring attractor* could interpret sensory cues and keep HD in active (short-term) memory. To understand it intuitively, imagine: one places head direction cells on a ring, each at the angle it is most responsive to (see Fig. 1.4), and the connections between the neurons are taken to have been strengthened by Hebbian plasticity, resulting in neurons close to each other on the imaginary ring exciting each other. What we can observe then, is a bump of activity or an "activity pocket" – it corresponds to the animal's head direction, wherever it is pointing, among the 2π directions on the ring. Fig. 1.4 illustrates a somewhat more sophisticated version of this concept, in which there are 3 rings, not one, and slightly asymmetric connections between the rings are used to update the angular position of the bump with velocity inputs. What remains true also in the sophisticated version, however, is that the interactions among the units – producing attractor dynamics – compactify, stabilize and can keep in short-term memory a position on the ring, but not select among alternative rings.

Could this system also include the selection of one among a number of rings? The question becomes very concrete, and easy to visualize, if applied to place cells,

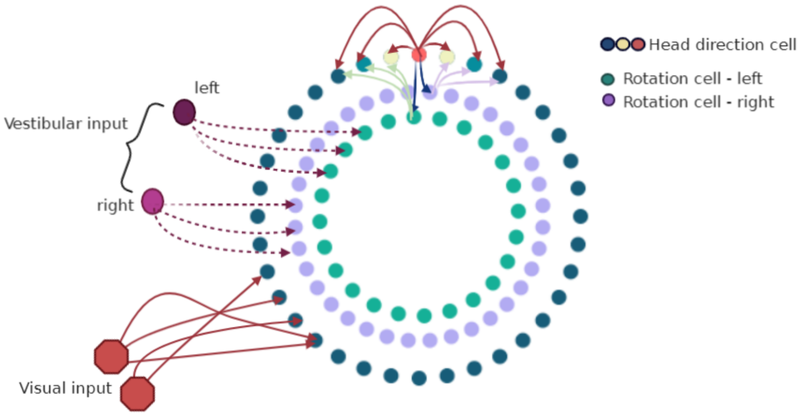


Figure 1.4: Head direction cell ring (adapted from [32], to which we refer for an explanation of the proposed mechanism.).

in 2 dimensions.

Remapping: a continuous attractor for each familiar environment

A fundamental discovery was reported the same year as the McNaughton and Morris review [33, 34], when it was found that place cells *remap* their activity from one spatial context to another: they change their firing patterns when the animal is moved to a different environment, in a manner that appears totally unpredictable from knowledge of its place field(s) in the original environment, or from the changes, or remapping, expressed by nearby cells.

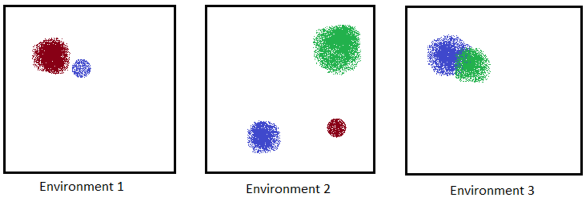


Figure 1.5: Remapping illustration: the place fields of three place cells (marked by different colors) as they may appear in three different environments.

As schematically and summarily illustrated in Fig. 1.5, in detail these experimental findings show that:

- place cells tend to form a new, seemingly entirely reshuffled configuration of activity for each new environment the animal is exposed to, unless it is identified with a previously familiar environment;

- a place cell may have one or more fields of activity in some environments, and remain silent in others;
- relations between place fields, whether expressed by the same or different cells, are not preserved by remapping;
- the switch between two representations is very abrupt, although in special conditions, when the animal is confused, remapping might teeter back and forth for a few seconds [35].

This last fact led to a strong hypothesis that the configuration of place cells activity, in a given environment, should serve as a continuous attractor for the network, i.e., comprise a manifold of all the spatial positions in that environment. Attractor dynamics would then unfold at two different levels. Within one environment, it would refine or interpret possibly conflicting sensory evidence, and keep in short-term memory the continuously varying position of the animal in the environment, much as with the ring model for HDs. When moving the animal to a new environment, attractor dynamics would again refine and interpret possibly conflicting evidence, but this time for selecting the representation of the environment among a set, possibly discrete, kept in long-term memory. Unlike the continuous updating of spatial position, this will then result in an abrupt "jump" from one attractor to another, that could be observed in the activity of single cells or of small groups, as in Fig. 1.6. The distinction between the two levels of operation, continuous and discrete, is clearly an oversimplification, which may well prove inadequate when moving outside artificial environments defined in the lab.

In 1997, McNaughton and Samsonovich [37] proposed a network model that accounts for the expression of multiple continuous attractors in 2-dimensional space, which they called *charts*. A chart can be conceived as an arrangement of place cells on an imaginary plane in such a way that each cell is represented at the location of

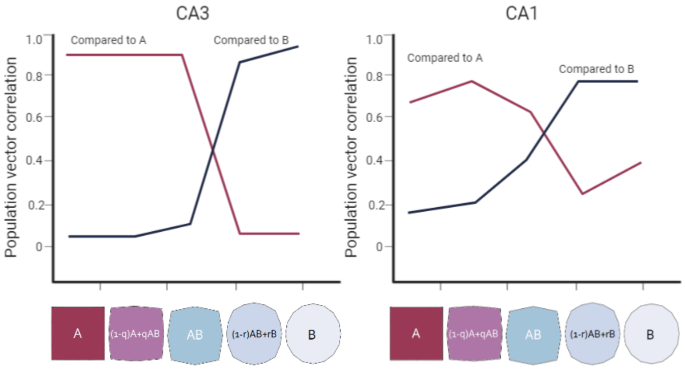


Figure 1.6: Adapted from [36]: Population vector correlation of firing rate maps of place cells in CA3 and in CA1, after making rats familiar with boxes A and B, and then testing them in *morphed environments* spanning a quasi-continuum in between. Population activity undergoes a sharp transition, especially in CA3, suggestive of attractor dynamics between the two discrete long-term maps of A and B.

its highest activity (just like the placement of head direction cells on a ring), and then the actual spatial position within a chart is represented as a bump of activity moving along a continuous attractor, with the motion suggested to be registered by path integration.

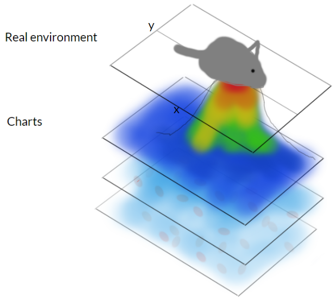


Figure 1.7: Adapted from [37]. Conceptualization of a chart: a place cell configuration on a plane, where each cell is placed in the location of its highest activity and the actual location of the animal is represented by a“bump” of activity.

Remapping between charts should let the system update its estimate of the position of the animal, as the latter navigates among familiar environments, but attractor dynamics can only be effective, if the system can hold in long-term memory a sufficient number of environments – therefore the chart model makes sense only if the storage capacity of the continuous attractor model, applied to the CA3 systems, turns out to be substantial.

A simple mathematical model of a recurrent network of threshold-linear units was analyzed by [38], in which a unit is assigned one field or none, in each chart, with an overall sparsity a of population activity (roughly, a is the fraction of significantly active units at any one time). The model allows calculating the maximum number p_c of charts that can be stored and individually retrieved, which scales up with the number C of distinct recurrent connections each unit receives, so that the result is expressed as usual with the ratio $\alpha_{max} = p_c/C$. α_{max} was found to depend mainly on a and on the degree of recurrence of the connectivity; interestingly, for a fully or densely connected networks α_{max} is seen to have a maximum for intermediate values of sparsity [38]. Fig. 1.7 shows the result for one-dimensional charts, but for two-dimensional ones the outcome of the calculation is similar, if quantitatively lower. The indication, therefore, is that a densely recurrent network with of order 10^4 connections per neuron can store up to roughly a hundred charts [39]. This provides one form of quantitative convergence between the two hippocampal narratives – episodic memory and spatial cognition.

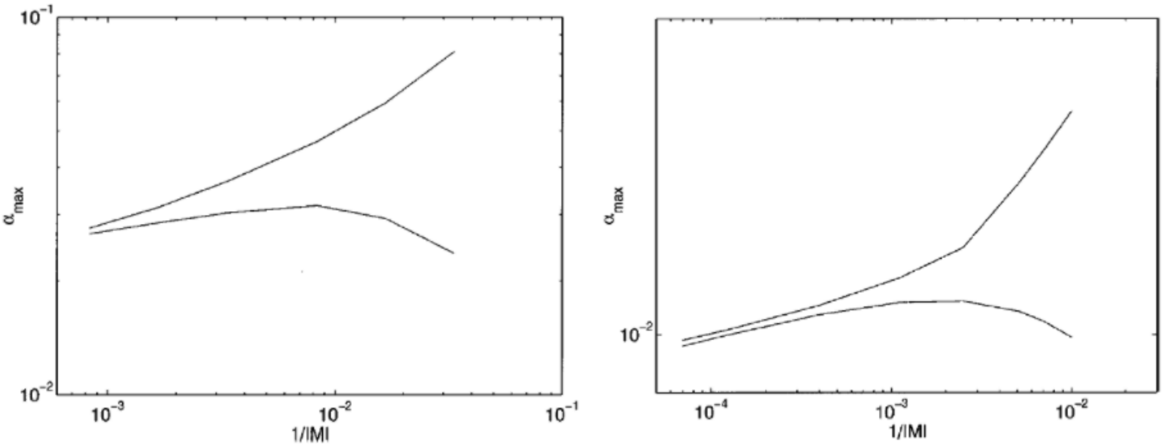


Figure 1.8: Memory capacity of the network for one-dimensional (left) and two-dimensional (right) charts as a function of chart sparsity, in the fully connected (lower curve) and extremely diluted (upper curve) limits [38].

Note that these calculations assume uncorrelated charts, as perhaps enabled by the dentate gyrus inputs to CA3. With correlations, the storage capacity but even more the remapping dynamics would be different, as already indicated by beautiful analytical work [40].

1.2.2 How can place fields be set up?

Theories of memory should conceivably devote at least as much attention to the issue of how memories are created, as to how they can be retrieved. Yet, the two dominant memory modelling narratives of the 80's shirked their responsibility towards memory formation in the brain, for different reasons. For networks trained with backpropagation [41], the artificial learning algorithm was an embarrassment and its plausibility was best glossed over, focusing instead on the ability of such networks to implement whatever mapping was requested, as if by an outside agent. For auto-associative networks analyzed from a statistical physics perspective [42], instead, the characteristics of the stored representations are a given, a bit like the constituents of a piece of condensed matter, and the focus is on analyzing their dynamical rearrangement – the retrieval process – not how they came to be in the first place. It was again the McNaughton and Morris review that raised the issue of how to set up spatial representations for memory.

Non-associative inputs in associative memory

As already mentioned above, the key idea was that of synapses acting as "detonators", a notion borrowed from the study of the neuromuscular junction [43]. It was proposed in [26] that a small subset of the synapses on the medial entorhinal input to DG act as detonators, those presumed to be much stronger than the rest. They would then essentially establish the primary selectivity of the receiving cells, with the remaining numerous mEC and lEC inputs relaying additional attributes that can be paired, through associative learning, with such primary selectivity. They also proposed that the very granule cells of the dentate gyrus, with their sparse and powerful mossy synapses, serve as "detonator cells" for the CA3 network. Then they could establish a place field, for example, to which performant path inputs to the apical dendrites could associate spatial context (from mEC) and object (lEC) information.

The idea that memory representations in CA3 are primarily established by DG inputs was cast in semi-mathematical form by Treves and Rolls [44], with a simple model that suggests that the perforant path, on its own, would not have the strength to prevail over the interference produced by already stored memories, reverberating in CA3 mainly on the recurrent collaterals; whereas the mossy fiber inputs have the appropriate quantitative characteristics to imbue new memories with sufficient information content. A logical inference from this model is that, once the new memory representation has been formed, removing the granule cells or just blocking their afferents to CA3 should not impair the retrieval of information already deposited there. Such a prediction was confirmed in two different experiments, in mice and rats, employing different manipulations to either transiently or permanently remove DG inputs to CA3 [45, 46].

The intuition that emerges from these findings is that for the CA3 network to be able to store multiple charts the input from DG *has* to be strong and sparse. And yes, there has to be additional input into the recurrent network in order to cue the retrieval of the memories – this could be the perforant path from entorhinal cortex [46]. Also in 2004, Leutgeb et al. in fact pointed out that the place fields of CA3 and CA1 cells, hitherto so strikingly similar, presented one major contrast in a suitable experimental paradigm: the former remap to orthogonal representations when changing environment, the latter show graded changes, that reflect the physical similarity of the two environments [47]. Note, however, that the mathematical model in [44] does not refer to place fields at all, and is framed, in fact, in terms of discrete patterns of activity. To gain insight into the formation of place fields, an even simpler computer model had been proposed a year earlier.

Associative model for DG place fields

Patricia Sharp [48] proposed an associative model, which can be taken to account for the formation of non-directional place fields wherever they appear first, in information flow, e.g., in the dentate gyrus. Imagine combining together sensory information from all possible orientations, i.e., all possible directions in a two-dimensional environment, which the animal can follow to traverse a particular location. A representation of the specific location, independent of direction, can be established by a straightforward variant of Hebbian learning (a *trace* learning rule) within a competitive associative network. The mechanism exploits the continuity of space: different viewpoints of the same environment from nearby directions can be smoothly associated together, as the animal changes its head direction. The resulting simulated place fields can be seen on Fig. 1.9: they resemble very non-noisy place cell signals. This could therefore be a mechanism to form DG place fields, although it might have to be extended to incorporate more than just visual information. Note that, in the models we discuss later, we take cellular selectivity to be already in the form of place fields, arguably inherited from those set up in the DG.

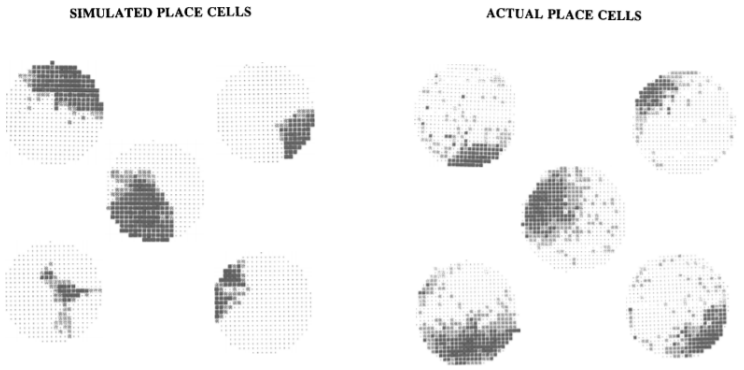


Figure 1.9: Examples of simulated "place cells" and real place cells firing maps corresponding to a floor of a cylinder rat cage, where the firing rates are binned for computational purposes (modified from [48]).

The Sharp model describes one mechanism for the formation of place fields from inputs of a different nature, a simple mechanism that may be selected for even by brain-less genetic algorithms [49], and that generalizes directly, eg., to primates, from the learning of arbitrary association [50] to the establishment of spatial view fields [51].There is no real need for such a mechanism if place cells are taken to emerge from the place fields of other cells; if anything, the computational problems that a model may try to explain are different. For example, if place fields are assumed to emerge from the summation of the place fields of many different grid cells of different phases and orientation, the challenge may be how the former dispose of the periodicity of the latter [52] (but see [53, 54]).

CA3 fields from DG fields

In the same logic, if CA3 place fields arise from those in DG, the question is not so much how they arise *ex nihilo*, but rather whether they can be sufficiently defined by the DG inputs to overcome the interference due to other memories, including other spatial charts, previously stored in the CA3 network. This question was addressed in [55, 56], with a study of an attractor neural network of CA3, in which DG inputs are in the form of spatial maps (Fig. 1.10).

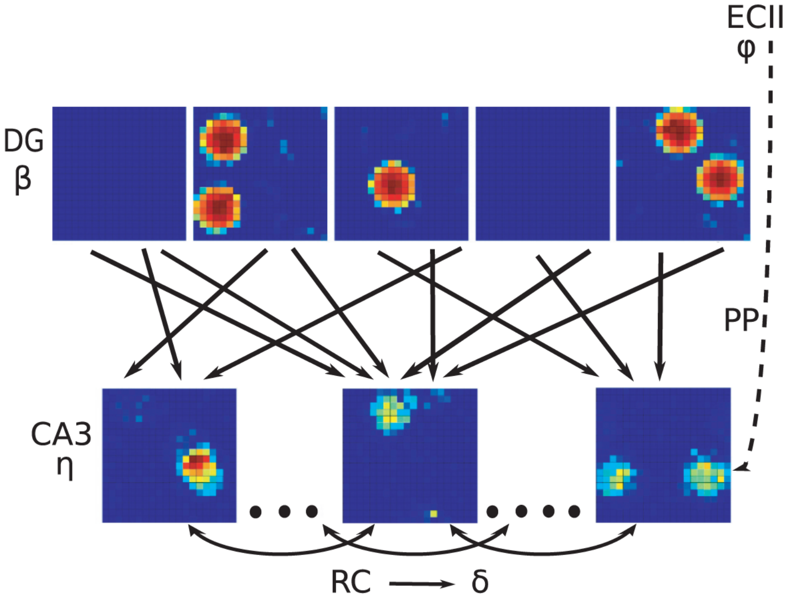


Figure 1.10: Schematic representation of the model wiring in [55].

The model assumes that DG granule cells encode position in a room of size $l \times l$, by a fraction of them having assigned, independently of each other, one or a few place fields each, but with most of them being silent. Their activity, denoted as β_i, is fed into a recurrent network corresponding to the CA3 region, whose pyramidal cells have activity η_i. They receive input connections from DG cells as well as recurrent inputs from each other, other afferents and inhibition, which are summarily described by a stochastic term δ_i and by a threshold T

$$\eta_i(\overrightarrow{x}) = g\left[\sum_j c_{ij}^{MF} J_{ij}^{MF} \beta_i(\overrightarrow{x}) + \sum_k c_{kj}^{RC} J_{kj}^{RC} \eta_k(\overrightarrow{x}) + \delta_i - T\right]^+, \qquad (1.1)$$

As the virtual rat follows a trajectory in the room, simulating exploration during free foraging, different sets of connections are modified, and a quantification of the amount of information in the population activity of CA3 cells allows to determine the influence of each parameter in the model.

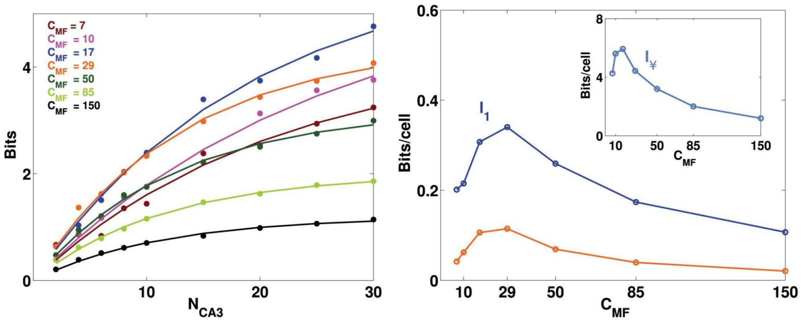


Figure 1.11: From [55]: computer simulations and analytical estimates converge on a quantification of the spatial information in a sample of N_{CA3} units as a function of network parameters, here C_{MF}, the number of DG inputs per CA3 cell, showing a maximum for a plausible value of C_{MF}.

Confirming the analytical calculations by [44], the results of both analytical calculations and simulations with varying parameters show that (Fig. 1.11) the spatial information in CA3 population activity does not depend on the number of fields per DG unit and is maximal when:

- the number of connections from DG to CA3 is low, but not too low;
- importantly, the activity of the CA3 network, which is also in the form of place cells, is sparse.

Moreover, plasticity on the MF synapses is shown not to increase the information content of CA3 representations – DG can exert its driving force through non-modifiable weights.

Representations of multiple spatial maps within CA3

Experimental evidence shows that CA3 cells, in line with theoretical predictions, form a representation of a novel space quite different from previously stored spatial memories, and that essentially orthogonal charts are produced for at least 11 (physically very similar) environments [57].

However, simulations with the same model, when trained to explore a number of different environments, and with associatively modifiable recurrent weights, indicate that several charts can be stored on the same synapses, but with a degree of granularity in the representation of each space – the would-be continuous attractors are in fact only quasi-continuous. In [56] the network is simulated to learn a number of two-dimensional environments, with the CA3 recurrent network allowed to self-organize, i.e. to adjust its synaptic weights with a simple Hebbian rule. This self-organizing model is then compared with a pre-wired version, where the connection strength is defined at the beginning as an exponential function of the distance between the place field centers of a pair of units.

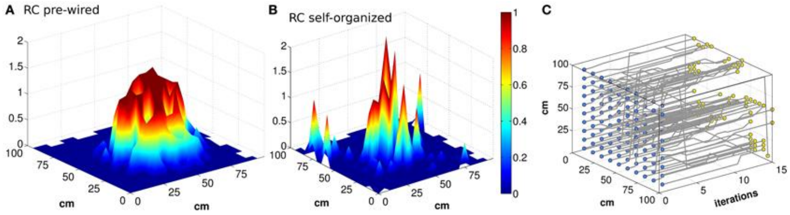


Figure 1.12: From [56]: Contrasted with a pre-wired chart (A), one that self-organizes during exploration (B) is more irregular and granular, in the sense that the continuous attractor is broken into a number of discrete attracting locations (C).

As shown in Fig. 1.12, the information about the newly explored environment can be stored in the self-organizing network, independently of the noise level, but the attractors of the population dynamics appear to have some granularity. In parallel, learning produces a refinement of the place fields that would have resulted from DG inputs alone, as shown in Fig. 1.13.

1.2.3 What happens within one chart?

The models above focus on the representation of multiple environments, coarse-grained in time. A most intriguing phenomenology emerges when looking at the representation of even a limited environment, but with finer temporal resolution. Only a few salient traits of this phenomenology will be mentioned in the following, to be considered in refinements of the simple models above.

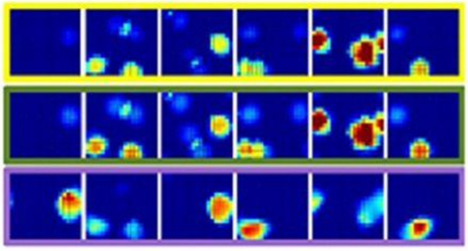


Figure 1.13: The progressive refinement of place fields in the model analyzed in [56]: Six examples of CA3 firing maps in the DG-CA3 model network with MF and RC connections. The top row shows CA3 place fields with no Hebbian learning; the middle row shows the same fields after learning; and the bottom row shows them after mossy fiber inputs are turned off.

Phase precession and its possible role in the memory process

A finding from rodents that must be taken into account in relation to place cells, is phase precession. Place cells are observed to fire action potentials in relation to local theta waves, and O'Keefe and Recce [58] noticed that in a one-dimensional track place cells tend to fire late in theta cycle when the animal enters the firing field of each cell and as it approaches the center of the firing field the firing occurs earlier and earlier in the theta period, often in a burst of action potentials – as if moving backward, i.e., precessing, within the theta cycle (Fig. 1.14).

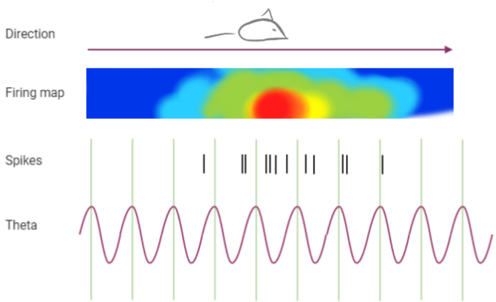


Figure 1.14: A place cell firing in relation to theta waves during one run of a rat.

In one-dimensional environments place cells are directional, meaning they normally have different fields when running in the two directions. Therefore each field is entered at roughly the same position in space, and so the spikes it elicits code for that position also via the exact theta phase at which they occur – an enrichment of the pure frequency code. In two dimensions, however, the phenomenon persists, but each field, typically non-directional, can be entered from multiple directions, hence the additional code loses its meaning as the correspondence between exact position and theta phase does not hold.

Nevertheless, phase precession may play a role in facilitating plasticity that promotes the learning of sequences. This can occur, as cells active at slightly displaced positions A and B can fire together within a theta cycle, i.e., within the appropriate plasticity window, with A, already at the center of its field, firing earlier and thus strengthening its connection and its influence on the firing of B. Blum and Abbott [59] suggest this may serve to store in memory a trajectory: a recent trajectory could then be retrieved by *replaying* it at a speed not necessarily similar to the one at which it was stored. Phase precession may thus be a way for place cells to deposit simple navigational "plans".

Replay, preplay and goal-directed behavior

During rest, whether awake or asleep, place cells can be observed to fire in sequences that roughly match those seen during locomotion, typically but not necessarily in one-dimensional environment. The phenomenon is called *replay* [60] when it occurs after the behaviour, and *preplay* if before. In replay, as the animal is sleeping, resting or before it starts another run on a track or in a box, the place cells corresponding to a learned trajectory would activate sequentially, in forward or reverse order, over short time scales (Fig. 1.15). This has been interpreted as a mechanism used to consolidate the trajectory in memory, and perhaps also to replay possible routes for future decision making.

Here the one-dimensional case from a mechanistic point of view is rather straightforward – a bump of population activity can easily be made to propagate following the remembered route, which has no alternative. In two dimensions there are alternatives, and a new set of intriguing questions arose following the discovery of preplay by Pfeiffer and Foster in 2013 [61]. What they reported was that, during rest, place cells would activate sequentially in relation to the trajectory to be followed shortly,

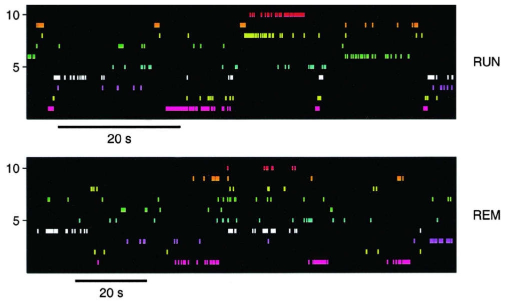


Figure 1.15: From [60]: a raster plot of place cell activity during a run on a linear track and during a period of REM sleep. The run section is scaled to correspond to the sleep period.

to a remembered goal location (Fig. 1.16).

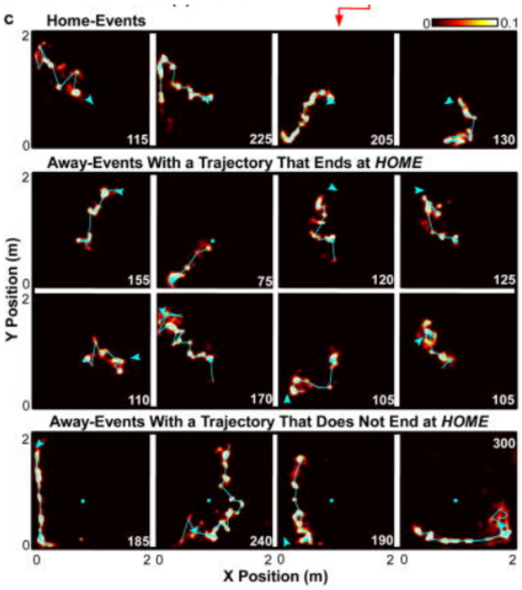


Figure 1.16: Preplay phenomenon from [61]: firing of place cells corresponding to locations in a familiar square box, as interpreted by a decoding algorithm: the frames are summed over time durations indicated in ms in the bottom right corners. Cyan circles correspond to the position of the rat home, cyan arrows to the current position of the rat.

Interestingly, in terms of neural network operations, this phenomenon can be seen as a goal-directed behavior driven by adaptation – the omnipresent characteristic of pyramidal cells, whereby they tend to decrease their firing rate after some time of activity, as if adapting to the input. Such a mechanism has been modeled [62] by adding a simple form of firing rate adaptation to the attractor CA3 network described above. Adaptation gives the place cell code some predictive power – the trajectory decoded from CA3 activity is shifted towards future steps, as shown in

Fig. 1.17.

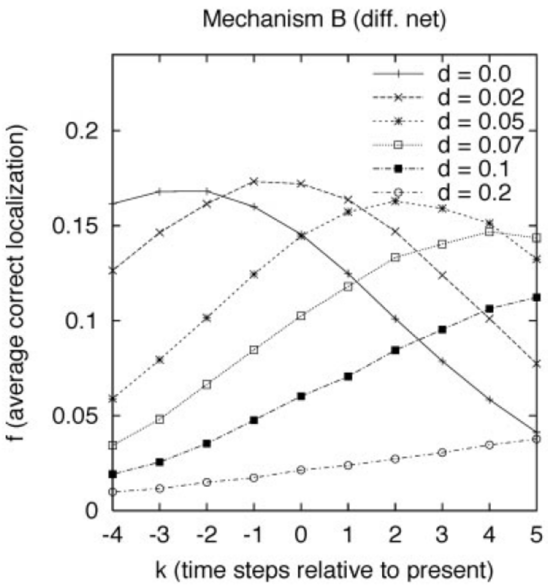


Figure 1.17: From [62]: the spatial location decoded from network activity relative to the present position of a virtual animal (point 0 on the x-axis) for different values of an adaptation parameter d. Steps in the past have negative values.

Many questions of course arise around goal-directed behavior in general: how are goals incorporated in a place cell code, and, most interesting, how is the corresponding neural dynamics operating in the presence of numerous goals. Such questions still await critical experimental advances.

1.3 Out and about: what happens in the real non-ideal world?

However complex the questions may seem, new experimental findings appear as we write and make it even more challenging to develop theoretical models on how spatial memory in the hippocampus may function, in rodents and in other species, including humans. Most intriguing results begin to appear when the experiments move from the conservative lab conditions to more ecological environmental settings. Described in numerous studies, place cells having one place field each in a perfectly square empty laboratory box cease to exist once given more of any relevant context, pointing towards the more complex nature of the memory function.

The hippocampal recordings in bats that started in the lab of Nachum Ulanovsky in the late 2000's have shown that three-dimensional space gives rise to three-dimensional place cells. This fact alone raises several questions from a theoretical point of view. One of their most recent studies, by Eliav et al. [9], focuses on the neural representation of a long one-dimensional tunnel, and the findings make us reconsider many assumptions used in modeling up to date: it appears that on a long track place cells acquire multiple receptive fields of various sizes and peak rates, in a spatial code that appears dominated by disorder. One wonders whether such disorderly representation can be deposited in memory *as is*, just by virtue of associative plasticity.

Several experimental results have shown that grid regularity can be distorted as soon as the environment becomes more complex, for example by non-standard shape of the walls [63] or the presence of goals [64] – on the other hand, the variability in the peak rates of the fields of the same cell, not just in their position, has been shown to be reliable, hence it probably carries some information [65, 66]. Such effects are expected to be huge in the natural environments in which the place and grid cell system has presumably evolved, for example the habitat of the Norway rat [67], Fig. 1.18. These observations call for a theoretical analysis, but defining a mathematical model of an arbitrarily shaped environments containing a number of arbitrary objects is a sure recipe for an arbitrary outcome.

Taking an alternative approach to the question, we ask: is the regularity an artifact of a sterile square environment? Do real memory representations make their living off in the irregularities?

1.3.1 Are charts pieced together by fragments?

To address these questions, let us consider the theory of navigation by fragment fitting introduced by Worden in [1]. In his theory, mammals store their representation of space as a number of independent fragments bound to specific landmarks. When

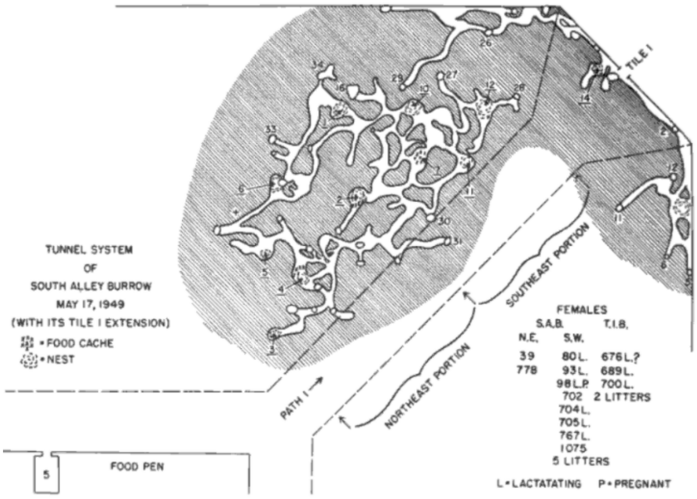


Figure 1.18: From [67]: An example of a burrow system of a Norway rat.

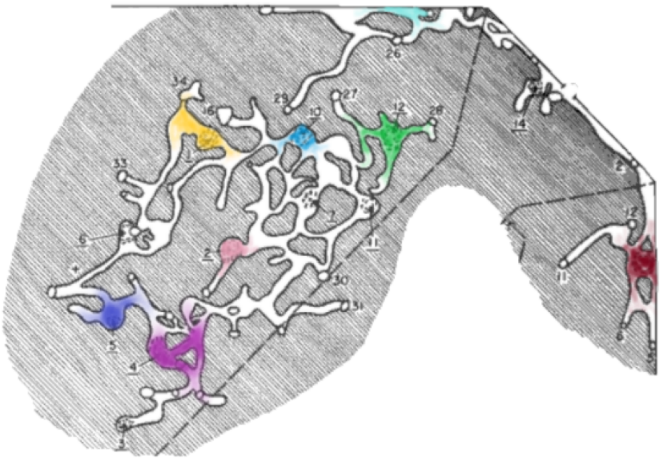


Figure 1.19: Our schematic illustration to the navigation by fragment theory by Worden [1]: separate sectors (in color) of a complex environment are stored in memory as fragments of a puzzle and are fitted together during navigation.

navigating, the animal puts these fragments together rather like solving an unbound jigsaw puzzle (Fig.1.19).

In this thesis we hypothesize that human memory relies on a similar principle, and it extends both to spatial and non-spatial memory. To illustrate this claim, we use both theoretical approaches and behavioral experiments in humans.

In chapter 2 we study our more *regular* model of CA3 from [56] and look closer at the quasi-continuous *maps* that form for each square room. We find that when the learning is asymmetric and sparse, more real-life like, then the memory patterns stored are fragmentary and reflect better the attractive nature of the learned bits.

Section	Main topic	Objectives
Chapter 2	Model of CA3	Study of spatial memory with an attractor neural network, its capacity and dynamics
Chapter 3	Potts network and free recall	Propose a model for short-term memory, study free and serial recall dynamics in humans and with the model
Chapter 4	Free recall	Characterize dynamics of recall, study common tendencies
Section 5.1	Mind wandering	Study the effects of schemata and episodic memories on a model of free thought dynamics
Section 5.2	Poetry memory	Study effects of poetic meter variables on poetry recall

Table 1.1: Summary of the different studies that are described in the following chapters.

In this context we call fragments the well memorized parts of the learning trajectory and test whether the localization accuracy is better within fragments.

In chapters 3&4 we study human behavior in 'lab-like' conditions: we cannot record human place cells in a regular box, but we can observe the more complex behavior arising in spatial memory tasks. We introduce such a task on a homogeneous hexagonal grid and describe the dynamics of recall as biased by the fragmentary schemata that already exist in human memory. For example, we hypothesize that seeing a familiar fragment, like a number of dots on a straight line, should facilitate fragment recall.

Finally, in chapter 5 we give two further approaches to studying the fragmentary nature of memory: in the first part we introduce an ongoing experiment on mind wandering – a schema-driven process of freely latching thought. We aim to test whether imposing a novel schema can bias a free association chain, and whether in absence of an important actor in schema formation, the ventromedial prefrontal cortex, such bias would persist.

In the second part of this chapter we describe how poetic meter and its components help memory when learning poetry. Here our goal is to underscore the disorderly arrangement of the schemata orchestrating the memory fragments.

Chapter 2

Can CA3 be rethought as a fragment assembly?

As described in the introduction, the CA3 network, with its distinctive DG inputs, has been widely considered to operate with attractor dynamics also in the spatial domain, as indicated by the simplified model studied by Erika Cerasti and Alessandro Treves [55]. They have shown that Hebbian learning applied to random exploration in a novel environment allows to form a quasi-continuous attractor in CA3 – a spatial map, and several maps can be learnt on top of each other within the same synapses of CA3.

Several questions remain, which we aim to address here:

- can learning improve indefinitely, or does it saturate?
- how many maps can the network store?
- in what sense are CA3 maps quasi-continuous?
- are the continuous bits related to the trajectories learnt?
- can we model replay in the same network?

2.1 The model

Here we build upon the mathematical model of CA3 proposed in [55]. The simulated CA3 consists of a network of 500 threshold-linear units [68] defined by the firing rate of the unit neurons and the weights of the connections between them. The input to the network is given as firing rate maps of simulated units of DG through sparse connections, with added noise δ and an integrated term of inhibition+threshold T. The firing rate of a CA3 unit i at a position $\vec{x}$ is given as a weighted sum of all the inputs to it at this position:

$$\eta_i(\vec{x}) = g\left[\sum_j c_{ij}^{DG} J_{ij}^{DG} \beta_j(\vec{x}) + \sum_k c_{ik}^{CA3} J_{ik}^{CA3} \eta_k(\vec{x}) + \delta_i - T\right]^+ \tag{2.1}$$

In the current simulations we take the gain term $g = 1$. The threshold T, defined to include inhibition term, is set to control for the sparsity of the CA3 firing a_{CA3} in the sense of the Treves-Rolls population code sparsity [69]:

$$a_{CA3} = \frac{(\sum_i \eta_i(\vec{x}))^2}{\sum_i \eta_i(\vec{x})^2} \tag{2.2}$$

In the reported simulations $a_{CA3} = 0.1$. Note that defined as in eq. 2.2 fixing the sparsity of population code provides for a peaked distribution of activity – with many small (or zero) values and few high values of the firing rate, whereas fixing the population mean would give uniformly low firing rate.

The connectivity c_{ij}^{DG} from DG to CA3 is set to be sparse, 50 random connections (10 % of the active DG units) per CA3 unit. The weights J_{ij}^{DG} of the existing connections are set to 1. The connections c_{ik}^{CA3} are set so that each CA3 unit gets input from 300 other CA3 units, so that the recurrent network is at its best in some sense [38, 55] – neither fully connected nor too diluted. The strength J_{ik}^{CA3} of a connection between i and k evolves following a Hebbian learning rule, increasing for the pairs of cells that fire together:

$$\Delta J_{ik}^{CA3} = \gamma \eta_i(t)(\eta_k(t) - \Lambda_k(t)). \tag{2.3}$$

Here $\Lambda(t)$ denotes a trace of the recent postsynaptic activity and γ is a learning rate.

2.2 Storage of multiple environments

We will start by addressing the first two questions together. In [56] Cerasti and Treves show that the model CA3 network, defined and with the same parameters above, can store multiple sample environments. Each environment is defined as a 1m x 1m square room, assigned a random coverage of place fields of DG units, independently for each separate room. Note that in [55] Cerasti & Treves demonstrated that in their model there is no effect on CA3 map storage if there is one or more fields per DG place cell in an environment. Ten years later, more findings suggest that in larger and more realistic settings place cells, also in CA3 itself, and definitely in CA1, tend to have multiple fields [9, 70, 71], often differing in their size and firing rate range. For this reason here we report the work done on the model defined as in [56], where each place cell of DG has number of place fields drawn from a Poisson distribution, averaging at $q = 1.7$ disorderly distributed place fields per cell, each of the same effective size and peak firing rate – the full parameters space to remains to be explored in future work.

The virtual rat runs through each room following a random trajectory long enough (3000 steps of 2.5 cm) to basically explore all positions multiple times; the firing of DG units is fed upstream to CA3, where it is seen to gradually form a quasi-continuous attractor state, or a *map* of this room. In [56] they let the network learn 4 rooms, and test map storage by giving the network partial cues along a new random trajectory and then comparing the firing of a sample of CA3 cells to the stored templates of activity in every location of each of the four rooms. Fig. 2.1A shows the resulting map classification along a testing trajectory: the correct environment is recognized more often than those coded in the other maps but the latter interfere, and as a result classification accuracy is low. Fig. 2.1B shows the proportion of the maps retrieved in each room and we see that the room learnt last is recognized best, while all the other rooms are classified correctly about half of the time.

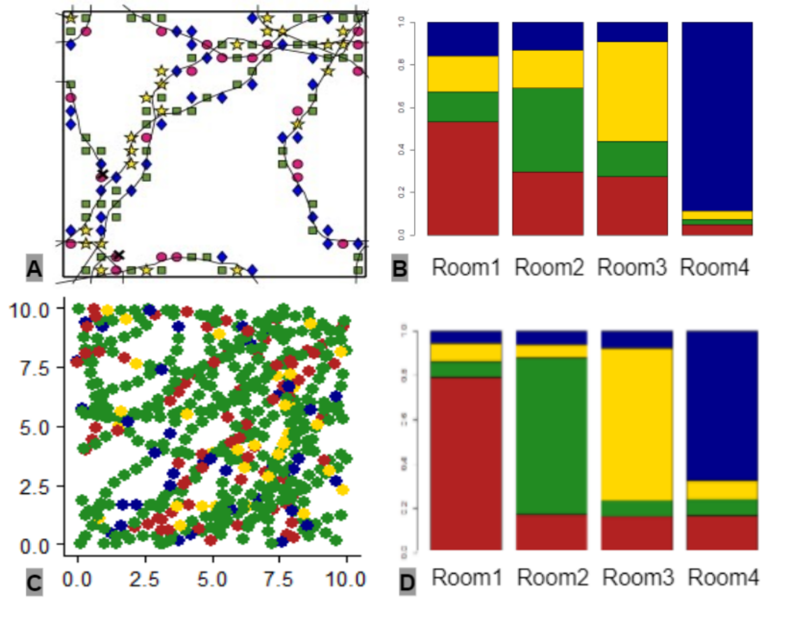


Figure 2.1: Testing the storage of 4 rooms. A – from [56]: environment decoded (color and shape) along a sample trajectory of 300 steps from CA3 activity, then compared to all stored templates; the correct environment is noted by a green square. C – same done with epochs of slower interleaved learning. B, D – the proportion of locations along the trajectory classified to templates in every room stored, B – with the original learning procedure (that of [56]), D – with the slower interleaved learning.

To increase the decoding accuracy uniformly across the environments, we modified the learning procedure: we broke the exploration phase in several epochs of learning and lowered the learning rate. The slower interleaved learning gave better results (see Fig. 2.1 C, D): decoding accuracy is stable over 75% for all the rooms learnt, after an equivalent of accumulated 30 minutes of exploration in each room. Note that all the parameters, except for learning rate, remain the same as in [56].

2.2.1 Storage capacity

We repeated the procedure for different number of rooms and different learning rates. For each number of environments we found an optimal learning rate. Taking the average correct decoding as the measure of stability of storage, we find that the network of 500 units, with the other parameters above, has a storage capacity limit of 10 *maps* (see Fig. 2.2): when the network "tries" to learn 11 rooms, the accuracy of retrieval drops to chance level for all the rooms.

This result is consistent with the independent remapping of the place cell activity in CA3 shown to happen at least for 11 experimental rooms in rats in [57], also after around 30 minutes of random exploration, even though the real CA3 networks differs from our overly simplified model in so many ways, starting from its connectivity.

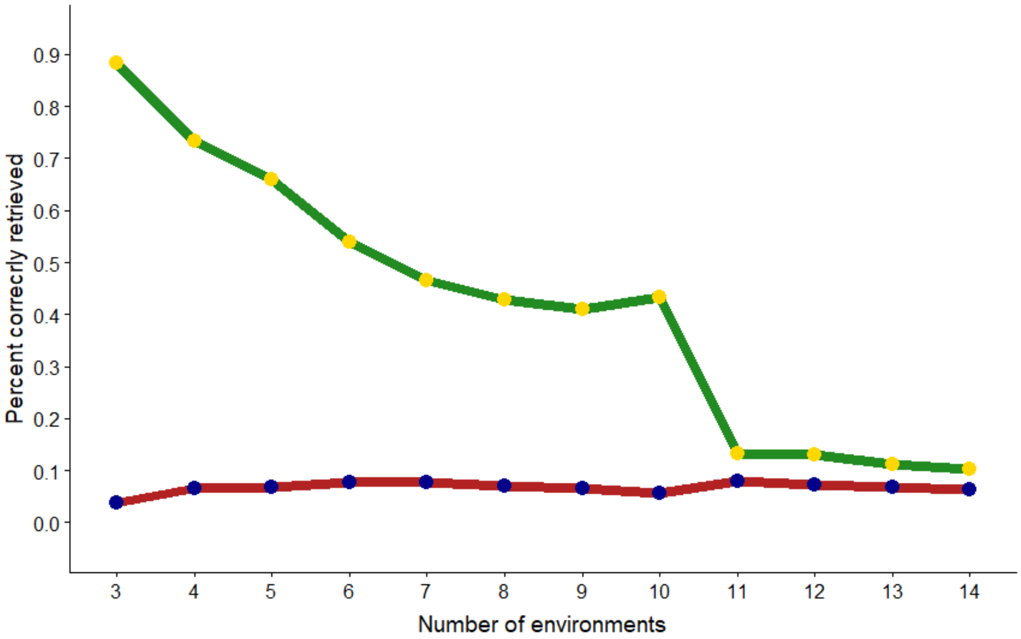


Figure 2.2: Average over the number of stored rooms percent correctly retrieved maps (yellow points, green line) and incorrectly retrieved maps (blue points, red line) for different number of rooms learnt.

The mathematical analysis of an idealized model indicates a substantial storage capacity for a CA3 network of recurrently connected place cells [38], in the order of a hundred charts, with realistic rat parameters, with an order of 10^4 connections per unit, as mentioned earlier [39], but in our case, with only 500 neurons, the estimate has to be scaled down accordingly.

As mentioned earlier, in our simulations the activity of CA3 neurons is set to be sparse with the parameter $a_{CA3} = 0.1$. Following the calculations of [38] and taking the factor $k_d = 3.6$ (also from [38]) the effective size $|M|$ of an environment used in our simulations reads:

$$\frac{1}{|M|} = \frac{a_{CA3}}{k_d} = \frac{0.1}{3.6} \approx 0.02778 \tag{2.4}$$

As our network is neither fully connected nor extremely diluted, with 300 recurrent connections per CA3 cell, according to Fig. 1.8 right the 2D capacity of our network should lie somewhere between the limits of ~ 0.01 and 0.04, which corresponds to being able to store between 3 and 12 sparse maps. The limit of 10 maps that we find for the network numerically (Fig. 2.2) falls in between.

The network storage capacity depends on the number of units in it and the recurrent connectivity, and we wanted to see how the capacity limit changes when we scale the network down. Decreasing the number of neurons by 100 and 200 (keeping the connectivity ratio at 3/5), we were surprised to find the drop in the correctly retrieved curve as on Fig. 2.2, at 10 maps. This observation leaves us puzzled as to whether what we find here is indeed a capacity storage limit, or a limit imposed by the parameters of learning and/or definition of the spatial maps. More work is needed to better study the question.

However, compared to the analytical calculations of [38], the simulation of a self-organizing, more detailed version of the model points at the role of disorder in determining granular charts, that represent space only quasi-continuously. Further studies are needed to better quantify this phenomenon and its influence on the storage capacity of the network. Such quantification appears to be important in interpreting the representation of natural habitats, for example the representation of extended, quasi-one-dimensional spaces by bats, currently being investigated in quasi-naturalistic conditions [9, 72].

We decided to look closer at the continuity issue, turning to the next two questions from our list together, namely: in what sense are CA3 maps quasi-continuous? And are the continuous bits related to the trajectories learnt?

2.3 Fragmentary 'recall' in a sparsely learnt environment

Even though the average fraction of steps classified correctly increases with the slower interleaved learning, the granularity of the attractor states persists (see Fig. 2.1 A and C where dots of the same color tend to cluster together): the network seems to *jump* between the stored maps when unsure. We hypothesised that the learning trajectories influence the local quality of storage and decoding of each map.

In order to understand better the difference between the parts of the environment that are decoded more reliably and those parts where all the different maps are decoded with the same probability (Fig. 2.1), we get back to teaching the network 4 environments and look at the relationship between the learning and testing trajectories. With the original learning procedure, the virtual rat explores each room extensively, as its learning trajectory covers the environment multiple times.

With a shorter exploration trajectory, covering only a part of the environment, we can separate the parts learned continuously and those left for generalization. On Fig. 2.3 we show this contrast: indeed, the testing locations that are close or coincide with the learning trajectory are classified more accurately than those in the area not covered in exploration – there, the decoding probability of each map is close to chance level.

This finding reflects the attractive *fragments* of memory that arise plainly from an uneven learning procedure. On Fig.2.4 we give a visual link between the featureless map learning in CA3 and the notion of navigation by fragment in an animal mind: we can consider each map of a room to have only formed partially, and so the fragments decoded correctly, salient to the parts learnt in said room, are interleaved with decoding an incorrect environment. In a more information-rich environment like that shown again on the right panel of the figure, the fragments are hypothesised to form in relation to important locations in a den (e.g., where food is stored, or where family sleeps).

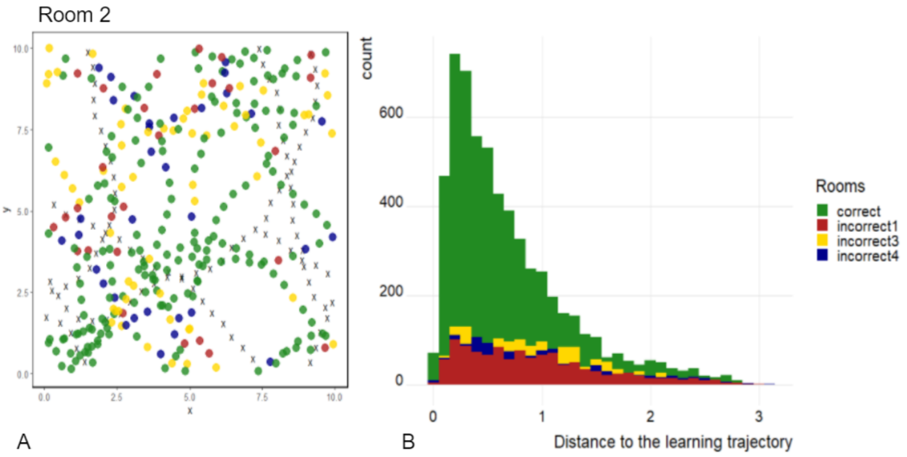


Figure 2.3: Fragmentary recall in decoding. A – map decoded along a testing trajectory of 300 steps(round points, color); the learning trajectory in the correct environment (room 2, green) is shown by grey crosses. B – the templates decoded along the testing trajectory as a function of the distance to the trajectory used in learning.

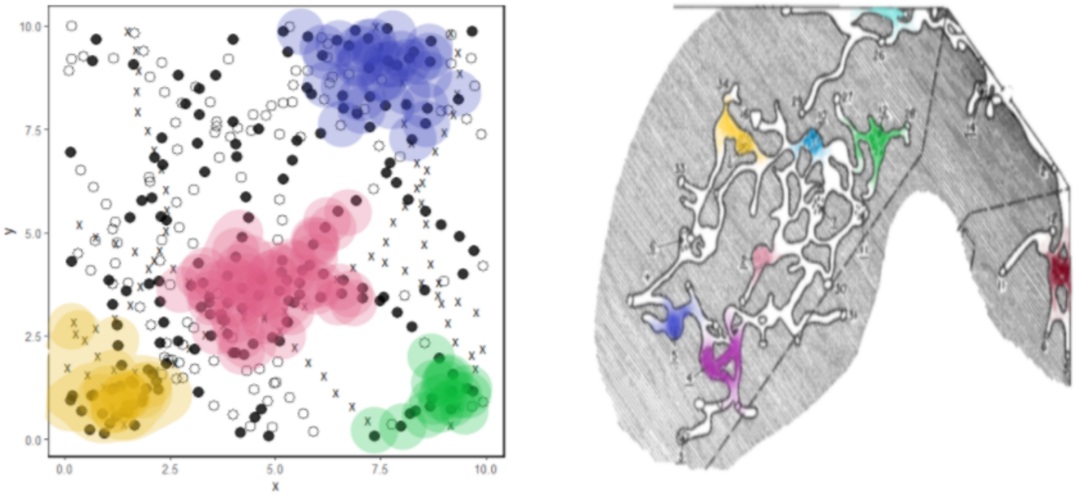


Figure 2.4: Left – A schematic illustration of fragmentary encoding of the map 2 in the model of CA3 (same as left of Fig. 2.3, but decolorized so that correctly decoded locations of the trajectory are now shown in black, and all the incorrect are shown as empty circles). The portions highlighted in color are the parts where the learning trajectory was and decoding is most reliable. Right – For a visual analogy, we show again the schematic illustration of the Norway rat den with most salient fragments of the map in color.

Interestingly, a recent study [73] suggests a mechanism in DG that could support switching between generalization to discrimination when novel information needs to be stored. The authors report observed depolarization of membrane potential in DG granule cells when the animal is exposed to a novel environment, and propose a model accounting for an external cholinergic signal to enable the above-mentioned switch. Incorporating such a mechanism within our framework could potentially help form a more stable map for a plain environment, like those that we show here, but in real-world conditions with multiple important novel objects such a mechanism should produce even more inbalance in learning, forming fragments of memory biased towards the new or the more salient information.

2.4 Neuronal adaptation and prediction

Another problem we wanted to address was the phenomenon of replay in neuronal firing, described in the introduction. Replay is the process of offline reactivation of the place cells active during exploration in the order they were activated in the actual trajectory, but on a shorter timescale [60]. Because it is usually observed during sleep or just before a task run, this phenomenon has been associated with memory consolidation and planning.

We wanted to see if within our framework we could model replay as a process that would arise within the recurrent connections without external input: again, because replay is observed *offline*, an intrinsic cue should be all that triggers the sequence of activation.

For that we borrow the idea from [62]: also described more in detail in the introduction, the idea is to add neuronal adaptation to the equation. This intrinsic property of neuronal firing in response to stable input is brought by fatigue, and computationally drives the activity forward. What happens physiologically is that, given monotonous input, a neuron first fires a lot, but over (quite a short) time its activity in response to the same stimulus decreases.

To model this effect within the current model we subtract from the neuron firing rate, summed at t as in Eq. 2.1, a fraction d of its own recent activity at each time step, as proposed in [62]:

$$\eta_i(t) = \eta_i(t) - d \sum_{k=-\infty}^{t-1} \eta_i(k) \tag{2.5}$$

As in [62], adding this term to our model gives the firing code some predictive power: while the animal moves along a trajectory, the location decoded from the activity of CA3 is shifting toward future steps, varying with the parameter d (Fig. 2.5). Without adaptation, $d = 0$, we decode the current position cued to the virtual animal, and with increasing d the location decoded moves to the future. We hypothesize that with a suitable value of d, given just a cue, the recurrent network will *replay* trajectories within a fragment of memory. However, within the current framework it is difficult to define the regions of interest, and the notion of a landmark or a goal is needed to further address this question.

Here the mechanism moving activity forward is based on firing rate adaptation. Somewhat similar results were obtained by Romani and Tsodyks in [74], but there the term of synaptic depression drives the propagation of the network activity in time. Both our and Romani & Tsodyks approaches rely passively on the networks inherent properties for modeling replay, while a new more *active* approach was suggested by Spalla et al. in [75]: the authors consider an asymmetric component to the the synaptic plasticity, that is acquired through learning and produces dynamic retrieval similar to online replay.

Stella et al. in [76] observed that the activation of place cells in CA1 during sleep after aimless foraging appears to be similar to random walks in the environment used in foraging. Our model, while not suited for the retrieval of trajectories leading to specific goals, can well serve to produce these random trajectories, but to understand better the dynamics of aimless foraging, we have first turned to a behavioral task in humans that might rely on a similar process — free memory recall, studied in the next two chapters. The described mechanism of adaptation-driven replay aiding Hebbian learning could serve human memory for more complex multidimensional material: in chapter 5 we will dive deeper into this conceptual link, suggesting that a process similar to replay could drive the schemata involved in remembering poetry.

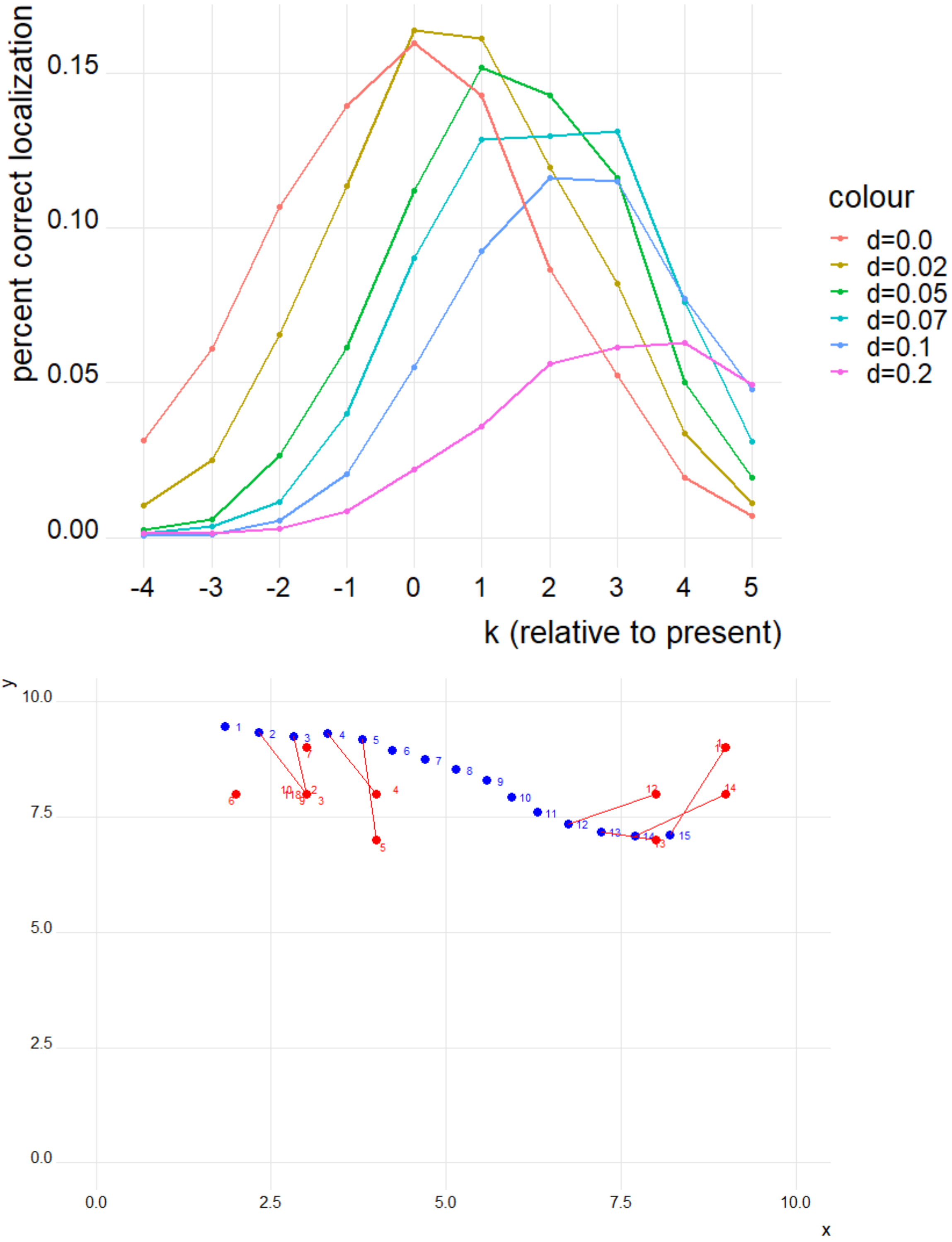


Figure 2.5: Predictive effect of firing rate adaptation. Top - decoded location relative to the current position of the animal (0 on the x-axis) for different values of the adaptation parameter d. Bottom - an example of a simulated trajectory of the animal (in blue) with respective locations (in red) decoded from the activity with $d = 0.07$ (we do not add all the links in red for better visibility: e.g.,the middle points of the trajectory are all classified to be the interim location linked to 2 and 3)

Chapter 3

Recall and Potts network

In an attempt to describe human spatial memory if given as plain and dull environment as the ones traditionally used for animal studies, we introduce a simple task. Our CA3 model is based on the neuronal recording data of freely moving rodents, and recording the activity of single neurons in freely moving humans, even if it were useful, is not feasible. On the bright side, in behavioral tasks in humans we can ask for much more than we can ask from animals, for example, to memorize and recall different types of material.

Several recent studies using intracranial recordings in human participants [77, 78] find signatures of spatially selective cells in the parahippocampal area, that are also activated during *free recall*. This gives us additional motivation to study dynamics of recall in order to characterize the general properties of memory across species.

Here we will discuss the relevance of a notable model of memory recall, argue in favor of our own Potts network[1] as an alternative STM mechanism, capable of immediate recall, and describe some prominent features of recall arising from the experimental evidence we gathered.

[1] Note: the work on the Potts network was done by Kwang Il Ryom and Vezha Boboeva in our joint paper [79], and in this thesis I will not go into great detail of this study. Instead I will focus mainly on the results of the experiments, conducted and analyzed by me, using the results of the modeling to illustrate further the thesis of fragmentary understanding of memory.

3.1 Models of free recall

Studies of free recall have been showing a common trend since the 1960s [80, 81]: when asked to name as many as possible out of a list of M random words, participants give on average $k\sqrt{M}$ correct answers.

A simple attractive account of such behavior is that offered in [82]: a parameter-free associative model gives a scaling law of $\sqrt{1.5\pi M}$ (see Fig.3.1). It is worth pausing for a moment to consider what the data in Fig.3.1 imply. With the number of items recalled always growing, it is striking how much a person can remember, if never stopped. But were not we told that short-term memory is limited in capacity? (by, say, seven items: [83], or four: [84]).

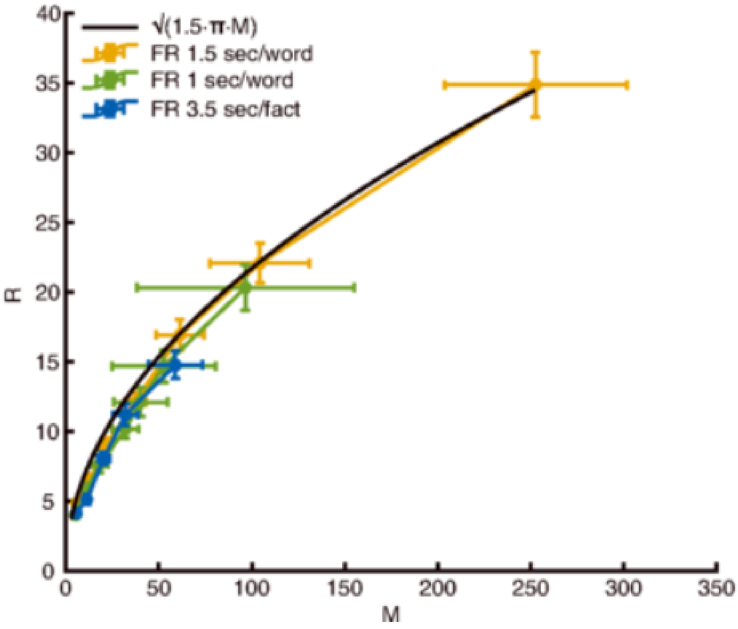


Figure 3.1: From [82]: number of recalled items R for various list length M in free recall of words and facts with different presentation rate.

3.1.1 Is free recall – navigation by fragments?

It may well be so when the task is to recall words: one cannot control for all the possible associations each participant can have between any of individual high-dimensional stimuli, like words. It is in itself a very interesting research question and we try to approach it with a free association paradigm in chapter 5. What about the contradiction between the limited capacity of STM and the unbounded recall in Fig.3.1? Perhaps it could in fact be addressed through a fragmentary understanding of memory: the more there are fragments or schemata that bind the stimuli from the list, the better is *the navigation* through the list, and the higher the recall capacity. In contrast, short-term memory experiments, for example, serial recall with short presentation time and penalizing mistakes in recall order might yield very limited recall performance, because *there is no time for fragment activation* – a hypothesis that comes from the results of our experiment described in part 2 of chapter 5.

To be able to test these hypotheses, however, we need to reduce the dimensionality of the stimuli, so as to bring individual fragments to the light without the need to account for all possible semantic associations. This is why we have chosen to run a spatial memory task, aiming to look closely at what happens during recall.

We have designed an experiment in which participants are shown a progressively increasing number of spatial locations on a hexagonal grid and are asked to recall as many as they can by clicking on the corresponding positions on the screen (Fig. 3.3). This way we could record every (mis)click and follow the series of recall. First, we needed to see if the unboundedness of the free recall capacity holds for our simple spatial stimuli and what constrains this capacity. Secondly, we have explored giving a trajectory structure (like navigating in structured environments in rodent experiments) and introducing simple non-salient landmarks (also, like in some rodent experiments) to see whether it improves recall performance in restrained conditions (serial recall). Then we have compared the performance of the participants with that extracted from latching dynamics in the Potts network (see next subsection). Lastly, we have analyzed the recall sequences in free recall for the signatures of schema-driven fragments.

3.1.2 Potts network

The Potts network has been studied as a model of interacting cortical patches and its *latching* as a model of retrieval of memory patterns distributed across the cortex

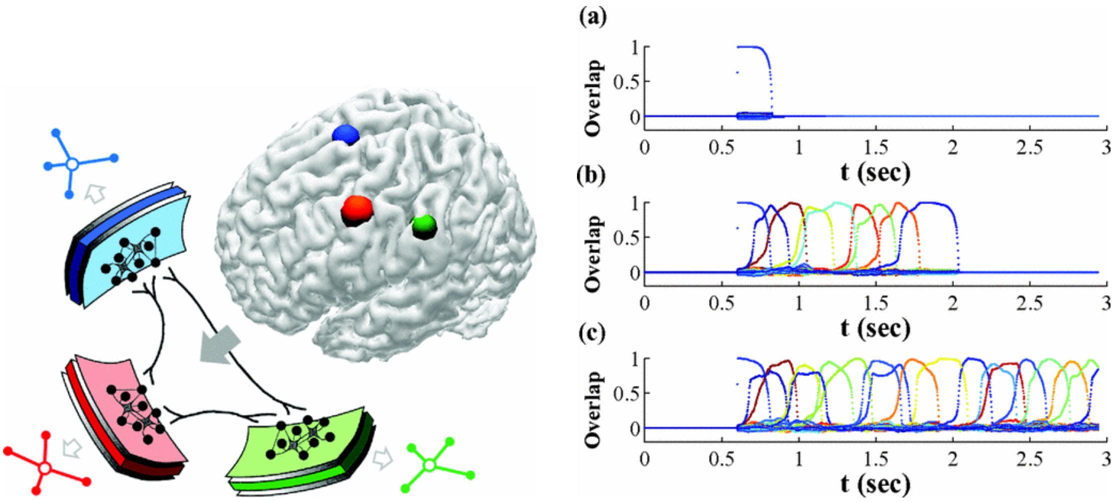


Figure 3.2: From [85]. Left: A schematic illustration of a Potts neural network. Here each of three Potts units (colors) can be in 4 states. Right: examples of latching dynamics in a Potts network: the correlation of the current state of the network with the stored patterns (each of a different color): (a) absense of latching after one retrieved pattern; (b) latching that eventually dies out; (c) self-sustained infinite latching between patterns.

[85–87]. Earlier research has demonstrated that this model helps characterize the structure and dynamics of an extended memory system, for example, the complex problem of the storage of correlated semantic memories [87] or impairments in the short-term storage of phonemes in the phonemic output buffer [88].

The Potts network is a distributed memory model, that comprises several interconnected cortical patches, called Potts units, each consisting of many model neurons that, as a collective, can be active in one of S local attractor states (see Fig. 3.2, left panel). Attractor dynamics, due to the local and global plasticity of the network, leads to the capability to retrieve p global attractor states of the cortical activity. The network acts as an auto-associator, in fact, capable of retrieving these p patterns when given a partial cue. In Braitenberg's skeleton model [89] focusing on the N pyramidal cells of the cortex, there are $\sqrt{N}$ compartments, each with $\sqrt{N}$ pyramidal cells fully connected with each other; this organization would lead to a number S of local attractor states per compartment limited to be at most of order $\sqrt{N}$, while the number C of effective connections each compartments receives from other patches is left undetermined, but at most again of order $\sqrt{N}$.

Irrespective of Braitenberg model, Iddo Kanter analyzed a specific version of the Potts autoassociative network, back in the 1980's [90], and showed that the maximum numer of (global) memory patterns it could retrieved scaled like $p_c \sim CS^2$, therefore potentially supralinearly in N.

Previous studies of the Potts network [85, 86] show that with an added feedback term the network retrieves a pattern – it reaches a global attractor state – and stays in it until neuronal fatigue destabilizes it, and then its activity either dies out or reaches another attractor state (see Fig. 3.2, right panel). This saltatory dynamics among attractive network states was named *latching*.

In [79] we hypothesize that, when restrained to a subset of patterns, Potts network latching can serve as a basis for short-term memory retrieval, or *recall*. We find that a Potts network can indeed reproduce some of the features of the human behavior in recall for different types of material, provided we use suitable measures of recall performance and, in the case of serial recall, we add to the basic plasticity model an extra heteroassociative component, as discussed below.

We can look at latching dynamics as a semi-stochastic process linking retrieved memory fragments. If the extra heteroassociative component is assumed to be localized in prefrontal cortex (see Chapter 5), successive patterns can be interpreted as reactivating the context needed for a full episodic memory replay.

3.1.3 Definition of the model and parameters

In [79] we consider different models for STM, but here, for comparison with the experimental data we will only refer to one of the proposed models, and call it Model 2, so that we keep in mind that it is only one a number of possibilities for modeling short-term memory in free recall. Let us define the model in general terms.

As in the basic definition, in Model 2 each Potts unit has S active states, indexed as $1; 2; \ldots; S$, representing local attractors in that patch, and one background-firing state (no local attractor is activated), the 0 state. The N units interact with each other via tensor connections, that represent associative long-range interactions through axons that travel through the white matter, while local, within-gray-matter interactions are assumed to be governed by attractor dynamics in each patch. The values of the tensor components are pre-determined by the Hebbian learning rule, which can be contrued as derived from Hebbian plasticity at the synaptic level:

$$J_{ij}^{kl} = \frac{c_{ij}}{c_m a(1-\frac{a}{S})} \sum_{\mu=1}^{p} (\delta_{\xi_i^\mu k} - \frac{a}{S})(\delta_{\xi_i^\mu l} - \frac{a}{S})(1-\delta_{k0})(1-\delta_{l0}), \tag{3.1}$$

where c_{ij} is either 1 if unit j gives input to unit i or 0 otherwise, allowing for asymmetric connections between units, and the δ's are the Kronecker symbols. The number of input connections per unit is c_m. The p distributed activity patterns which represent LTM items are assigned, in the simplest model, as composition of local attractor states ξ_i^μ ($i = 1; 2; \ldots; N$ and $\mu = 1; 2; \ldots; p$). The variable ξ_i^μ indicates the state of unit i in pattern μ and is randomly sampled, independently on the unit index i and the pattern index μ, from $0; 1; 2; \ldots; S$ with probability

$$P(\xi_i^\mu = k) = \frac{a}{S}(1 - \delta k, 0) + (1-a)\delta_{k,0} \tag{3.2}$$

The parameter a is the sparsity of patterns – fraction of active units in each pattern; the average number of active units in any pattern μ is therefore given by Na. In the simulations reported here $S = 7, a = 0.25$.

Local network dynamics within a patch are taken to be driven by the input that the unit i in state k receives

$$h_i^k(t) = \sum_{j\neq i}^{N} \sum_{l=1}^{S} J_{ij}^{kl} \sigma_j^l(t) + w[\sigma_i^k(t) - \frac{1}{S}\sum_{l=1}^{S} \sigma_i^l(t)], \tag{3.3}$$

where the local feedback w models the depth of attractors in a patch – it helps the corresponding Potts unit converge to its most active state. The activation along each state for a given Potts unit is updated with a *soft* max rule

$$\sigma_i^k(t) = \frac{exp[\beta r_i^k(t)]}{\sum_{k=1}^{S} exp[\beta r_i^k(t)] + exp(\beta[U + \theta_i^A(t) + \theta_i^B(t)])} \text{ if } k > 0, \tag{3.4}$$

$$\sigma_i^0(t) = \frac{exp(\beta[U + \theta_i^A(t) + \theta_i^B(t)])}{\sum_{k=1}^S exp[\beta r_i^k(t)] + exp(\beta[U + \theta_i^A(t) + \theta_i^B(t)])} \text{ if } k = 0, \tag{3.5}$$

where U is a fixed threshold common for all units and β measures the level of noise in the system. Note that σ_i^k takes continuous values in $(0, 1)$ and that $\sum_{k=0}^S \sigma_i^k = 1$ for any i. The variables r_i^k, θ_i^A and θ_i^B parametrize, respectively, the state-specific potential, fast inhibition and slow inhibition in patch i. The state-specific potential r_i^k integrates the input h_i^k by

$$\tau_1 \frac{dr_i^k(t)}{dt} = h_i^k(t) - \theta_i^k(t) - r_i^k(t), \tag{3.6}$$

where the variable θ_i^k is a specific threshold for unit i and for state k.

Taking the threshold θ_i^k i to vary in time to model adaptation, i.e. synaptic or neural fatigue selectively affecting the neurons active in state k, and not all neurons subsumed by Potts unit i

$$\tau_2 \frac{d\theta_i^k(t)}{dt} = \sigma_i^k(t) - \theta_i^k(t), \tag{3.7}$$

the Potts network additionally expresses latching dynamics, the key to its possible role in short-term memory.

The unit-specific thresholds θ_i^A and θ_i^B describe local inhibition, which in the cortex is relayed by at least 3 main classes of inhibitory interneurons [91] acting on $GABA_A$ and $GABA_B$ receptors, with widely different time courses, from very short to very long.

In the Potts network it has proved convenient, in order to separate time scales, to consider either very slow or very fast inhibition [66]. Here, we consider a case in which both slow and fast inhibition are taken into account. Formally, we have two inhibitory thresholds θ_i^A and θ_i^B (to denote fast, $GABA_A$ and slow, $GABA_B$ inhibition, respectively) that vary in the following way:

$$\tau_A \frac{d\theta_i^A(t)}{dt} = \gamma_A \sum_{k=1}^S \sigma_i^k(t) - \theta_i^A(t), \tag{3.8}$$

$$\tau_B \frac{d\theta_i^B(t)}{dt} = (1-\gamma_A)\sum_{k=1}^{S} \sigma_i^k(t) - \theta_i^B(t), \tag{3.9}$$

where one sets $\tau_A < \tau_1 \ll \tau_2 \ll \tau_B$ and the parameter γ_A sets the balance of fast and slow inhibition. In the presented simulations we fix it intermediate regime $\gamma_A = 0.5$.

In the Model 2 that we will present here a parameter regulating firing rate adaptation is reduced selectively for the neurons that are active, in those patches, in the representation of the L items. That is, we decrease adaptation, by subtracting from the adapted threshold (θ_i^k) a term $\Delta\theta$, for the Potts states that are active in any one of the L patterns,

$$\tau_2 \frac{d\theta_i^k(t)}{dt} = \sigma_i^k(t) - \theta_i^k(t) - \Delta\theta\Omega(\sum_{\mu=1}^{L} \delta_{\xi_i^{mu},k}). \tag{3.10}$$

In the following sections we will only refer to this version of the model, calling it Model 2 and discuss how varying the adaptation decrease term $\Delta_t heta$ we can observe latching dynamics resembling that of human recall. To model serial recall we add a heteroassociative term within the synaptic weights definition (Eq. 3.1) with varying strength γ.

3.2 Experiment: free recall of spatial locations

The first experiment described here is a free recall task of locations on a hexagonal grid (see Fig. 3.3). The aim of the experiment was to characterize human recall capacity in a spatial recall task, compare it with the known limit for the recall performance with other material and with what could be presumed would be the corresponding measure, from latching dynamics of Potts network.

First we will describe the procedure for conducting experiments online, adopted just before the covid-19 pandemic and adapted to most of the experiments that are described in this thesis, unless noted otherwise.

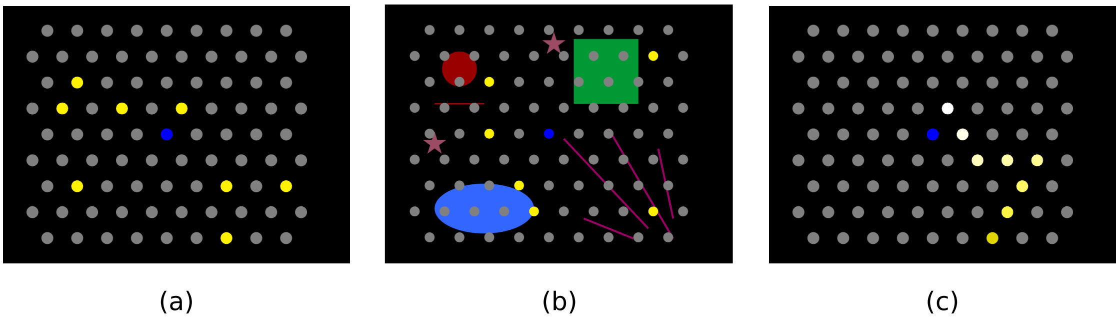


Figure 3.3: Sample stimuli used in the experiments. Participants were presented with a grid of grey dots on a screen, after which a series of yellow dots appeared. Subsequently, after they had disappeared, participants had to recall the locations by clicking on their positions. This experiment was carried out under several different conditions. **(a)** The dots appeared simultaneously and then disappeared all together. The participants then had to freely recall their positions. **(b)** Same as in (a) but with additional landmarks, intended to probe whether landmarks help memory recall. **(c)** In this case, the dots appeared one by one (white to yellow) and formed a continuous trajectory, contrary to (a) and (b), after which participants performed serial recall.

3.2.1 Note on the general procedure for online experiments

All of the experiments described in this chapter were conducted online, with participants recruited through https://www.prolific.co/. The platform provides a pool of participants from all over the world that can chose to take part in an experiment for a remuneration proportional to the time that the task requires. For the experiments with words, like the one described in Chapter 4.2, we only recruited native speakers of Italian. For the other tasks, no filter was applied to the sample, except for *not having taken part in our other experiments of the same type*, i.e., in other recall tasks.

Through Prolific we were able to control that the participants were using desktop screens (not smartphones or tablets) to complete all the tasks. The experimental setting constrained to fixate the relative dimensions of the task boards and fonts.

In all the experiments, we asked the participants to agree to the experimental procedure with a consent form approved by the Ethical Commitee of SISSA.

Before the experiments described in this chapter, and before the Covid-19 pandemic, we tested the participants in a lab setting using the same experimental paradigms, so when we moved to the online setting we had a baseline to compare to.

Unlike the participants in the lab, the online participants had the ability to cheat, unobserved, e.g. to take pictures of the screen and then ”recall” perfectly. To account for that, we discarded the blocks of data where the participants' performance deviated more than 2 standard deviations above or below the sample mean. If more than 20/35 trials were to fall into this category, we discarded the participants' data. Overall we discarded in this way almost 9% of the data, 4% due to deviations above and 5% below. If not noted otherwise, the number of observations is reported after exclusion. The resulting data was not significantly different between the online and offline participants.

Software

All the experiments reported in this work have been written in Javascript by me – using D3.js for visual features and everything written from scratch, unless stated otherwise (the experiments of Chapter 5.2). No additional software was used for the experimental designs. Data analysis was conducted mainly using R and Python.

On sample size

The two experiments described in this chapter were designed in parallel, and the sample size was defined to rend them comparable. The experiment on serial recall needed balancing participants across order of conditions, which is described more in detail in the experimental setting, and we chose to test 3 participants in each of 12 condition-presentation time order combinations, so that to obtain a balanced design with 36 participants. For free recall experiment we collected data from 40 participants to have comparable results.

3.2.2 Experimental procedure

In this experiment we tested the participants' ability to recall spatial locations on a grid in any order. The stimuli were the locations highlighted in yellow on a hexagonal grid (see Fig. 3.3). The sets of stimuli were presented all at once, and the participants (N = 40, F = 12, Age: mean = 25.5, sd = 6.26) were instructed to repeat as many as they could recall, by clicking on the dots in the grid. For each set size[2] L in {4, 6, 8, 12, 16, 24, 32}, the participants had 5 trials to do, each trial allowing for $2L-$*(number of correctly recalled items)* clicks overall, unless all correct items were recalled already (or $2(L - \#correct(t))$ of remaining clicks at time t). For example, if participants correctly clicked 3 dots in a trial with $L = 4$, they were given another 2(4 - 3) = 2 clicks. Instead if they clicked incorrectly once and then correctly once, they had 2(4 - 1) - 1 = 5 clicks left.

A set of size L was presented for $\log_2 L$ seconds. The choice of presentation time was motivated by the results of the experiment described in Section 3.3, allowing the same time for memorizing each item of a set as given time per item in serial presentation for any value of L.

3.2.3 Results

We refer to the the joint paper [79] for the modelling results, and focus here on the experimental ones. Let us just say that we compare the quantitative measures on latching dynamics, constrained to a small number L of the p activity patterns stored in the network, serving as an STM storage of long-term memory items, with the same measures of performance of human participants in a free recall task. For example, in the intermediate regime of latching shown in Fig.3.2b we count 9 *recalled* items, two of which are repeated twice and one is repeated thrice (the blue curve). Then, in the measures limited by repetition, this simulation scores 1, and a mistake would be counted for each item not from the pool of L STM items.

As it is problematic to establish a correspondence between human recall time and simulation time in the Potts model, we define another quantity: we compute the number of correctly retrieved items, ignoring errors and repetitions, M_R, within a given number of consecutive latches, denoted by $g(L)$.

Clearly, the parameters of the experimental protocol can be expected to affect

[2] when we described the model by [82, 92] list length was referred to as M; here we passed to calling L list length in our experiments and number of items in STM in Potts network.

recall, including the amount of time allocated for recall. However, in our experiment, participants only need to click on the correct locations (as opposed to typing in the words they recall [82]), and setting a fixed recall time may seem *ad hoc*. As an alternative, and to further explore the validity of latching dynamics as a model for this experiment, we give participants a limited number of clicks adjusted in the trial, set as $2(L-h(t|L))$, where $h(t|L)$ is the number of correctly recalled dots up to that point in time. Then we computed M_R, defined as the number of correctly recalled dots for a given L ignoring errors and repetitions, and compute the same measure from simulations with the Potts model, setting $g(L) = 2(L-h(t|L))$.

We find a reasonable agreement between the performance of the Potts model and human subjects in our experiment, where both cases show a slope of approximately 0.5 (Fig. 3.4). This suggests that latching dynamics capture some aspects of the underlying neural mechanisms of free memory recall, perhaps related to the random walk nature of the trajectory, although the exact details depend on the paradigm.

If limited by errors, the network cannot recall beyond its STM capacity

Now we want to look at the errors, which often interfere in recall, altering its dynamics [93]. We hypothesize that the limited capacity of recall, often seen in short-term memory tasks [83, 84], may be due to the interference of other long-term memories.

In the scaling model [82], a quantity R is defined as the number of recalled items until the searching trajectory enters a loop, which is then iterated indefinitely. STM items are drawn, in their framework, from a virtually unlimited reservoir of (LTM) memory items. Since they define transitions between items as being completely deterministic and based on the largest representational similarity between them, trajectories always enter a loop. Given such simple transition rules, the relation $R \propto \sqrt{L}$ can be derived, where L is the number of items to recall.

So far we have ignored errors (extra-list items) in order to compare with [82, 92]. Note that errors are not discussed in their conceptual model and experiment, in which retrieval of extra-list words is simply dismissed as irrelevant. The beauty of

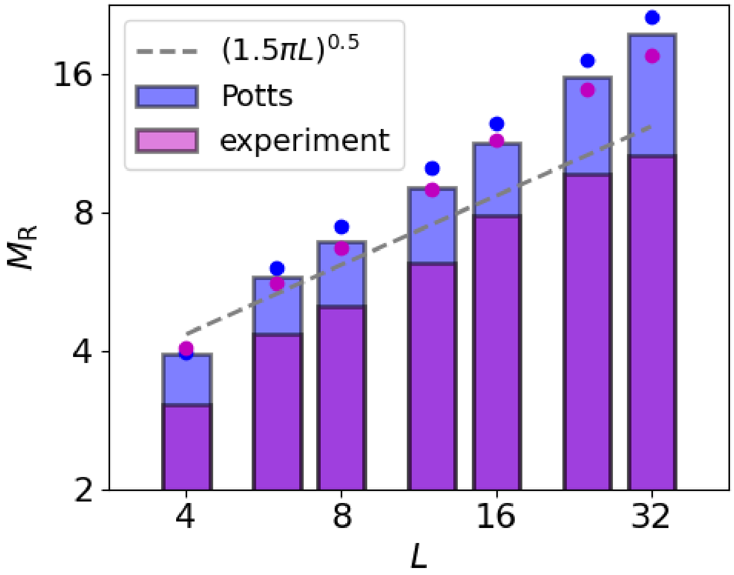


Figure 3.4: Free recall of locations in a 2D grid also shows an approximate $\sqrt{L}$ dependence (note: L was noted M when citing [82]). M_R, the average number of correctly recalled locations in our experiment, is shown by the height of pink bars in a log-log scale. The distance from the bar to the dot of the same colour corresponds to the standard deviation of the mean. Results of 40 participants are pooled together. The same quantity M_R is computed, from simulating Model 2 (one of the versions of a Potts network we describe in [79]), as the number of correctly retrieved STM items within a given number of consecutive latches set as $2(L-h(t|L))$, where $h(t|L)$ is the number of correctly recalled STM items up to that point in time (blue bars). The dashed gray line is the theoretical prediction of R in [82]. Both results, from our experiment and the Potts model, show an approximate $\sqrt{L}$ trend.

their treatment, in fact, stems from the simple question they pose, without getting into how the recall process happens dynamically in the brain and how LTMs affect performance of free recall. These questions are our own interest in this work.

Again, we consider lengths of $g(L) = 2(L - h(t|L))$, where $h(t|L)$ is the number of correctly retrieved STM items up to that point in time; within this sequence we count the number of correctly recalled STM items until there is either an error or a repetition. We compute this quantity M_{corr} for several values of $\Delta\theta$, a parameter decreasing the adaptation in the Potts model. We find that the behaviour of M_{corr} with respect to L is qualitatively similar to that of the experimental curve for a broad range of $\Delta\theta$ values (see Fig.3.5). For all values of $\Delta\theta$, M_{corr} saturates reaching a maximum that is similar to that of the experimental data, of around 8 items correctly recalled. Exceptions are at the two extremes: too small and too large values lead to

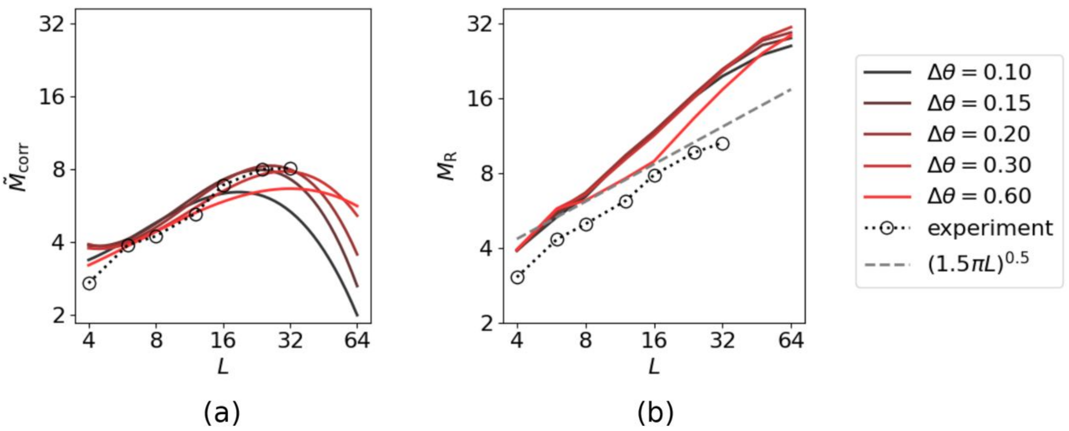


Figure 3.5: Two measures, M_{corr} and M_R, are shown for several values of $\Delta\theta$, coded by colours. Black dotted curves are the experimental results of free recall of locations in a 2-dimensional grid. (a): M_{corr} has a maximum value. It is the number of recalled STM items until the network either revisits one of the already-recalled STM items or visits one of the LTM items, but within a given number of latches $2(L-h(t|L))$, where $h(t|L)$ is the number of correctly recalled STM items up to that point in time. (b): M_R shows a scaling behaviour. M_R is the number of recalled STM items, ignoring repetitions and errors, within a given number of consecutive latches set as again, $2(L - h(t|L))$

lower capacity of the Potts model, below 7 items.

The saturation behaviour, and hence the notion of memory capacity, reflected in the measure M_{corr}, again contrasts with the scaling behaviour approximated by the various measures such as M_R. This contrast holds irrespective of the values of network parameters used in simulations. Indeed the scaling behaviour of M_R is almost independent on the value of $\Delta\theta$ except when it is too large, $\Delta\theta = 0.6$ (Fig. 3.5b).

"Performance" therefore depends very differently on L, if recall is taken to be terminated by errors, i.e. by the recall of an item that is not in STM. Thus, while if ignoring errors the notion of STM capacity appears irrelevant (evident from the scaling behaviour of the various quantities discussed above), it becomes relevant if errors are considered to be critical in the task.

In summary, we have shown in this section that whether we get scaling or saturation in the performance depends on the specific metric we use to measure the performance, both in the Potts model and in our experiment. In free recall experiments, performance has often been quantified through the M_R index, thereby ignoring errors. The scaling behaviour of this index has recently been corroborated for lists up to 512 words [82]. In contrast, taking our experiment as an example, we have shown that if errors are considered critical, in our case through the M_{corr} measure, then the performance of human subjects actually expresses a saturation at about 8 items. In our model, that expresses a similar behaviour, this saturation is brought about by the interference from long-term memories.

3.3 Serial recall of different stimuli

We have found that repetition-limited and duration-limited measures of performance in free recall in the Potts model endowed with short term memory function can express quasi-square-root behaviour in the number of items in the list. One question that naturally arises is whether the same model can express behaviour similar to serial recall, a paradigm very similar to free recall, but with a crucial difference. Here, participants are instructed to recall items in the same order as they have been presented, making the task more difficult.

We have run serial recall experiments with three different types of material. We asked participants to observe and repeat sequences of stimuli presented to them on the screen - either digits or spatial locations on a 2-dimensional grid (Fig. 3.3), and varied the time of presentation of the stimuli in the observed sequence. There were two conditions for the spatial locations, referred to as Locations and Trajectories: in the Locations condition, considered to involve only "discrete" items, the six chosen locations around the centre of the grid were highlighted in any order; while in the Trajectories condition, every next location was one of the six consecutive locations around the previous one, thus suggesting a "continuous" trajectory.

3.3.1 Experimental procedure

The 36 participants (F = 11, age: mean = 30.8, sd = 10.8) were instructed to watch a sequence appear on the computer screen and repeat the sequence just after, by clicking on the screen. They had to repeat sequences of L stimuli (L starting from 3). Contrary to the previous experiment reported in Section 3.2, in this task participants had to recall the material in the correct order, otherwise the trial was dismissed as incorrect. In one of the conditions, 5 trials are available for each length L, with L growing until 3 out of 5 trials are incorrect; the last L before the one with 3 incorrect trials is then called the limit capacity for this participant in this condition. For each participant the sequences were of all three stimulus variants:

- (D) Digits out of 1, 2, 3, 4, 5, 6 on a black screen, presented one at a time;
- (L) Locations on a hexagonal grid (Fig. 3.3) highlighted one by one, out of 6 around the central (blue) dot;
- (T) Trajectories on the same hexagonal grid: now each consecutively highlighted dot is one of 6 neighbors of the previous one (the first one is always one of the six around the center), thus suggesting a “continuous” trajectory.

Each stimulus was presented for one of the time durations (in separate blocks): 400ms, 200ms, 100 ms. First always came the 400 ms training session, then either 200 ms or 100 ms (balanced), and then the remaining duration. Presentation order was balanced across duration and stimulus material.

3.3.2 Variations of the experiment

In earlier versions of the experiment we had also tested the effect of adding spatial cues or *landmarks* to the background of the grid (Fig. 3.3b) on memory capacity for the trajectories condition. No significant effect on the recall performance was observed.

Further, testing the different conditions including the additional presentation times of 500 ms, 300 ms and 50 ms indicated the same consistent trend – recall capacity falling with decreased presentation time for all types of stimuli – so for the final version of the experiment described above we only included three representative presentation times, to make the experiment shorter, and so less tiring for the participants.

3.3.3 Results

One block of Condition x Presentation time was defined as follows: participants started with short sequences of length 3; if they recalled them correctly in at least 3 out of 5 trials, the sequence length increased, until a memory capacity limit for this stimulus type and presentation time was reached. In this way we measure the memory capacity for serial recall, taking as a measure the Area Under the Curve (AUC), often used in the field [94].

Our experiment yields two main results (Fig. 3.6). The first is that the type of stimulus does not affect the recall probability, except for a slight disadvantage in the *discrete* Locations condition, suggesting a universal mechanism for recall independent of the material, which manifests itself at the systems level. The second, which is instead pronounced, is the effect of presentation time per stimulus, that, when shortened, makes it more difficult to correctly remember and repeat the longer sequences, suggesting a disadvantage at the encoding stage. We ask whether latching dynamics in the Potts model can express this finding. Given that our results, as well as those from other studies [95], show very little dependence on stimulus material, hereafter we only consider the result with digits in order to establish a comparison with our model.

Adding a heteroassociative rule to the network dynamics, we find a good agreement between our experimental data and the model (Fig. 3.7). In addition, we find that human subjects perform better if the to-be-memorised digit series include ABA or AA (Figs. 3.7a, 3.7c), in line with the notion that the repetition of an item aids memory [96–99]. Such sequences are not produced by our model, due to firing rate adaptation and inhibition preventing the network from falling back onto the same network state. Due to this, we refer to stimulus sets excluding such sequences as *Potts-compatible.*

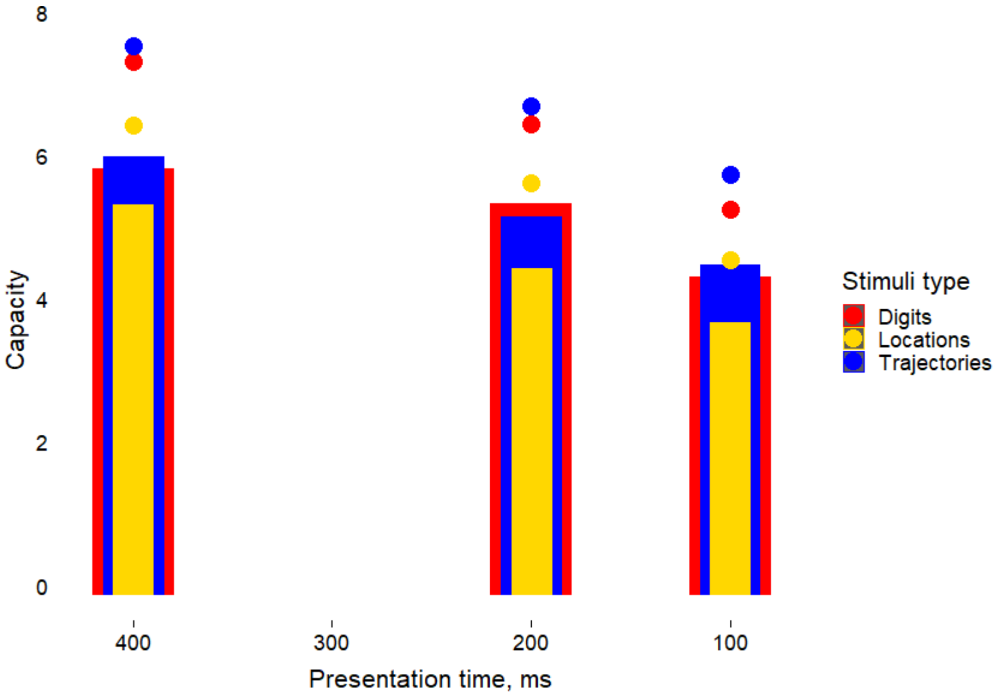


Figure 3.6: Short-term memory capacity for serial recall does not markedly depend on stimuli type. Memory capacity (from the AUC measure) for serially presented stimuli for different presentation times: bars correspond to the average across participants of the longest correctly recalled sequence, while the distance from the bar to the dot of the same colour corresponds to the standard deviation of the mean. We performed the experiment for three different stimulus types, shown in different colours.

The heteroassociative component of the learning rule provides “instructions” to the network regarding the sequential order of recall, allowing it to perform serial recall (this is to be contrasted with the model with a purely autoassociative learning rule, performing free recall). The strength of such instructions is expressed through the parameter λ. We find that this parameter plays a role similar to that of presentation time in our experiments; increasing it enhances performance, just as increasing the presentation time increases the performance of human subjects (Fig. 3.7). However, values of λ that are too large again make performance worse and deteriorate the quality of latching (Fig. 3.7d).

Therefore, the most functional scenario is when the heteroassociative instruction acts as a bias or a perturbation to the spontaneous latching dynamics rather than enforcing strictly guided latching in the Potts model, which could be the Potts network analogue to *fragments of memory*. This is in sharp contrast with the mech-

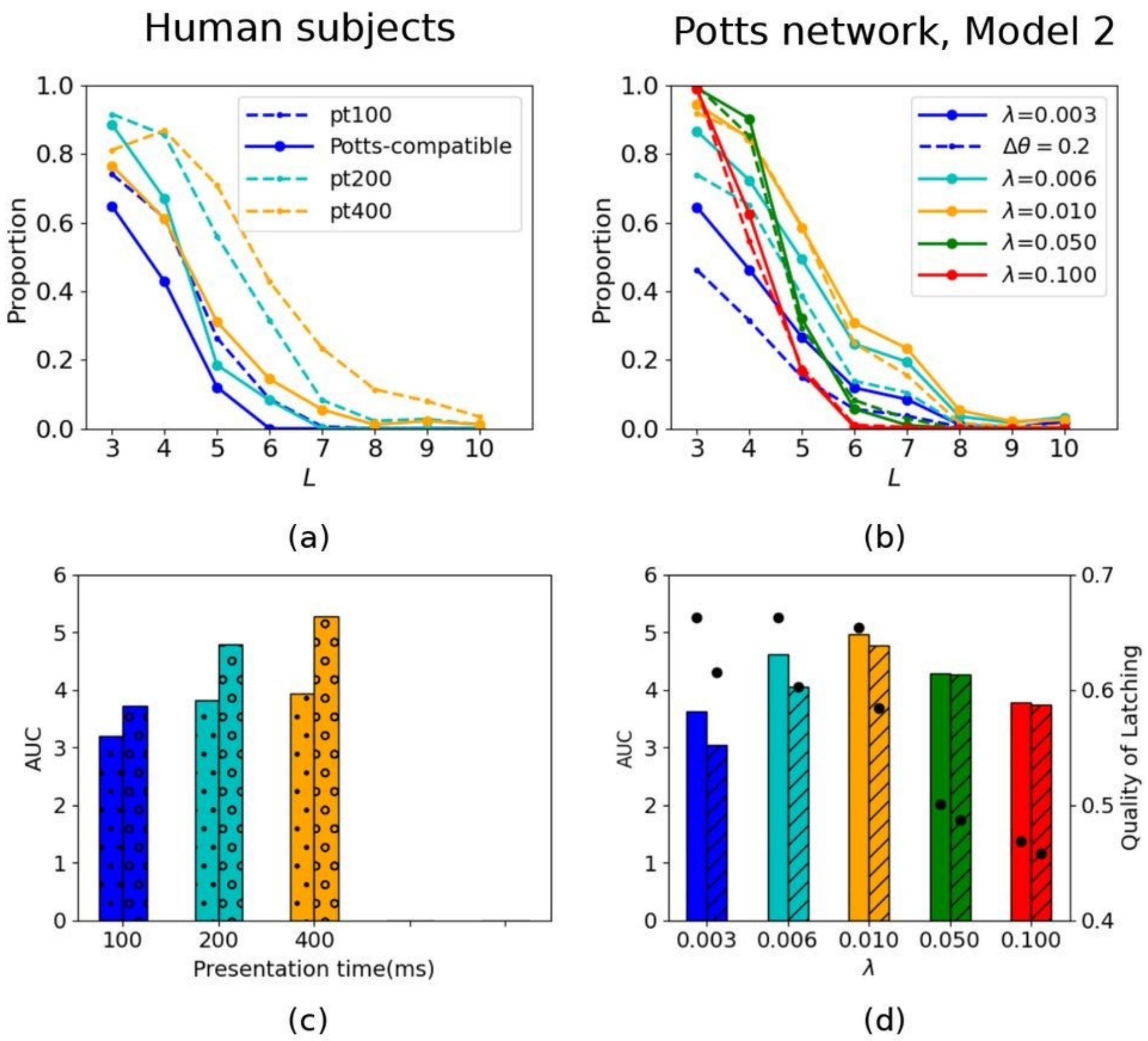


Figure 3.7: (a) Proportion of correct trials in the serial recall task with digits. Data for all subjects ($n = 36$) are pooled together. Colour codes for presentation time in units of milliseconds. Solid and dashed curves for each colour show the result of "Potts-compatible" trials and of all trials, respectively. (b) Proportion of correct subsequences in a latching sequence of the Potts model. Colour codes for values of λ. Solid (dashed) curves are for $\Delta\theta = 0.1(0.2)$. (c) Area Under the Curve (AUC) computed from the curves of (a). Colour-coding is the same as in panel (a) and the bars filled with dots (open circles) correspond to solid (dashed) curves in (a). (d) AUC for latching sequences of the Potts model. Same colour-coding as in (b) is used. Bars without hatches are for solid curves in (b) and those filled with oblique lines are for dashed curves in (b). Black dots indicate the quality of latching on the right y-scale.

anism for sequential retrieval envisaged in the model considered in [100], where the heteroassociative connections are the main and only factor driving the sequential dynamics; in that case, without it, there are no dynamics but rather, at most, the retrieval of only the first item. The effect of lower adaptive threshold (expressed by $\Delta\theta$) on latching sequences is to constrain the dynamics to a subset of presented items among p patterns, but values of $\Delta\theta$ that are too high degrade the performance as well as the quality of latching (Fig. 3.7b, 3.7d). As mentioned above, the Potts model produces latching sequences even without any heteroassociative instructions. This means that the free transition dynamics of the model may or may not coincide with the "instructions" provided by the heteroassociative weights. Then one question naturally arises. How does the congruity between spontaneous, endogenous sequences and instructed ones affect the performance of the model? To see this effect, we obtain some intrinsic sequences by running simulations with $\lambda = 0$; from these sequences, we generate a set of instructions. These instructions are congruous, as they reproduce latching sequences emerging without any heteroassociative

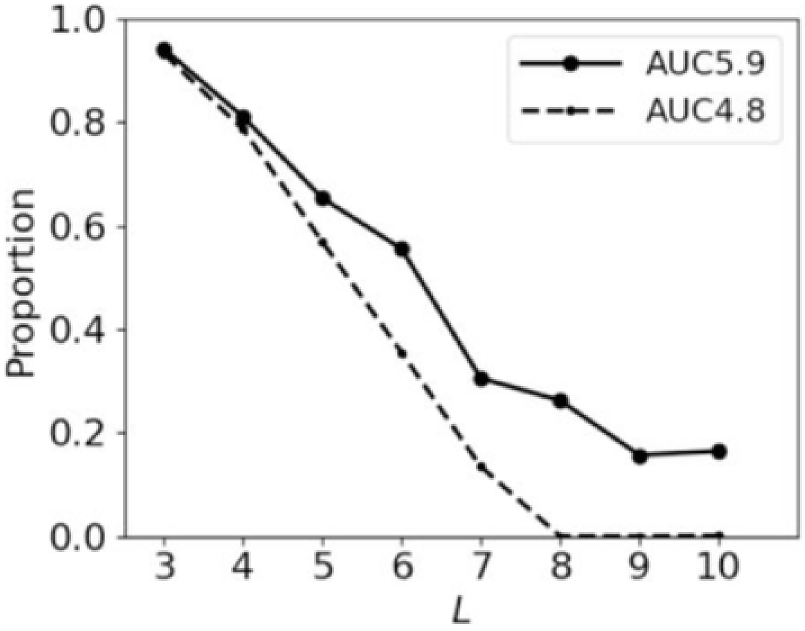


Figure 3.8: Serial recall of digits by human subjects and the Potts model. Proportion of correct subsequences in a latching sequence of the Potts model. The solid curve is for congruent instructions only and the dashed curve is for a shuffled version of intrinsic sequences.

instructions. Then we compare the performance for these congruous instructions with those of incongruous instructions, which we obtain by shuffling the congruous ones. We find that the capacity of the model (denoted as AUC in the legend in Fig. 3.8) increases by as much as 1 item for the congruous case relative to the incongruous case.

These results together with those from the previous section indicate that intrinsic latching dynamics, similar to a random walk, can serve short-term memory (e.g., free recall). Furthermore latching dynamics can also serve serial recall, if supplemented by biases that modify the random walk trajectory; the modification (or perturbation) should be a quantitative one, which biases the random walk character of the trajectories, rather than an all-or-none, or qualitative one, that inhibits it. This is consistent with our experimental result from the next chapter (see Fig. 4.2), where "assisted" serial recall leads to poorer performance than a non-guided control. Whereas, like a congruent bias improves "performance" of the Potts network, an established *schema* trace should be what differentiates human recall dynamics from that of a random walk.

3.4 Unfolding the dynamics of recall

While in a Potts network it is us who tune the parameters, in human behavior it is the data that should reveal a bias in recall, if there should be any. That is why we conducted further analysis on the data from the experiment on free recall.

The spatial recall task has an advantage compared to the word-recall tasks, in which any word or non-word or word fragment could be typed by participants: since an answer is a click on a spatial location, the mistakes are as good as the data as the correct clicks. On Fig. 3.5 we showed that the well established scaling law of immediate memory capacity holds on average if measuring the recall capacity before a mistake. Looking at all the data together, rather than only at the average, gives out a great variability in recall dynamics (Fig. 3.9): a lot of participants make a mistake on their first click, more so in the longer trials, and some recall the whole set of locations in one perfect streak of clicks, or at least for the first m of L STM locations.

This observation motivates further questions: what makes a sequence memorable? How far off are the mistakes? What is there common in the recall sequences of different participants?

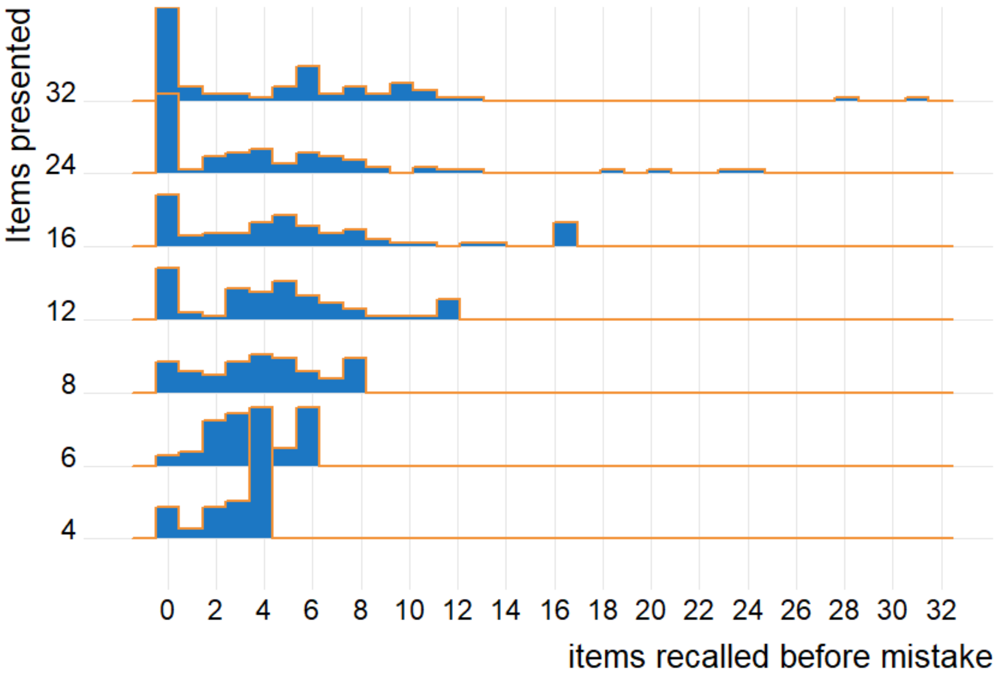


Figure 3.9: First recall sequence across all participants: frequency of the number of items recalled before a mistake for different number of items presented (normalized separately for different list length).

In the experiments described in this chapter the configurations of stimuli were fully randomized, so the apparent characteristics of the recall process did not depend on the exact patterns. And we look at the common trends across the very different patterns. Addressing the questions above, we plot the statistics of individual clicks during recall sequences on different trials (Fig. 3.10, Fig. A.1,A.2), and we observe a number of common properties:

1. on shorter trials, the clicks are close to the correct locations and far from each other: the participants remember well the pattern and follow it if they avoid mistakes;

2. on longer trials, the clicks are only very precise at first, for 3-4 well remembered points;

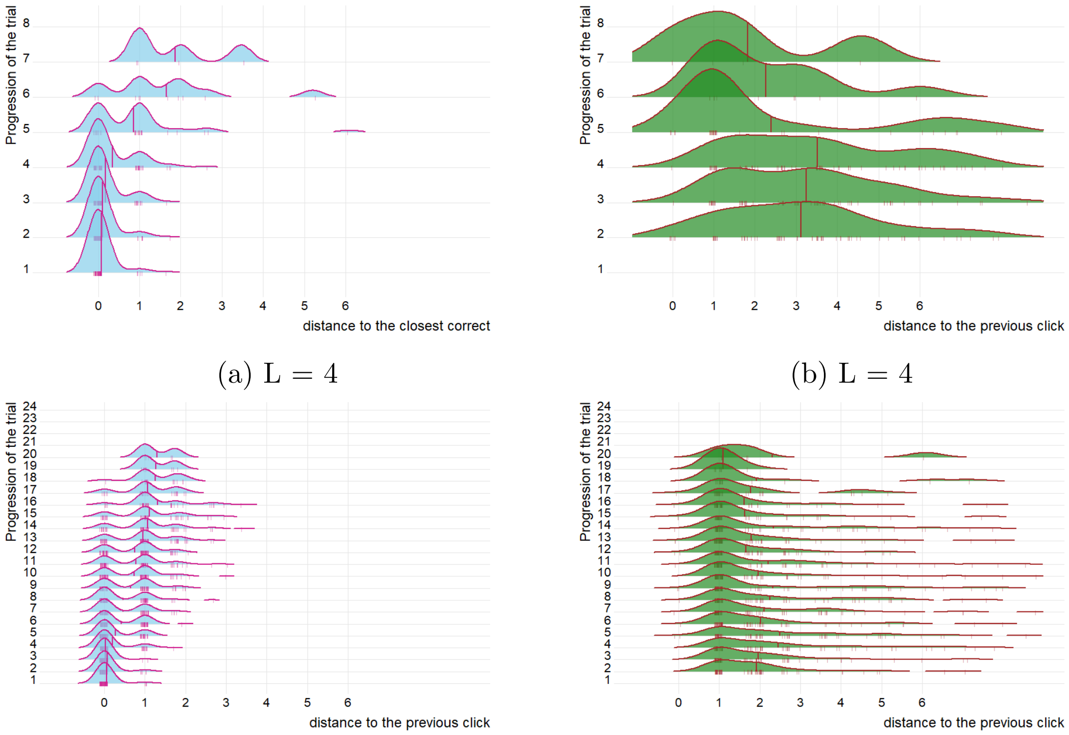
3. after the first correct streak, the clicks are still close to the correct ones, indicating that participants now vaguely remember the overall schema of the shown pattern;

4. on very long trials, after all they recalled correctly, the participants just click everywhere, probably in what one may dub a random foraging behavior.

This combination of schemata – smaller(2) and larger(3) – and foraging motivated our further investigation of schemata in recall, described in the next chapter. In this chapter we introduced a spatial recall task and demonstrated that, quite like a stored map of a randomly explored environment is biased by the learning trajectory in the model of CA3, human recall of random material seems to be a semi-stochastic process, resembling latching of a Potts network, and relying on partial schemata in its partial order.

In the next chapter we will explore this topic further, focusing on how instruction and/or freedom of choice inspire the selective activation of fragmentary memory, whether mnemonic techniques help *connect* the fragments, and whether there are common schemata, or features of fragments that help memory recall.

To address these questions we needed a possibility to add words to the memory board and a compartmentalized setting, where we could define a spatial inaccuracy associated with any spatial error and thus characterize general memory schemata that may be faulty.



(a) L = 4

(b) L = 4

(c) L = 12

(d) L = 12

Figure 3.10: The individual clicks by the participants as the trial progresses (y axis) for two trial types – of length 4 and of length 12. The figures in the left column show the distribution for all participants of distance of the current click to the closest correct. The figures in the right column show distance of the current click to the previous click. Distances are normalized to grid units (as in Fig. 3.3). Vertical lines show the average value, while the short vertical dashes show individual points (with jitter). Note: there was a maximum of $2L$ clicks available for each trial of length L, but the trial ended sooner, e.g. if the participant recalled all the correct locations before reaching $2L$. For other length of trials see Appendix (Fig. A.1, A.2).

Chapter 4

Common fragments: a closer look at recall dynamics in a compartmentalized environment

In order to add a linguistic dimension to the spatial memory experiments, we have designed a beehive-like grid shown on Fig. 4.1. This setting was developed initially for the experiment on mind wandering, described in chapter 5. The board consists of 18 6-petal *flakes* of identical cells, which allows for:

- multiple symmetries that help balance the design,
- recording the general area of recall (or semantic association in case of using words).

4.1 Instructions and fragmentary recall

First, we aimed to address two of the questions arising from the previous experimental results: to what extent do explicit instructions interfere with fragment activation and where does the limit on recall capacity arise from?

For this aim we designed an experiment on spatial memory, where only the presentation mode varied between the conditions – serial in the first and third or

Condition	A	B	C
	Free Recall	Serial Recall	Assisted Serial Recall
Presentation mode	Simultaneous	Serial	Serial
Presentation time	$\log_2 L$ s	315 ms per stimulus	315 ms per stimulus
Allowed clicks	$2[L - h(t)]$	$2[L - h(t)]$	L

Table 4.1: The three experimental conditions of experiment 1. Setting described for a trial with L stimuli.

simultaneous in the second – while in the third condition the participants were cued as to where was positioned the next item they were asked to retrieve from memory.

Unlike in the experiment on serial recall in Chapter 3, in this experiment we did not terminate each trial as a function of the number of mistakes, and the locations for recall were randomly distributed, like in the simultaneous presentation condition of Section 3.2. Because of no structure now guiding the encoding during the serial presentation, but also no penalty for order mistakes, we expected to see similar results for serial recall as in the previous experiment, *Trajectory* condition (Fig. 3.6). The *assisted* recall condition consisted in giving the participant the order of recall areas, thus leaving them only to remember the relative *petal* positions, which with probability 1/6 match the *Locations* condition of experiment of Section 3.3.

Our hypothesis was that in the simultaneous presentation condition, see Table 4.1 and Fig. 4.1, a participant can pick up on a set of fragments, e.g., a sequence of two-three positions in a particular relation to one another, for example two next to each other, or three in a straight line, they would memorize more easily, associating them with a simple schema in long-term memory, e.g., "three points on a straight line" and that would be reflected in a higher number of recalled items, growing sublinearly with the number of items on the list, as in [80–82] and Experiment 1 of Chapter 3. Limited recall capacity for serially presented items should then remain low for any number of items presented.

4.1.1 Experimental procedure

We tested 90 participants from *www.prolific.co*, 30 in each of the three experimental conditions of Table 4.1.

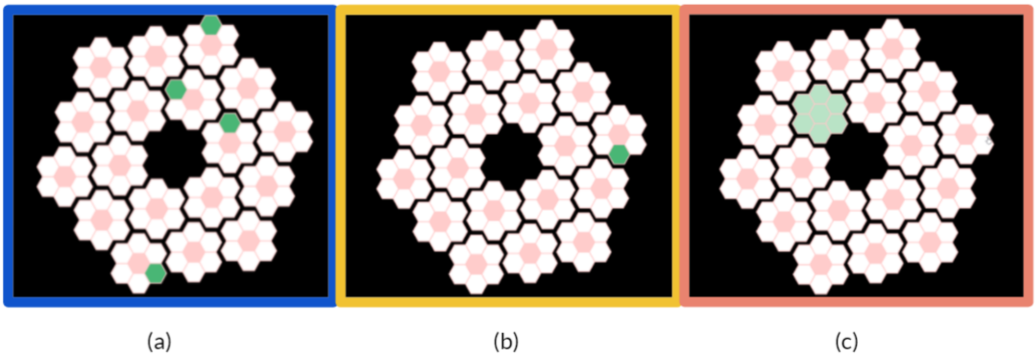


Figure 4.1: The spatial arrangement of the hexagonal cells used in the experiments. We refer to the arrangement as a *snowflake*, each separate six-*petal* unit of which we call a *flake*. The spatial positions to be memorized were shown in green, either simultaneously (a, condition A) or serially (b, conditions B and C). In condition C, at the recall phase, the correct flake is highlighted in light green and the participant has to choose one of the six petals for each flake in the order of presentation.

The general setting

In all the tasks of Section 4.1, the participants were instructed to watch subsets of L out of 108 hexagonal cells (petals) turn green on the screen and reproduce as many as they could recall by clicking on the corresponding cells. The sets of stimuli of sizes L = {4, 6, 8, 12, 16, 24, 32} were presented in growing order, 5 trials per length.

There were 3 separate experimental settings, as shown in Table 4.1:

- In condition A, the Free Recall experimental condition, the sets of L cells were highlighted simultaneously for $\log_2 L$ seconds. Participants were instructed to recall freely as many locations as possible. While the participants were recalling the cells by clicking on them, they had $2[L - h(t)]$ available clicks. $h(t)$ was the number of correctly retrieved cells by the time t. This way we could observe both how many items people recall overall and how many they could recall in the first L clicks.

- In condition B, the Serial Recall condition, the L cells were highlighted one by one, each for 315 ms (the mean over L of $\frac{\log_2 L}{L}$). The participants were instructed to recall as many locations as they could in the same order as they were presented. While the participants were recalling the cells by clicking on them, they again had $2[L - h(t)]$ available clicks.

- In condition C, the Assisted Serial Recall condition, the sets of L cells were highlighted one by one, each for 315 ms (the mean over L of $\frac{\log_2 L}{L}$). The participants were instructed to recall as many locations as they could in the same order as they were presented. Each flake (the separate six-petal hexagonal flower) was highlighted in light green and the participant had to choose one of the six petals. The participants only had L clicks for recall in this condition.

We excluded 4 participants (2 in condition A, and 1 in each of B and C) for doing the task 'too well' (suspected cheating) - on at least 20/35 trials they exceeded mean+2sd of raw correct responses (the overall number of correct clicks before chance correction). In Section 3.2 we explain the motivation for this exclusion criterion.

4.1.2 Results

Participants. The population sample after exclusion consisted of 86 participants (gender: F = 36, age: mean = 27.08, sd = 9.19) recruited through Prolific. Since this was an exploratory study, we aimed first at having 30 participants in each experimental condition, so that the results could be comparable to the results of the experiments of Chapter 3.

On Fig. 4.2 we show the average recall performance per condition for different measures of correct recall sequence and the theoretical prediction for recall of words from [82].

For all of the experimental conditions we corrected the recall measures by the differing chance level in each condition. In condition C, we chose to only include the sequences and answers that were reported in the different flakes, in order to be able to estimate the chance level with simple calculations.

Free recall performance roughly follows the theoretical $\propto \sqrt{L}$ trend in simultaneous presentation condition (both measures), with a lower slope if stopped at L items

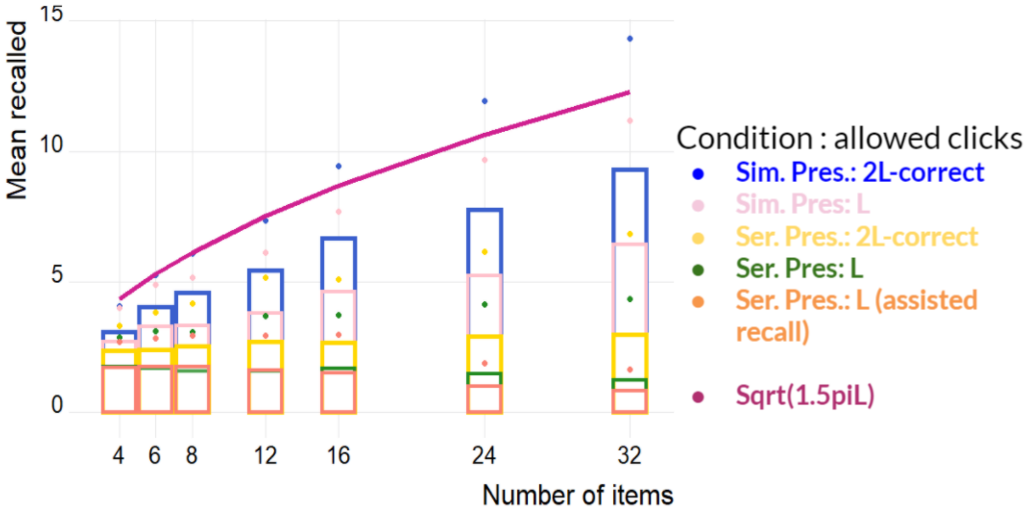


Figure 4.2: Mean number of items correctly recalled for different number of L presented spatial locations. Color marks the different conditions and different measures of recall: blue and light pink show the results in condition A; yellow and green for condition B; orange for the condition C (only one measure of correct recall applied). The dark pink line depicts the theoretical prediction from [82].

recalled. Allowing space for mistakes, given with the $2[L - h(t)]$ clicks allocated for recall of a sequence of length L, allows for additional correct recall that may come from the 'foraging' for correct answers in longer trials.

For serially presented items the theoretical prediction does not hold. As already shown in the experiments of Chapter 3, and in line with the existing literature, serial recall is worse in performance than free recall and imposed order yields limited recall capacity [95]. The present experiment extends and qualifies the results presented in Chapter 3, in that it indicates that allowing for mistakes, to a smaller or larger extent depending on the measure, is not sufficient to recover the $\propto \sqrt{L}$ trend – suggesting that serial encoding has, at least in this spatial paradigm, a major detrimental effect *per se*, independently of whether the recall attempt is terminated by mistakes.

Moreover, the assisted order of recall turned out to be the most difficult condition for correct recall. This condition imposed an order in both presentation and recall, and, irrespective of that they could 'see' in the spatial array of locations to click, a participant could only rely on an artificially imposed schema for successful recall.

This result is consistent with the hypothesis that a major contribution to free recall comes from the activation of LTM schemata, which are specific to each participant; they are more difficult to activate with serial presentation, and are made irrelevant anyway by imposing recall order in the ‘assisted’ condition. Comparing with our observations on Potts model latching, reported in detail in a forthcoming publication (not in this Thesis), we note that both in Potts network simulations and in human behavior a congruent instruction could improve the memorability of a sequence, while arbitrarily imposed instructions can only make recall more difficult. While with the Potts model we implemented ‘congruent instructions’ by first observing the latching sequences spontaneously produced by a particular network, an interesting challenge is now how to operationally define instructions, in terms of serial order, which are congruent with the specific schemata of a particular participant.

One strong strategy for congruent instructions is using mnemonic techniques. For example, the method of *loci* suggests using navigation among (non-spatial) items, having first anchored them to spatial locations. In this case, the instructions are internally driven, they are in effect self-instructions, congruent with a fragment of memory well-learnt already: usually, a trajectory through one’s own house or their route to work. It is natural to hypothesize that activating well encoded LTM fragments helps learning additional information, and we thus wondered what is the special role of spatial memory in it.

4.2 Mnemonic techniques and free recall

In their recent study [101], our colleagues have shown, in collaboration with mentalist Vanni De Luca, that just two hours of a lecture on mnemonic techniques can have a major impact on the ability of participants to memorize and correctly recall lists of 16 items, in 5 different tests with material varying from digits to images. The main mnemonic technique used in the training was the well-known Cicero method, or method of *loci*, where one learns to associate objects with spatial locations on a familiar route. It is natural to assume that memorizing the spatial positions of food

Condition	A1	A2	B1	B2
Presentation mode	Simultaneous	Simultaneous	Serial	Serial
Presentation time	$4 \log_2 L$ s	$4 \log_2 L$ s	315*3 ms/stim.	315*3 ms/stim.
Allowed clicks	L	L	L	L
Petal layout	Random	Fixed	Random	Fixed
Presentation order	–	–	2 orders	2 orders

Table 4.2: The four experimental conditions of experiment 2. Setting described for a trial with L stimuli.

and of safe spaces when foraging was a primary function in the evolution of memory systems, across mammalian species and perhaps beyond.

Using our experimental setting, we decided to test whether anchoring words to spatial positions increases the memory capacity in immediate free recall, too.

4.2.1 Experimental procedure

For this experiment we took the same snowflake board as in the previous section. Again, we highlighted subsets of hexagonal cells in green, but now words also appeared on them. We took the words from the word lists that have beeb used in the mind wandering experiment described in chapter 5. From the design of that experiment we also took the six possible arrangements of the words on the board to balance out the relative spatial positions of words between participants. We will explain the choice of word lists and their arrangement a couple of sections below.

In these experiments, participants were asked again to click on remembered locations, but now on each click they were also requested to type a word corresponding to this location.

Sample size definition

In these experiments the sets of words on cells were of sizes L2 = {4, 6, 8, 12, 16} and the number of trials of these sizes were respectively, $T_L = \{4, 3, 2, 1, 1\}$. This way we could define 6 conditions to balance combinations of words and their arrangements on the board across participants; for each participant not one cell, and so not one word, was repeated twice. Given 6 arrangements of the words on the board and 6 petals in each flake, we have 36 variants of the task, here for each condition we tested 18 variants with one participant per variant (so, 18 participants); while in the serial task we tested 9 out of 18 variants from free recall in 2 orders, resulting in overall 18 participants in the serial presentation condition, too – 9 seeing the location-word pairings in one order, and 9 in another.

In order to counter balance any possible direct associations between words, the 2 orders for serial presentation were defined so that no pair of words was used consecutively in both and the relative order 1 of all words is inverse with respect to the order 2. This is achieved by constructing order 2 from order 1 by first taking the even positions of order 1 in reverse order and then the odd positions. For example, if order 1 was ABCDEF, order 2 would be FDBECA, and two separate experimental groups did exactly the same task, but with different presentation orders.

Thus under each of the 4 experimental conditions we tested 18 participants, resulting in sample size of 72 participants (gender: females = 37, age: mean = 27.36, sd = 7.64) recruited through Prolific.

Choice of stimuli: locations and words

As mentioned earlier, in these experiments we fixed the spatial configurations of stimuli so that none of the locations were used twice throughout a testing session. This was motivated by the way the words were arranged on the board.

The word lists used in this experiments were initially chosen for the mind wandering experiment, and the rationale behind this choice is explained more in detail in Chapter 5. The current experiment was partly motivated by the design of the former: we wanted to test whether the spatial arrangement of words will affect their recall with and without additional associative schemata introduced to participants of the mind wandering experiment. In the current section we will only describe the initial attempt to find the space-word associations within the stimuli set shared between the two experimental settings.

The word lists were taken from Deese/Roediger– McDermott (DRM) [102, 103] lists for Italian from [104]. These are lists of words reported as a first association to a given word. For example, for a word "Trash" in the original DRM lists for English language people most often named "garbage", "waste", "can" etc. It has been shown that if another group of people are presented with the list "garbage", "waste", "can"..., they falsely recall having heard the *lure* word "Trash" with a probability of 49%, and falsely recognize the lure with a probability of 78% [105, 106].

In the experiment of this section we use words from the list not including the lure. Each of the flake has six petals, so a flake can accommodate six semantically associated words. In one of the six subconditions described above, a participant would get words "garbage" and "tea" and another participant, given a different subcondition, gets "waste" and "mug", thus between these participants we can account for somehow parallel, although distinct association that could affect recall and/or mind wandering. We aim to use this paradigm for further exploration along these axes, but for now we treat all the words as independent stimuli. Note that because no two petals of the same flake are one a trial, a participant can never have two highly associated words within a trial.

Experimental conditions

The 4 experimental conditions are summarized in Table 4.2. Additionally:

- In the experimental condition A the sets of L cells were highlighted simultaneously for $4 * \log_2 L$ seconds. (4 times as much as for the purely spatial task, simultaneous presentation).

  Now we also wanted to test the different resolution levels in the visual spatial memory, and:

  - in version A1 random locations on the screen were presented. The participants had to remember both the flake and the petal for each word.
  - in version A2, all the cells that were highlighted in one trial belonged to

the same flakes as in the version A1 (administered to other participants), but they were also on petals of the same relative petals of each flake(see Fig. 4.3 for a visual explanation).

- In experiments B1 and B2, each cell with a word was shown for 315*3 ms. (again, in accordance with the spatial task, serial presentation).

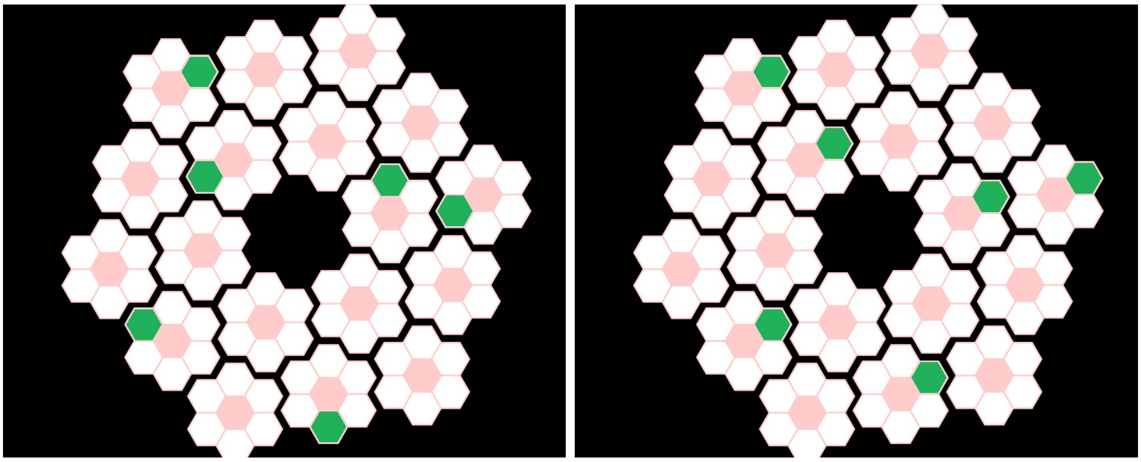


Figure 4.3: Two subconditions of stimulus presentation (in green): random petal (left) and fixed petal (right).

The typed words were counted as correct if they had less than two typos – insertions, deletions and substitutions – with respect to the form of the word presented.

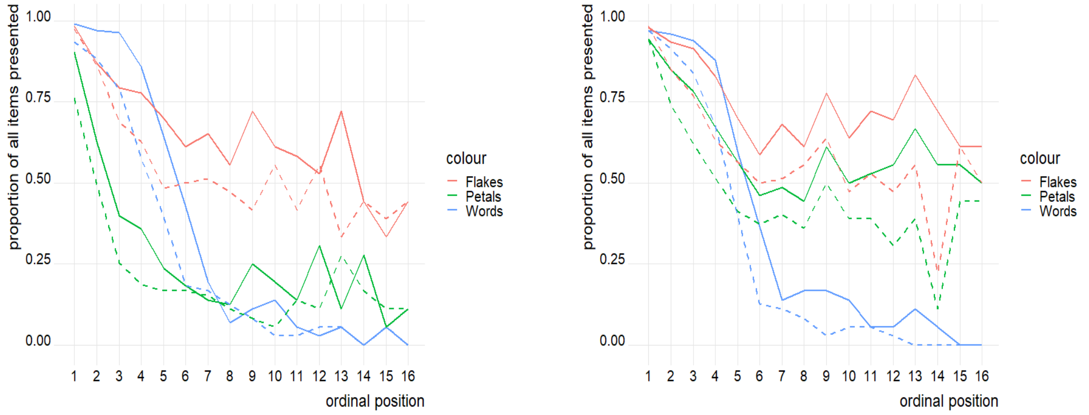
After each of the word experiments the participants also did the Free Recall version of the purely spatial task – this way we made sure they matched the average performance of the other participants.

4.2.2 Results

In this experiment we could directly compare the participants recall capacity in the number of retained locations, flakes (the regions of the locations) and words.

From the results of recall capacity under different conditions (Fig. 4.4) we observe that:

- again, simultaneous presentation is beneficial to recall capacity;



(a) Random petal

(b) Fixed petal

Figure 4.4: Average recall capacity as proportion of items recalled over number presented along the recall sequence for the two conditions: random petal (a) and fixed petal (b). Solid lines show the results of recall after simultaneous presentation of stimuli and the dashed lines – after serial presentation.

- the participants obviously remember Flakes better than Petals, under all conditions, but the difference is minor when the petals are in a fixed position;
- the memory capacity for words is effectively limited to 4-6 items in all conditions.

These findings should be complemented by testing participants on the recall of only the words, but at this stage we have no support for the idea that linking words to locations help memory in immediate recall task, and we are inclined to think that, at least in this paradigm, the spatial anchoring *per se* is not all there is to the mnemonic techniques. Indeed, a very recent study [107] suggests that using an Australian Aboriginal mnemonic method that resembles the method of loci, but relies even more on an episodic context given to anchor the memory item, improves recall even more than the method of loci. We are then led to presume that successful implementation of a memory technique relies on internalizing and consciously binding the recently acquired material to the existing episodic memory, forming an enriched fragmentary memory, rather than just on separating the words in space.

The result on the Flakes vs. Petals difference in recall performance points toward the memorization of the general schema as a main strategy for successful recall, more so when the order of encoding is decided by the participant, then when imposed by the presentation modality.

4.3 Common biases

For a more fine-grained analysis of the dynamics of recall, we ask: are there common dynamics of recall across participants? Is the random walk random? Or are there spatial schemata we all tend to use to help us memorize?

To address these questions we took *one* layout of spatial arrangements from the previous experiments, consisting of 10 trials of varying sequence length, and asked participants to recall freely, again, by clicking on the positions on the screen. When we look at the sequences of correct choices, we see that among all possible orders for a sequence of length M, that is $A(M) = M!$, participants pick a very limited number of orders (Fig. 4.5).

On sample size. In this experiment we tested 37 participants (gender:male = 22, age not recorded) from Prolific. This has been an exploratory study and we will test more participants to validate the models proposed in the following sections. We will have to calculate needed sample size depending on procedures of validation.

In some trials, over 80% of the correct answers were given in the exact same spatial order. Analysing these orders offers a number of insights, including:

- On average, participants' trajectories are shorter than random ones;
- When there is a way to approximately do it, participants prefer to memorize the locations along a circular path, clockwise or counterclockwise;
- Otherwise, there is a common tendency to go in a straight or quasi-straight line, and proceed from left to right.

So, generally, there are common schemata shared among people, and they seem to be activated when applicable. We hypothesize that these schemata are the learned attractor states that help (and bias) the recall dynamics.

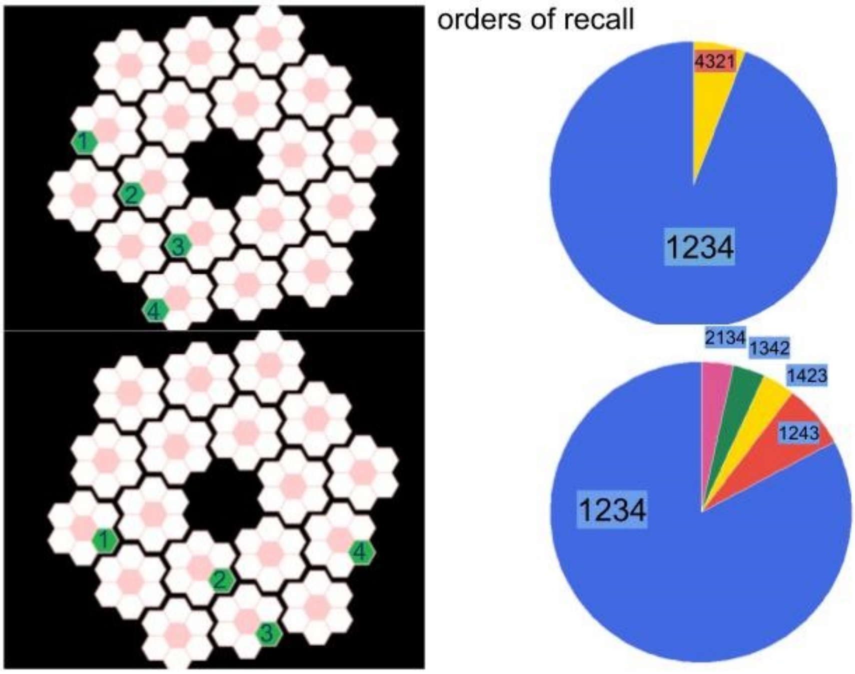


Figure 4.5: Two examples of layouts of 4 points given to participants in the experiment (left column). The highlighted green cells were shown empty and the order numbers show the most common recall order across participants. Right column: the frequency of choosing the order of recall across participants (there were 4!=24 possible orders of 4 points).

To test this point with the spatial setup data we ran a Monte Carlo-like simulation procedure that is aimed at approximating human behavior by adding schema-like biases S_i imposed on the probability of an order ν being picked from the pool of all possible orders, by applying schema-dictated measures M_i:

$$P(\nu) = \exp(-\sum_i \beta_i (M_i(\nu) - M_{i,0})^2) \tag{4.1}$$

Here $M_{i,0}$ denotes the bias-favored order.

From the experiments in chapters 3 and 4 we learned that when the sequences to remember are short, participants can recall them very well, and the first 3-4 items of all trials are recalled more accurately. Together with the notion of chunking [84], which suggests that when dealing with longer sequences participants tend to deal with them by first grouping or chunking them into fragments, this motivated us to look at sliding windows of sequences of recalled locations of length 3 and 4.

Now, for the recall data and for a simulated pool of randomly generated recall sequences we calculate the statistical measures that we drew from the data on the subsequences of length K = 3 and K = 4:

- total length $M_1 = \sum_t (\vec{x_t} - \vec{x_{t-1}})^2$
- absolute angle change $M_2 = \frac{|\theta_{12} - \theta_{23}|}{\pi}$ (for length 3) and $M_2\prime = \sum_t \frac{|\theta_{t-1} - \theta_t|}{\pi}$
- total angle change $M_3 = \frac{\theta_{12} - \theta_{23}}{\pi}$ (for length 3) and $M_3\prime = \frac{\theta_2 - \theta_0}{2\pi}$,

where $\vec{x_t}$ is the location chosen on step t and θ_{ij} is the angle between the line connecting locations $\vec{x_i}$ and $\vec{x_j}$ and the *x-axis*, denoted by θ_t in the case of quadruplets.

β_i was calculated for the different subsequences of length 3 and 4 for each measure M_i to match the simulated mean to the average bias on that measure in the data $\bar{M_i}$:

$$\beta_i = \frac{M_i^0 - \bar{M_i}}{\sigma_i^2},$$

where σ_i is standard deviation of the means of the 1000 simulations in measure M_i.

Adding the biases we observed – the shorter length of the trajectory, the circularity or the minimal angle change – to the probability of picking a trajectory through the remembered positions, we chose 1000 distributions of sequences. We show these distributions for the four trials of length 4 together on Fig. 4.6. For contrast we show the distributions of entropy of random orders (in grey): if participants recalled spatial positions in an unorganized manner, the red line (entropy of participants responses) would be within the grey histogram. If we constrain the probability of choosing a long trajectory to be low with eq. 4.1 when picking orders from a random sample, the entropy of the resulting sample (pink) is lower on average, approaching the value for the data in trials 1 and 2 (rows 1-2), but overshooting for trials 3 and 4 (rows 3-4). Similar trend can be seen when we add a constraint on total angle change (resulting distribution shown in green).

We thus concluded that accounting for biases can give some approximation of the human behavior, but in the present form it does not reflect its variability. This indicates that while such biases seem to affect participants' recall dynamics across trials, they affect it in different ways than how we attempted to model it here.

While in the approach expressed by eq. 4.1 we seek to optimize choosing order with respect to common biases *altogether*, it may be that the memory mechanism behind successful recall is in fact biased selectively. What we are suggesting is that there may be common schemata that act as biases with respect to the statistical measures described above, and when these schemata are partially present in a configuration of dots, participants pick up on them. For example, in trial 1 (top) on Fig. 4.5 one can see three dots in a straight line – and there minimizing the absolute angle change should only be applied partially; while it could be (although not seen in the data) that some participants would have a schema corresponding to ">" and they would start the recall with this pattern. To account for this computationally we need to reward similarity to a common schema instead of penalizing the distance to it as in eq. 4.1. If the reward or penalty were linear, the two approaches would be equivalent but as in eq. 4.1 the penalty is quadratic in the distance, and the reward quadratic in the similarity, they are not.

For that reason we next assumed that the actual schemata are independent of optimization rules and they are in fact traces of experience that *attract* the patterns close to them. For a visual explanation on the difference between the two approaches see Fig. 4.7.

4.4 Attractive traces of experience

Let us assume that all the different traces described by measures M_i are stored in a memory like patterns of activity in an associative network described in Chapter 1. We want to show that, presented with a layout of spatial locations, the network may associate its subconfiguration (in our case, a subsequence of length K=3 or K=4) to a previously stored trace or traces that as attractors bias the recall.

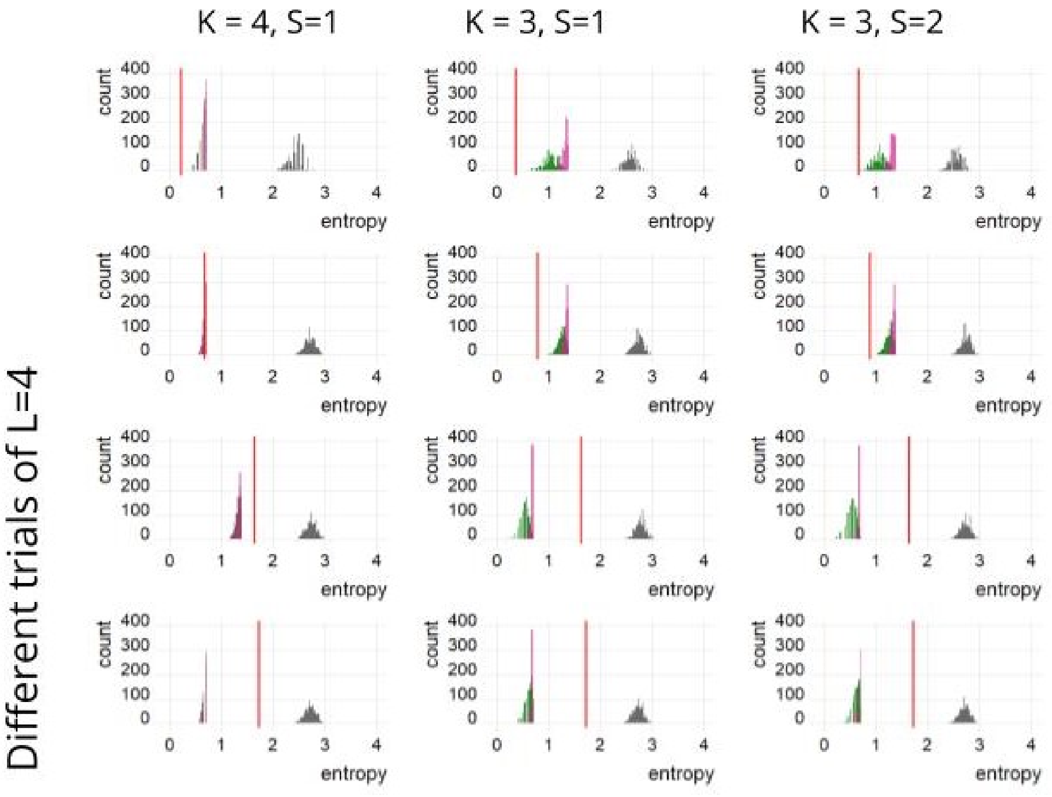


Figure 4.6: Distributions of entropy calculated for the samples of recall sequences for different trials of length 4 (rows). The red line marks the entropy of the experiment sample. In grey, the histogram of entropy for the randomly sampled orders of recall; in pink – for the orders with total length of trajectory constrained by eq. 4.1; in green – for the orders with the length of trajectory and total angle change constrained be eq. 4.1. K=3 and K=4 are the different lengths of subsequences of recall starting at S.

We consider all the measures M_i described above, and add a few more: shift along x-axis, shift along y-axis, change in shift along each of the axis and absolute angle change.

For each of these measures M_i, we calculate its normalized version $\mu_i(\nu_{KS})$:

$$\mu_i(\nu_{KS}) = \frac{M_i - \min_\nu(M_i)}{\max_\nu(M_i) - \min_\nu(M_i)}$$

(to reach the attractive end for each of the measures, we take $\mu_i = 1 - \mu_i$ if $\bar{\mu_i}) > 0.5$, where $\bar{\mu_i}$ is the mean of orders ν_{KS} of $\mu_i(\nu_{KS})$.).

Next we define the associative term (equivalent to a 'coupling' strength) for all

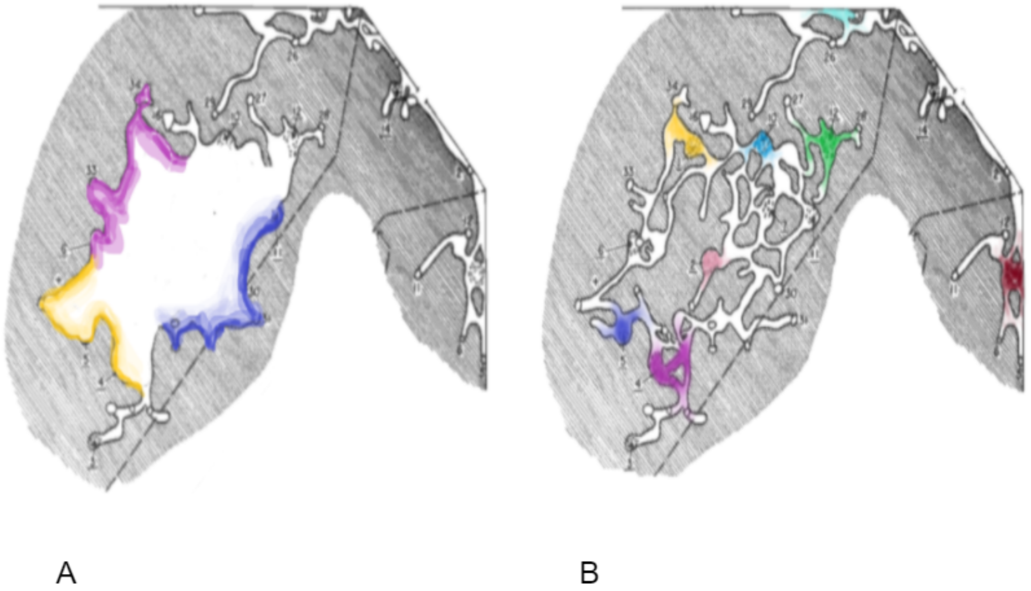


Figure 4.7: The two approaches to understanding biases in recall: A - the first approach: here the quadratic distance to an external bias is penalized, leaving space for all the variety of choices in between, not representing the real complexity of memory space. Here we have tuned the parameter β to adjust it to represent the outer borders. B - the second approach: now we reward similarity to internal memory schemata that are inherent to all participants and independent from each other, so only leave space in the transitions in between. Here instead we look for β as a measure of adjustment to common biases.

the pairs of measures μ_i for each of the subsequences of length K starting at S, which is summed over the subsequences ν_{KS} chosen by participants:

$$J_{ii'}^{KS} = \sum_{\nu} (\mu_i(\nu_{KS}) - \bar{\mu_i})(\mu_{i'}(\nu_{KS}) - \bar{\mu_{i'}})$$

Let us note that all the calculations are done separately for each K and S, so for now we will drop the index KS for the ease of reading.

We perform cross-validation within the sample, leaving each time a participant's subsequence $\nu(p)$ out when calculating $J_{ii'(-p)}$ and computing the probability that the network should pick the suborder $\nu(p)$:

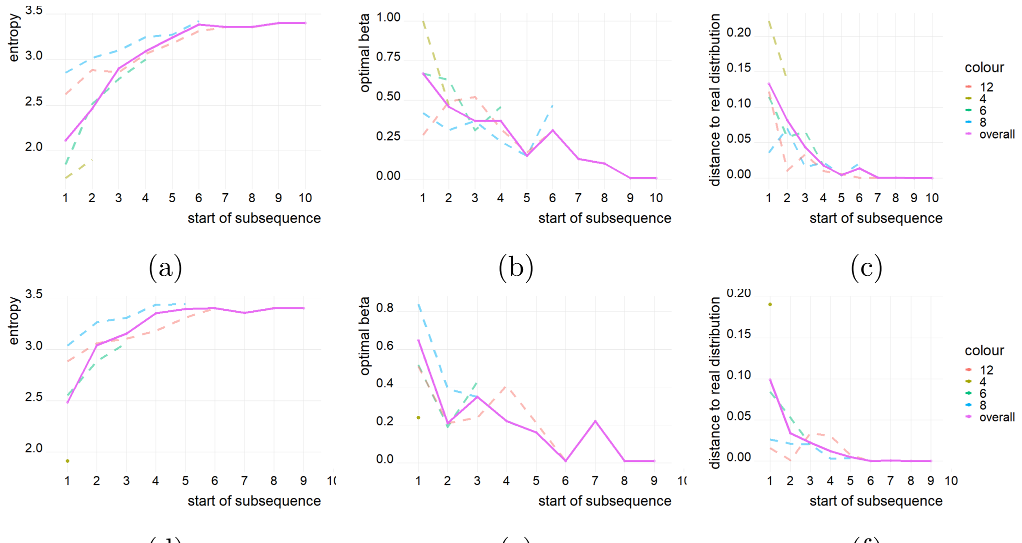
$$P(\nu(p)) = \exp -\beta[\mu(\nu) J_{ii'(-p)} \mu^T(\nu)].$$

The reason why we use cross-validation now and not comparing the experimental sample to all possible orders of recall, is because we wanted to include mistakes in recall, too. Notably, even incorrect answers given by participants shared a lot in common.

Next question was to find a β such that the entropy of the distribution $P(\nu(p))$ across participants approaches the entropy of the distribution of the data. On Fig. 4.8 we show the results of the search, obtained by minimizing the square error.

Note that for different trial length there is a different number of starting subsequences and there was only one trial of length 12 by design, so naturally the optimal β for last positions in recall coincides with the overall optimal, hence there is no error of prediction in the final steps.

Looking at Fig. 4.8, let us reiterate that the entropy of participants' choices grows with recall progression. Furthermore, Fig. 4.8 gives quantifiable evidence to the feature of recall that has been following us throughout this thesis: optimal *beta* decreases throughout recall, meaning that the fragmentary schemata that may help recall are indeed helpful but are only activated near the beginning of the process, leaving us 'foraging' towards the end of a trial (as hypothesized in Section 3.4).



(a) (b) (c)

(d) (e) (f)

Figure 4.8: Modeling results for subsequences of length $K = 3$ (top row) and $K = 4$ (bottom row). (a), (d) – entropy of the subsequence distributions in the data as a sliding window along the recall sequence, calculated for different trial lengths (color) and overall (violet), for each starting position. (b), (e) – optimal β calculated for starting positions separately, for different trial lengths (color) and overall (violet), (c), (f) - normalized distance from the prediction with optimal β to the data, taken as the overall optimal for each starting position (violet on the plot (b), (e)).

We wondered nevertheless whether there could also be any general schemata that may guide us throughout the trial, although with a blur.

4.5 Fragments in recall

To test this idea, we calculated the occurrence of the participants' recall sequences in steps, that is, individual transitions between two locations, and looked at the distributions of these transitions. This way we could trace the overall schemata that guide the encoding and recall, testing the idea motivated by the results of the experiment of Chapter 3, Fig. 3.10 and of experiment 2 of this chapter, Fig. 4.4: people often remember the general regions of the locations and the transitions between them as one fragment, so also the close misses can be informative.

An example of common transitions on a trial is drawn on Fig. 4.9: we show all the transitions (simulated from the probability distribution of the data), that take place from one of the two most commonly chosen locations as first in the sequence

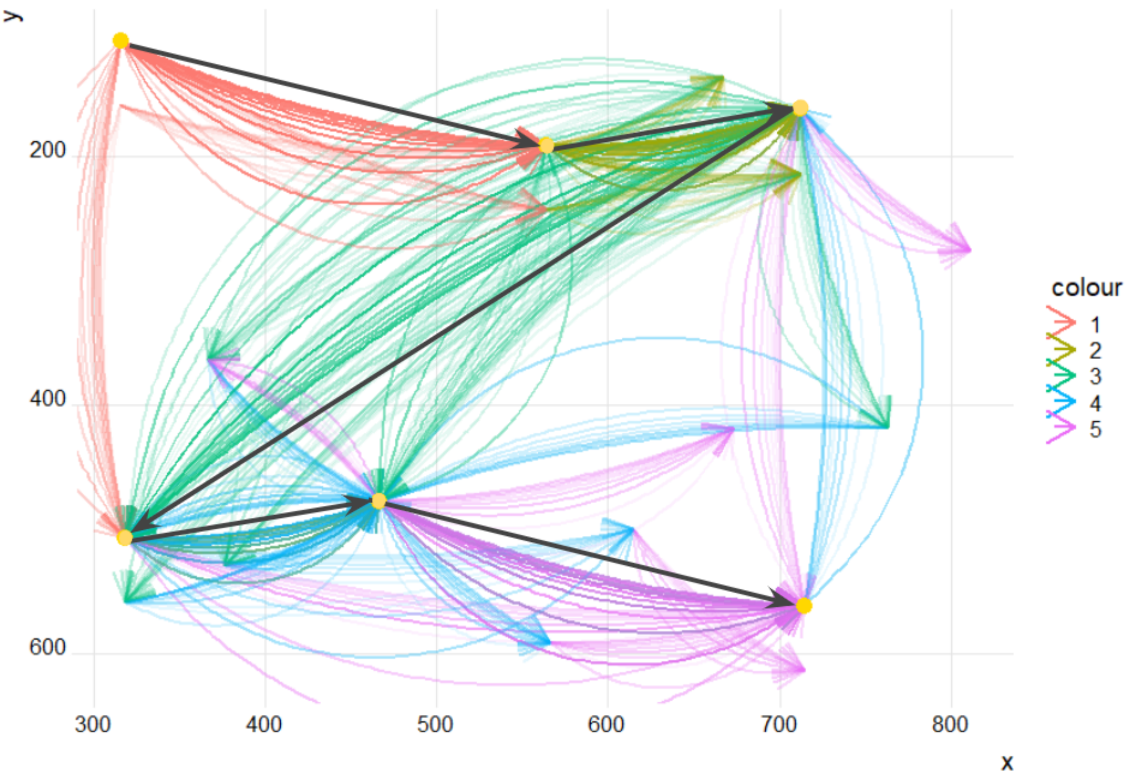


Figure 4.9: Common schemata of recall: 300 simulated trajectories from the training set of the data in a colored order for a trial with 6 positions to be recalled (yellow circles). The first choice is fixed and taken as the most commonly chosen first location to be recalled (top left yellow circle). Most common (correct) transitions are marked with the straight grey arrows.

of recall of 6 locations. We note the 2 common trends of recall orders at the first transition (in red). The main observation here however is in the common *schemata* of transitions, that hit close to the correct items and continue in parallel to the other participants' recall sequences.

4.5.1 Do schema fragments help correct recall?

The core issue that this thesis has been building up for is what is the interaction between LTM memory fragments and successful recall. To approach this question, we take inspiration in the observations from the previous sections.

We calculated the number of subsequences of length 3 for each trial, wherever in the recall sequence they start, and group these orders of choice by proximity to the most common answers (see Fig. 4.10 for a visual explanation). Because of the way

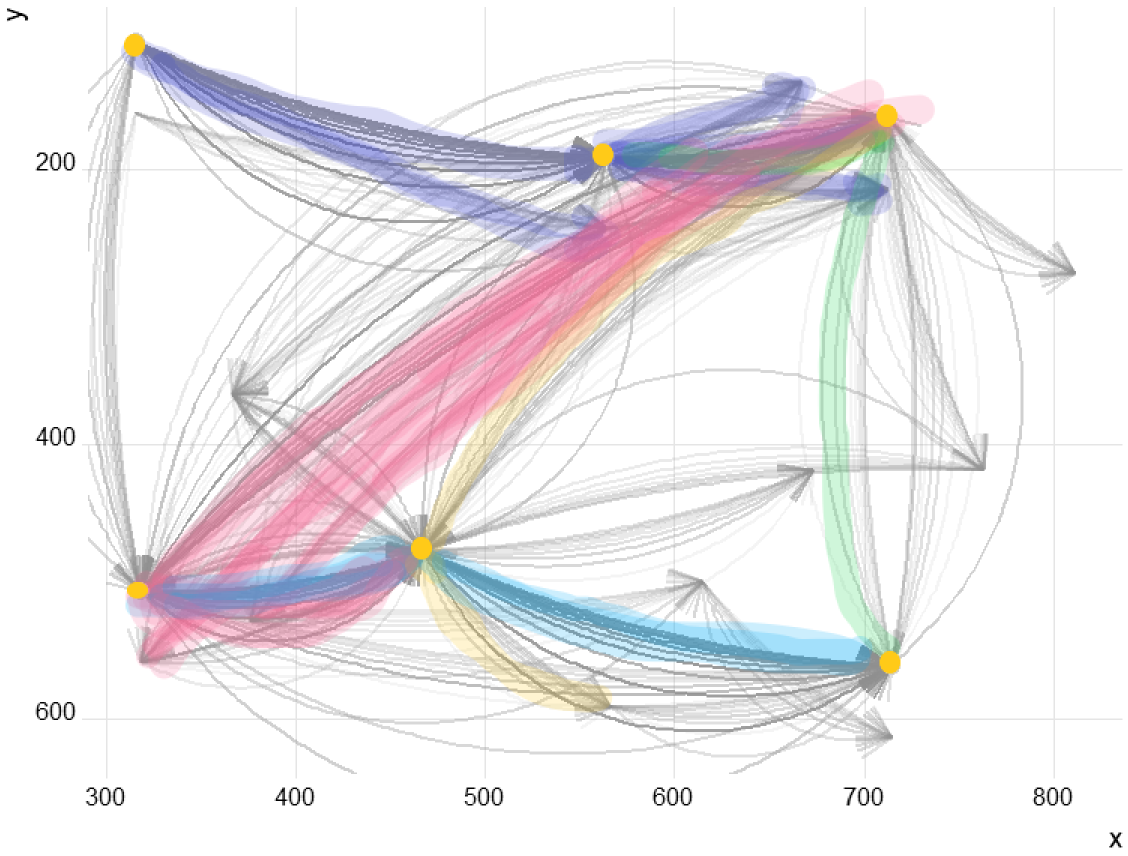


Figure 4.10: An illustration for the procedure of extracting loose common schemata. We take a decolorized Fig. 4.9 as an example, and its most common fragments of length 3 are highlighted in color: to be grouped together, the exact clicks did not have to be correct (match the yellow circle), but they had to be at most one unit away from other common answers at each node.

the board (Fig. 4.3) is constructed and the trial sequences are defined, there are no 2 petals of the same flake in the sequence of stimuli, so we can define a region of *quasi*-correct recall as those petals that are at maximum 1 petal away from the most common choice in a similar subsequence (which, in theory, could also be completely incorrect, but chosen by many). For each subsequence of length 3, we calculate its occurrence frequency as follows:

$$f_i = (\sum_{subs}(1 - 0.25w))/\sum_j f_j,$$

where a *wiggle* parameter $w = 0$ if the subsequence chosen is the same as the subsequence of reference, substracting $w = 1, 2, 3$ for each of $1, 2, 3$ deviating nodes in the 1-petal neighbourhood of the nodes of the subsequence of reference, and $w = 4$ for all the other subsequences. On this weighted frequency measure we can define a measure of entropy for the *common schemata* used in recall:

$$Entropy(f) = -\sum_j f_j * \log f_j.$$

Fig. 4.11 shows the percentage of correct answers (fully recalled sequences) as a function of the entropy measure above: the lower the entropy, the fewer there are groups of response patterns chosen between the participants. Compared within the same trial-size groups, the results suggest that the unanimity in choosing orders of recall is predictive of correct recall. In other words, if we group the fragments of responses by proximity to their respective nodes, we find common schemata bits even in incorrect answers, and they seem to help memory.

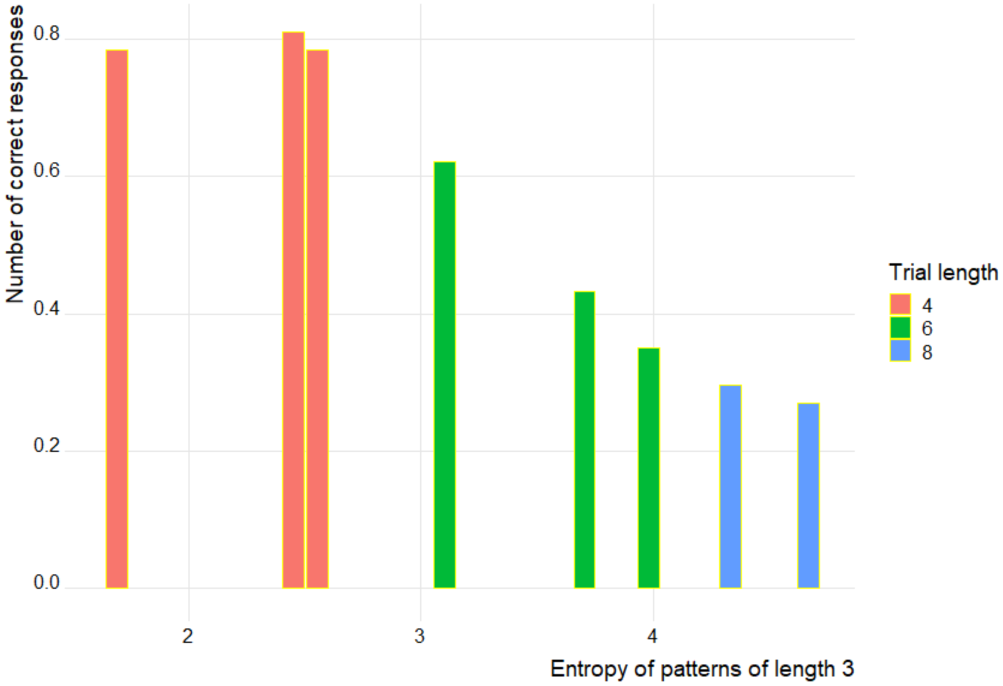


Figure 4.11: Percent of correctly recalled sequences as a function of entropy for different trials. Colors denote different trial length, that are shown on the same plot for convenience: the longer trials naturally had more possible subsequences, so the measures should be only compared within the same color-size-group.

Chapter 5

Schemata in mind wandering and in poetry

In this chapter we will describe two projects which focus on forms of the human behavior which can be argued to rely on fragmentary memory, bound in schemata: mind wandering and remembering poetry.

Recently, in the memory literature the notion of schemata, long seen as important (see e.g., [108]), has been discussed again [109], stimulated also by the analysis of its neurobiological basis in rodents [110]. A schema, whether directly functional like those involved in preparing coffee [111] or social/ornamental, like rituals of salutations [112], can be considered as a set of regularities that help organize and retrieve information [113]. In these two studies we will argue that both mind wandering and remembering poetry, processes that are not well understood from their functional point of view, but that are certainly more than ornamental for human cognitive function, rely on schemata of fragmentary memory in a rather similar way. We will attempt to characterize the functional role of schemata in both phenomena.

Mainly bound by their relation to schemata theory and the fact that my role in both of the projects was, as for schemata, functional but partial (the first one is lead by Elisa Ciaramelli and the second one by Sara Andreetta [114]), the two studies are attempts to address the question on how schemata support human memory and thought process.

In the first project described in this chapter, the mind wandering experiment, my role was to design the experimental setting with the others and set it up. While the questionnaires and the direct interaction with the participants was done by Mariachiara Esposito, I developed the experimental environment, referred to as the Snowflake, and curated its functioning. We also analysed the preliminary data together with Mariachiara Esposito. We discussed and readjusted the experimental paradigm muptiple times together with her, Elisa Ciaramelli, Massimiliano Trippa, Aline Viol and Alessandro Treves.

In the second project described here I joined at the stage when the experiment was already running and helped Sara Andreetta with its transferring online, participated in the discussions on the experiment readjustments and assisted the data analysis.

5.1 Schemata in fragmentary thought: mind wandering

In this project, still in its initial stages, we have prepared a setup to analyse in control conditions, to the extent that it is at all possible, the spontaneous dynamics of mind wandering. Mind wandering is the process of freely latching from one thought to another, planning the future or reverberating past experiences. Already from this definition, mind wandering is a great illustration to a fragmentary perspective on memory – bringing up a fragmentary episode from among one's own thoughts and jumping to the next in a schema-driven association. Like free recall, it relies on internal dynamics, but in the case of mind wandering, the whole process is internal, driven by endogenous schemata. Humans spend 25-50% of their wake time mind wandering [115], and studying the underlying neuronal mechanisms may help understand a lot about human cognitive function in general.

In fact, we already know where mind wandering lives. The brain default mode network (DMN), that is the network including medial prefrontal cortex (PFC), posterior cingulate cortex and the temporoparietal junction, was given its name for being active independently of any ongoing task [116]. DMN activity has long been associated with internal thought processes, or mind wandering [117, 118]. Moreover, recent studies have brought to light the particular roles of parts of the DMN in mind wandering. In [119] the authors compare the free thought patterns reported by participants with lesions to vmPFC to those in healthy and control-lesion participants and find that the former group tends to mind wander significantly less, but when they do, their thoughts are focused on the present and self, while the participants with an intact PFC often time travel in their thoughts, spread evenly in focus on the self and on others. In contrast, patients with a lesion to the hippocampus, in a similar experiment [120], mind wander almost as much as their healthy counterparts, but their thoughts latching lack episodicity and visual details, and can be described as plainly semantic and poor in context.

Bringing these observations together, it was further hypothesized how these parts of the DMN interact during mind wandering [121]: the vmPFC, known to be involved in schemata construction[122], initiates a stream of thought by activating a relevant schema, which is given context and schema-congruent details through iterative connections to the hippocampus-neocortex network. Assuming that mind wandering is schema-driven, it is interesting to take a closer look at an effect of individual schemata on mind wandering dynamics, and how a disruption in the DMN would alter it.

A disadvantage of the studies described above, as of most of the studies on mind wandering, is that they have mainly relied on self-reports by the participants, who would drift off in their thoughts when given a boring task to do. The common procedure, named Experience Sampling, consists in interrupting the task and asking participants about the current whereabouts of their thoughts.

Our experiment, designed in collaboration with Elisa Ciaramelli, is an effort to quantify mind wandering, independently of self-reports, in order to be able to study its dynamics as it unfolds. The idea is to introduce new episodic schemata to a group of participants by giving them a task where they have to build strong episodic associations between fixed words and test how these 'episodic memories' affect their train of thought in a free association task. Our hypothesis is that an effect of the episodic simulation should be stronger in healthy population than in vmPFC-lesioned participants.

5.1.1 Experimental design

We have designed the flake-petal wordboard (Fig. 4.1, Fig. 5.1), used in the experiments described in Chapter 4, and filled it with words taken from the Deese/Roediger–McDermott (DRM) [102, 103] lists for Italian from [104]. The DRM lists are semantically associated word lists that are created as follows: for a given word, called *lure*, a number of participants is asked to name their first association. The 12 most frequent answers are a DRM list for the lure. These lists have been used to induce false recall and false recognition: when a participant is given all or most of the 12 named words, but not the lure, and asked to recall as many words as possible, they recall or recognize the lure as frequently as the seen words, suggesting that they have formed a false memory [102, 103].

For our task, instead, we took as the stimuli the lure and the first 6 associated words from the list, thus creating basins of highly semantically associated words. The idea is to be able to contrast the effects of these semantic associations, presumed to hold across subjects, with episodic associations, both those that the participants have in their own idiosyncratic, experience-derived schemata and those that we introduce in the experiment.

To remove any outliers within stimuli, making sure all the words are distributed equally in the semantic space and are not too abstract, we have slightly adjusted the lists balancing their semantic similarity based on [123] and imageability score based on [124].

Due to its six-fold symmetries, the snowflake setting allows us to define 6 independent arrangements of the flakes that counterbalance any effect on association that could arise from the physical distance between the word groups.

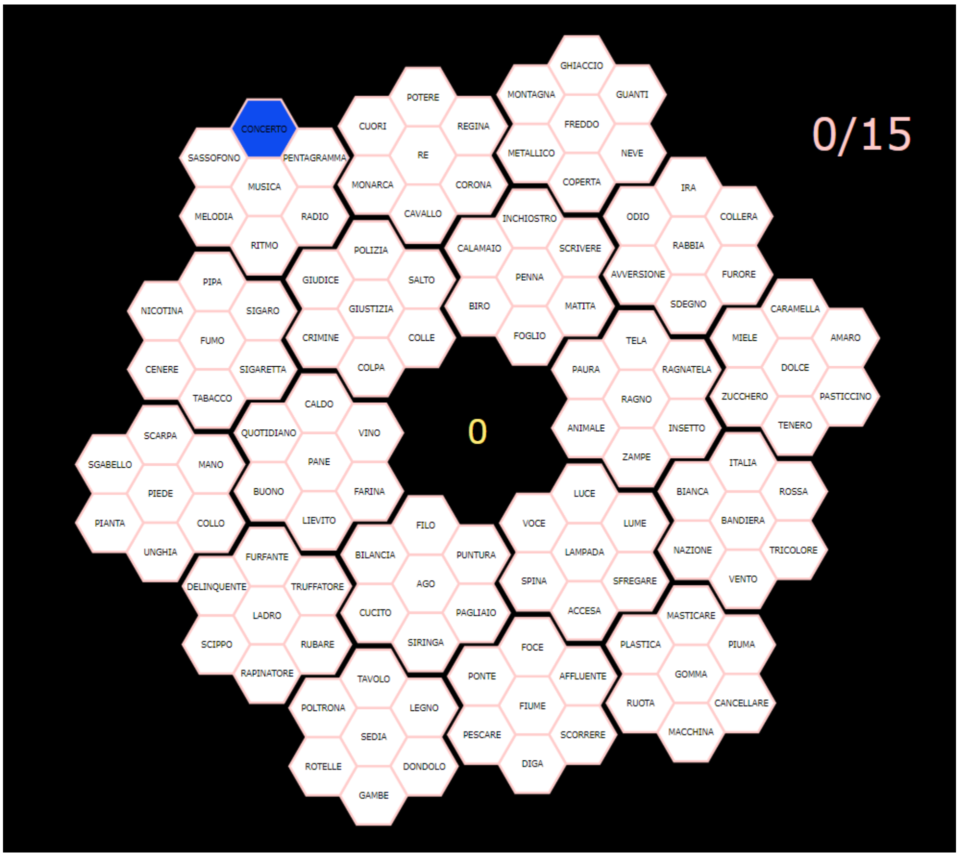


Figure 5.1: The arena with words used in the mind wandering experiment. We refer to the arrangement as a *snowflake*, each separate six-*petal* unit of which we call a *flake*. In each trial the participants are given a starting word *cue* (in blue) and they are asked to click on the words following their free associations.

Experimental procedure

The test is done in two days (see Fig. 5.2): first we attempt to introduce episodic associations, that could be construed as novel schemata, to the participants by asking them to imagine and write an episode that would link three words, that we assign specifically to each participant, from the lists above: for the experimental group the word combinations also belong to the wordboard, while for the control group the words are from lists not used in the creation of the board.

We have defined 4 sets of 9 *episodic cues* for 4 experimental groups, all including only the words, which are central words of the flakes. Note that in order to coun-

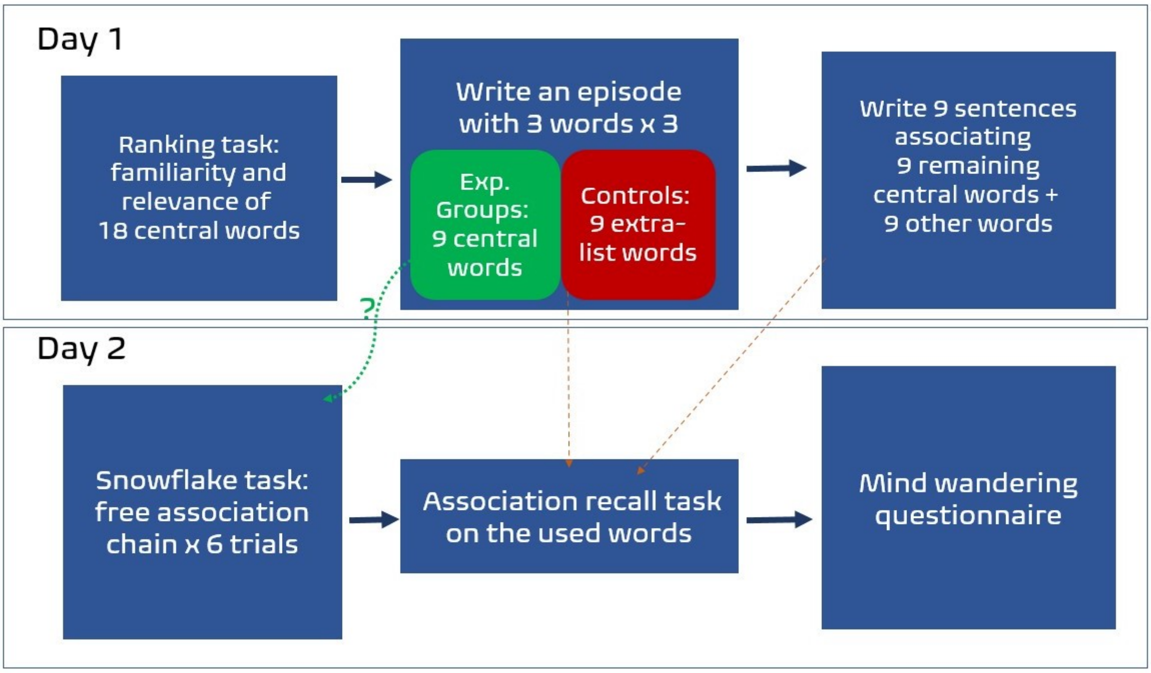
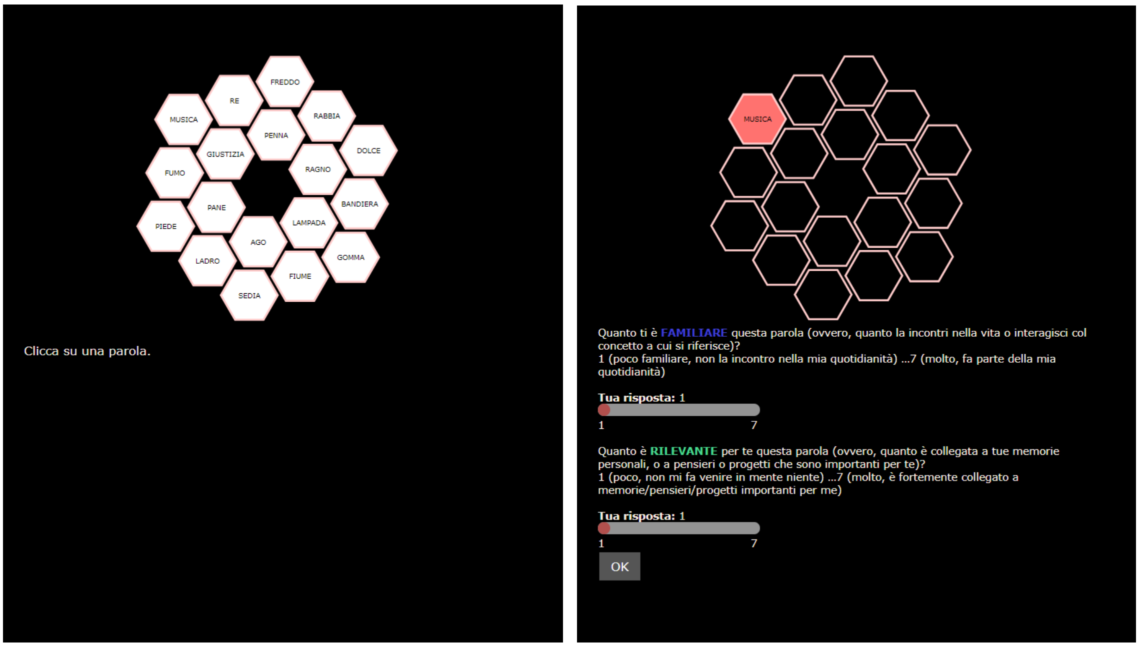

Day 1
Ranking task: familiarity and relevance of 18 central words
Write an episode with 3 words x 3
Exp. Groups: 9 central words
Controls: 9 extra-list words
Write 9 sentences associating 9 remaining central words + 9 other words
?
Day 2
Snowflake task: free association chain x 6 trials
Association recall task on the used words
Mind wandering questionnaire

Figure 5.2: The timeline of a pilot of a mind wandering experiment. The testing is done in two days and consists of the same steps for all the experimental and control groups with a difference only in the words used for creating episodes (Task 2 on Day 1).

terbalance a novelty effect on the unused central words, we give all the participants a ranking task on all the central words (see Fig. 5.3): the participants are given the layout of the central words – the lures of the DRM lists – and are asked to rank each in two categories – familiarity and relevance.

This allows the participants to get used to the arrangement of the semantic groups of words (different between the 6 layout-groups). Moreover, we get additional data on the possible closer associations of used words for the individual participants of the study.

The next day participants are presented with the wordboard. Cued with a word on it (blue on Fig. 5.1), they are asked to freely associate by choosing words on the board (clicking on them). To give participants motivation to explore the board, we give them a delayed score that is proportional to semantic similarity and physical distance of each transition between the words. The procedure is repeated 6 times with different starting words. In order to prevent falling into the same association



Clicca su una parola.

Quanto ti è **FAMILIARE** questa parola (ovvero, quanto la incontri nella vita o interagisci col concetto a cui si riferisce)?
1 (poco familiare, non la incontro nella mia quotidianità) ...7 (molto, fa parte della mia quotidianità)

Tua risposta: 1

1 7

Quanto è **RILEVANTE** per te questa parola (ovvero, quanto è collegata a tue memorie personali, o a pensieri o progetti che sono importanti per te)?
1 (poco, non mi fa venire in mente niente) ...7 (molto, è fortemente collegato a memorie/pensieri/progetti importanti per me)

Tua risposta: 1

1 7

Figure 5.3: The separate ranking task given to participants in the mind wandering experiment: given an arrangement of the central words (left) from the snowflake board (Fig. 5.1), participants are asked to chose them in a free order and rank the words in their familiarity and relevance (right).

stream, the used words are temporarily disabled in the adjacent trials. After the task the participants are asked to recall the associations formed the day before, so we can differentiate between the explicitly and implicitly encoded schemata.

5.1.2 Preliminary results and predictions

We ran 4 pilots for this study, varying additional tasks and their importance. For example, when we found novelty effect affecting significantly the chance of choosing unseen words with respect to those used in the first-day tasks, we added a ranking task and a sentence-association task (Fig. 5.2).

Sample. In each pilot, the general procedure consisted in testing 16 participants in the experimental group – 2 in each subgroups of episodic cues (in the pilots we used 2 sets of words out of 4 we prepared), using 4 different layouts of the snowflake; and 8 or 16 participants (in different pilots) in the control group, also using 4 different layouts of the snowflake. All the participants were taken from the Cognitive Neuroscience master course at the University of Bologna, that the population samples of different pilots were equally distributed in age, gender and education. In the experiment planned to be run with the patients, the sample will be matched in all demographic variables accordingly.

Data analysis. In the initial stages of our study we have been aiming to analyse the dynamics of free thought latching in its unfolding. First of all, we want to see whether there is any effect of the schemata imposed by the episodic simulation task. To test that, we compare the number of times the *episodic cues* (central snowflake words used in the writing task by experimental groups) were chosen in associative chains within experimental and control groups. Further, we look at the individual 'jumps' – the frequency of transitions between episodic cues within schema-triplets and towards other words on the board. We also look at the possible covariates in the schema-driven associations: the coherence of constructed episodes, association recall on the next day, familiarity and relevance rankings, and the results of the mind wandering questionnaire. In parallel, we run a more elaborate analysis on graphs of transitions in order to find less visible effects of the episodic memory manipulation on free association chains.

Preliminary results. Our prediction was that the experimental group will have association dynamics altered by the newly introduced schemata, i.e., the "episodic" associations among the sets of 3 words assigned to them to construct an episode with, the day before; compared to the control group. This prediction was not borne out in the 4 pilots we have run. Fig. 5.4 shows one of the preliminary results of pilot 4: contrary to our predictions, the experimental group does not show in their free association chains a preference towards the words used in episodic simulations, even though the participants of the experimental groups remember the associations, according to the recall tasks (results not shown). With respect to the control group (who did not use these words in episode writing), the experimental group uses less central words overall. We also do not find associative 'jumps' within the episodically bound triplets that we expected to see (not shown). We hypothesize that the episodic simulation task in the form that it is used now may be too straightforward, and

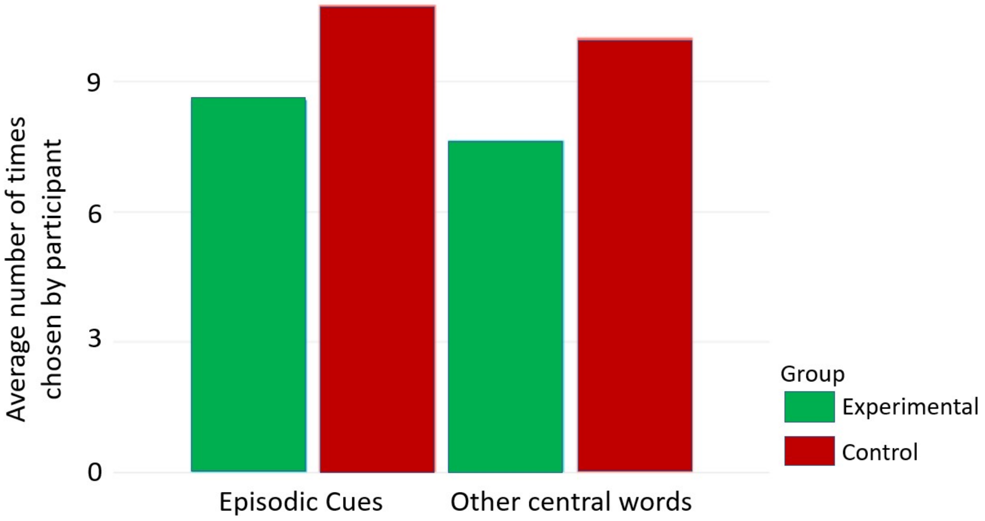


Figure 5.4: Preliminary results of a pilot: a total over 6 trials of the number of times a participant chose a cue used in writing an episode the day before, averaged across participants for each cue group within experimental and control groups of participants.

the participants understand what is expected of them and feel inclined to do the opposite.

Running pilots of the study, we have attempted with successive modifications to make these episodic associations stronger but more subtle, and we are currently revising the paradigm to explore more effective ways to associate the "episodic words" in the mind of the participants. The plan is to further test a group of patients.

Given the abundant evidence of the role of the vmPFC and hippocampus in forming episodic memories, schemata and mind wandering [118, 120, 121, 125], we expect to see different effects of the episodic simulation on the free association patterns in these groups. In addition, this new paradigm for assessing mind wandering dynamics may allow us to see more in detail differences in free thought processes between patients and healthy participants.

5.2 Schemata in poetry

Poems, nursery rhymes, traditional songs: they are found in every culture, and they have been around for ages, well before the advent of writing systems. Sometimes, they have or had the crucial mission of carrying an important message for the listeners: a list to know by heart, an event happening every year, a warning of a potential danger. What do these texts have in common? At least one aspect: they adopt a variety of devices that help hold verbal material in memory.

Human memory can, in fact, fail spectacularly at times. Writing systems have helped safely store verbal information, in a format relatively difficult to tamper with; before, when our ancestors had to rely on their fallible memory, a number of linguistic devices crystallized to help them remember words and verbal material. Cultural transmission, then, has depended for ages on these devices, which in poetry we can broadly refer to as “meter”. These devices may range from the use of repeated metaphors: “rosy-fingered dawn” in Homer [126], to the ring composition in the Zoroastrian Yasna [127] to semantic repetition as in Biblical poetry: “In the way of righteousness is life; and in the pathway thereof there is no death.” in Prov. 12:28 (King James Version).

In several Western literary traditions, including the Italian one, the local structure of poetry revolves around the verse, and includes a constant number of syllables, a limited variability in the pattern of accents, and a specific organization of rhymes.

Can the role of metrical features be explained from a neurocognitive point of view, with respect to memory? In this context, we consider the components of meter as schemata which, by encouraging regularities, facilitate the recall of verses. They operate as schemata insofar as they help us recruit, and possibly produce, the next element of a sequence stored in our memory.

In facilitating verbal sequence *replay*, metrical features appear to be effective with extended “trajectories”, lasting even several verses. These are extended relative to the short trajectories thought to be produced by the phonological loop of Baddeley’s model of working memory, which are presumed to last only a couple of seconds, precisely because of the lack of specific devices that extend their range[128]. To the best of our knowledge, though, the effectiveness of these features has never been quantified. In this study, we aim at measuring the strength of some metric devices.

Specifically, we focus on the three main characteristics of classical Italian meter: rhyme, pattern of accents, and verse length.

5.2.1 The two experiments

We extracted passages from two masterpieces of Italian literature: the Divina Commedia by Dante Alighieri (1265–1321), and the Orlando Furioso by Ludovico Ariosto (1474–1533). From the latter we chose ottave (octaves, stanzas of eight verses) from canti XIII, XV, XIX and XXX, and one from canto I to train subjects, while from the former we selected sequences of three consecutive terzine (hence nine verses) from two canti from Inferno (XXIV for the experiment, V for training), two from Purgatorio (VI and XVI) and one from Paradiso (XXVII). All passages had only proper (Italian) hendecasyllables with an accented 10th syllable, and were, to our arbitrary judgement, devoid of explicit or easily reconstructed memorable content or references.

Poem manipulation

The original texts were manipulated in a number of different ways. Firstly, most content words were converted into non-words in order to eliminate discernible semantic content, hence semantic effects on memory; an effort was made to change phonemes with similar ones, while maintaining the original prosody. Function words were not modified. This applied equally to all texts and resulted in "original non-poems" (ONPs).

The second stage of manipulations focused on metrical patterns. We created three conditions:

- a condition where we eliminated rhymes ("NPR" – Non-Poem without Rhyme)
- a condition where the accent patterns of four-five verses per passage were replaced with less standard ones ("NPA" – Non-Poem with modified Accents). To validate proper original accents, we consulted with an expert scholar for

Orlando Furioso, whereas for Divina Commedia we referred to the "Archivio Metrico Italiano", a database collecting masterpieces of Italian literature with their accents annotated (www.maldura.unipd.it).

- a condition where the number of syllables per verse, which in the ONPs were strictly 11 throughout (regular hendecasyllables) was altered again in four-five verses, to nine, 10, 12 or 13 –("NPS" – Non-Poem with wrong numbers of Syllables). Note that by adding or subtracting one or two syllables, also the pattern of accents was perforce altered, but we attempted to make the alteration less noticeable than the number change, in contrast to the NPA condition in which, while there were strictly 11 syllables/verse, the accents followed more unusual patterns.

These manipulations were applied to all texts. All texts were then recited by a professional actor and audio recorded.

For the experiment, every subject was administered four texts in total, by the same (original) author: one per canto and one per condition. Therefore, twenty-four combinations were created.

An example of an NPS we used, from the Commedia, is presented in Fig. 5.5 together with its original spectrogram,

Ranking

We conducted an online survey about how these manipulated poems were perceived by a group of Italian native speakers. Participants were asked to listen to the four conditions and give a ranking of preference, from the one that sounded the most "poetically plausible" to them, to the one they perceived as the strangest.

Subjects. 62 people participated in the online survey for Ariosto (F=32, M=30, mean age = 29.06, sd= 8.13) and 65 people for Dante (F=35, M=30, mean age = 26.48, sd = 6.26). Part of either cohort were the participants in the main experiment below, but tested with the other author, and they were asked to complete this survey after the end of the second session of the main experiment. Another group

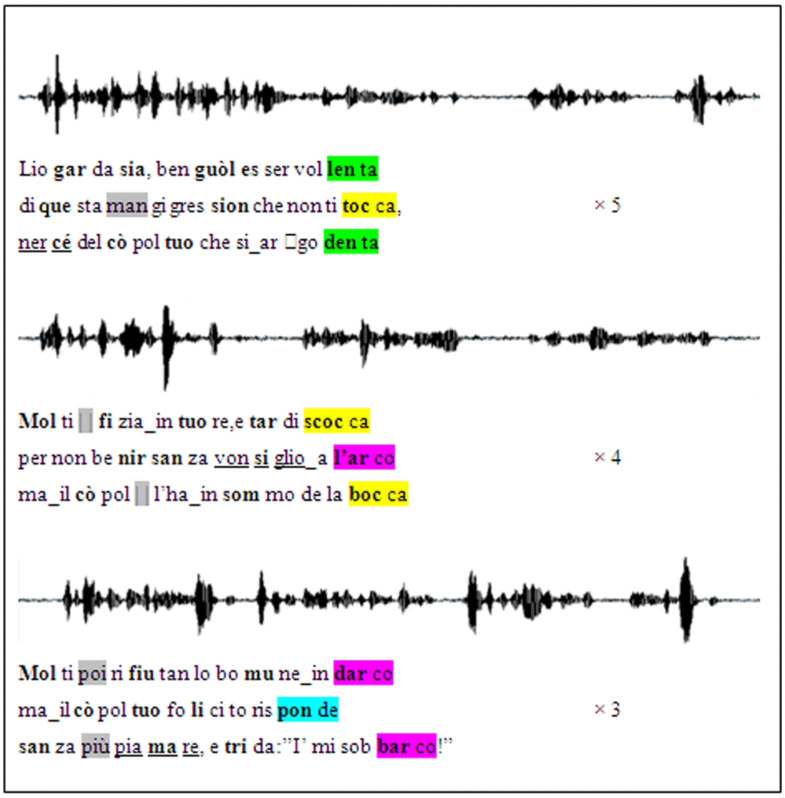

Lio gar da sia, ben guòl es ser vol len ta
di que sta man gi gres sion che non ti toc ca, × 5
ner cé del cò pol tuo che si_ar □go den ta
Mol ti fi zia_in tuo re,e tar di scoc ca
per non be nir san za von si glio_a l'ar co × 4
ma_il cò pol l'ha_in som mo de la boc ca
Mol ti poi ri fiu tan lo bo mu ne_in dar co
ma_il cò pol tuo fo li ci to ris pon de × 3
san za più pia ma re, e tri da:"I' mi sob bar co!"

Figure 5.5: **NPS example derived from Purgatorio, canto VI**. For each *terzina* (vv.127–135) the NPS text, shown below the sound wave by the professional actor, maintains rhymes, in color, and accents, in boldface, as in the ONP version; whereas overall three syllables have been added and two taken away, in gray. The underlined non-words were the targets of the memory test; underlined blanks denote *synizesis* (when two syllables are pronounced as one).

of participants was recruited through the online platform Prolific (www.prolific.co). This last group was compensated with five euros. We had aimed for 72 rankers in each cohort, but had to exclude some a posteriori, who failed to complete the ranking in full.

Experimental design. The online survey was designed with the open-source toolkit Psytoolkit [129]. After an example, presented as training also in the main experiment below, they listened to the four poems one at a time. Every poem was associated with a name, in order to help participants refer to that specific condition. If they wanted, they were allowed to listen again and again to the same poem before proceeding to the next.

At the end they were asked to rank the four poems: from the one they perceived as the best, to the one that sounded worse to them. From the rankings by all participating subjects we extracted an average index of metric plausibility by assigning a value 0.6 to the first–ranked condition (e.g., NPA), 0.3, to the second, 0.1 to the third, and 0 to the fourth. The logic behind this assignment is that subjects occasionally reported being unsure as to which passage sounded the strangest. The rankings were collapsed across canti, with the relatively large number of participants ensuring approximately even sampling (each canto was presented originally 18 times per condition, which came down to 16+/-2 after the exclusions). As a result, the average metric plausibility of each condition could in principle range from 0 to 0.6, but in practice was much more restricted, particularly with passages from Dante, to values around the average of 0.25 (see Fig. 5.6).

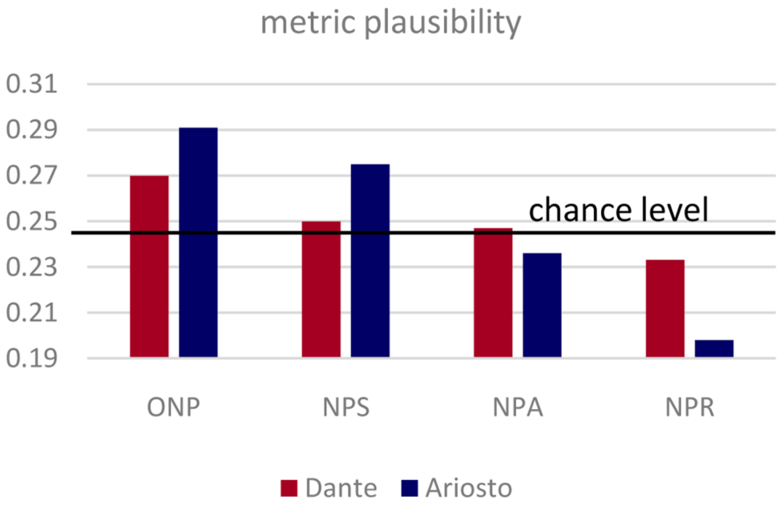


Figure 5.6: **Relative metric plausibility**. The different versions of the same four passages from the Divina Commedia (red) and Orlando Furioso (blue) were ranked in the same order, but the plausibility index is more spread out for Ariosto.

Memory experiment

Subjects. 48 native Italian speakers who had been exposed to Italian literature through one of the national high school curricula were recruited for the main experiment. Half of them were administered material from Ariosto (F= 15, M=9, mean age= 26.34, sd=4.02), the other half from Dante (F=15, M=9, mean age= 26.12, sd=3.61).

Experimental design. The experiment was designed with Psychopy Builder [130]. It included a study phase of about 30 minutes the first day, and the test of about

10 minutes the second day.

We aimed at an almost exclusively auditory experiment, in order to assess how memory relies on meter if listening is the only available channel to learn from [131]. Indeed, the material included audio files only, with the sole exception of written material when a fill in the gap task appeared.

Besides the four passages, verses from two other canti were used for training, as indicated above. However, these verses were presented only in the ONP condition, leaving meter intact.

Every passage, including the training, was associated with an image, taken from among Gustave Doré's illustrations of the Divina Commedia. The images were consistent for the same canto in different conditions and were intended to help engage memory without, at the same time, biasing the linguistic material. Every poem, including the training, was presented divided into three consecutive portions.

5.2.2 Results

Two separate cohorts or rankers, for Dante- and Ariosto-derived non-poems, were presented with a combination of the four passages from the same author, one in each of the ONP, NPS, NPS and NPR versions, and were asked to rank them in order of metric plausibility. The fully balanced design allowed us to extract a passage-independent plausibility score.

Data analysis. The outcome of interest is essentially the presence of a significant correlation between the dependent variables measured in the ranking (the plausibility index, see Fig. 5.6) and in the memory experiment (correct responses, and reaction times, see Fig. 5.8), that would indicate a joint dependence on the type of passage manipulation. Correlations were considered significant at $p < 0.05$.

The contribution of distinct components to metric plausibility

Both when derived from passages by Dante and Ariosto, non-sense poems were found most plausible in their fully metric ONP versions, somewhat less when the number of syllables was manipulated (NPS), even less when the pattern of accents was altered in the NPA renditions, and the least when rhymes were removed, NPR. Remarkably, however, differences in the plausibility index are shown in Fig. 5.6 to be quite limited, making the fully balanced design essential. The variance was particularly limited in passages from the Commedia, which may be due to Dante's taking more liberties with the meter he had adopted (the same hendecasyllables as Ariosto, but in terzine rather than ottave). To quantify this perception, at least in relation to accent patterns, which are more accessible to analysis, we applied two independent measures of accent variability to the four original passages by each author.

Dante appears to be slightly more variable in his accent patterns relative to Ariosto, but the main observation that can be gleaned off Fig. 5.7 is that both poets are far from using a fixed pattern, utilizing over half of the maximum entropy they had available in terms of accenting those passages.

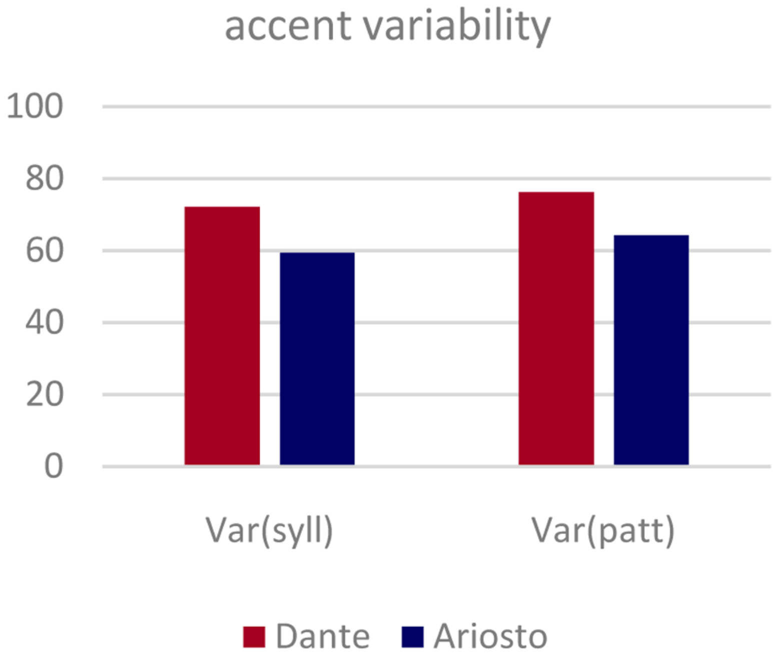


Figure 5.7: **Variability in the pattern of accented syllables in the eight original passages by the two authors**. Two independent entropy measures of variability, per syllable and per verse are both normalized to range from 0 (a single fixed pattern) to 100% (i.e., each syllable in the verse is accented half the time; or each verse follows a different accent pattern).

Meter can facilitate memory

Does such a loose structure help remember individual words? Fig. 5.8 (upper) shows that it does, only for the non-poems derived from Ariosto's octaves. Twenty-four subjects per author were asked, one day after repeatedly listening to one version of each passage, to remember non-word targets out of three alternatives, upon listening to the non-poem with selected non-words muted. There were three such targets in each non-poem. While in the case of those derived from passages in the Divina Commedia the overall number of correct responses per condition was unrelated to its metric plausibility (r2=0.04), seemingly fluctuating as much as the correct responses to the first, second, and third query taken alone, for the passages from Orlando Furioso the correlation with metric plausibility was remarkable (r2=0.98) and highly significant ($p < 0.01$). Interestingly, the total score of the two cohorts was nearly identical, 147 for Ariosto and 148 for Dante, out of a total of 288 (24×4×3).

Meter helps, but not for free

The analysis of reaction times helps interpret the above results. As shown in Fig. 5.8 (lower), overall it took longer for participants to pick a wrong answer over the correct answer (on average, 733 ms more), and it took longer for participants tested with Ariosto, relative to those tested with Dante, to respond (on average, 547 ms more). Most importantly, in each of the four types of trials above, the more metrically plausible the passage, the longer the reaction time. However, the trend is significant only with Ariosto, if data from the two authors are analyzed separately, and it is significant overall ($p < 0.004$) with a slope mainly determined by the Ariosto data, if analyzed together, as shown in Fig. 5.8. The slope for the Dante data alone would be higher, but not significant, likely because of the limited plausibility range spanned.

These findings suggest that processing meter in order to help retrieve a non-word heard the day before has a cognitive cost, and takes the order of hundreds of ms extra time, depending on exactly how much meter is "used up" in the process. For passages derived from Dante, it appears that although outwardly the metric

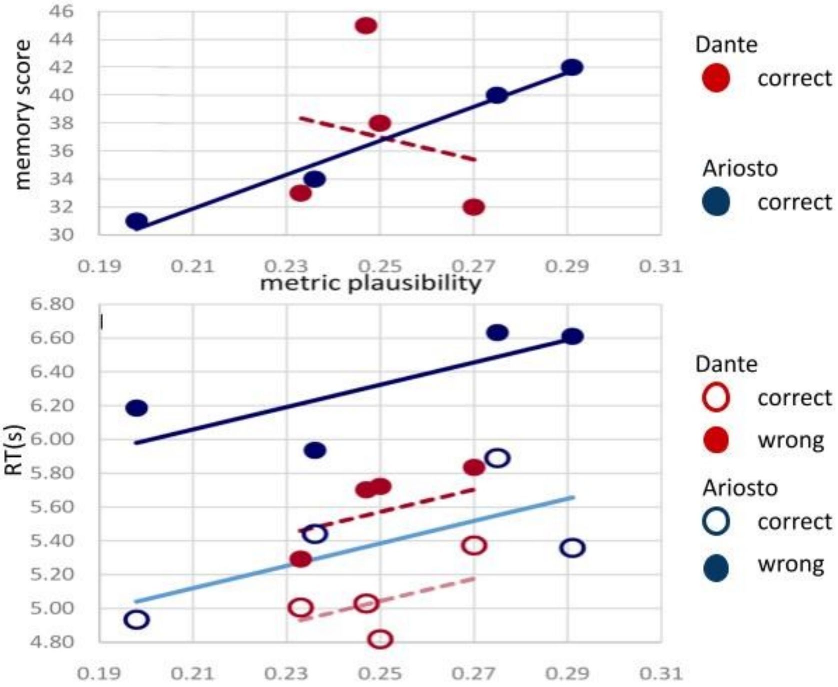


Figure 5.8: **Memory and reaction times both increase with metric plausibility**. (Upper) Overall correct responses (out of 72) for each condition, ordered in terms of their metric plausibility, as in Fig. 5.6, for passages from Dante (red) and Ariosto (blue). (Lower) Reaction times (in seconds) for correct (circles) and wrong responses (dots) are regressed against plausibility for each author, with a single slope parameter. The slope is significant and similar to that characterizing the Ariosto data alone, whereas it is denoted with a dashed line for the Dante data, because the latter would not produce a significant correlation on its own.

structure is essentially the same (with the slight qualification reported in Fig. 5.7, and the note that a passage is a sequence of three terzine rather than a single ottava), meter is used less, and the very same memory performance is attained on average in less time.

Chapter 6

Conclusions

In this work we study several aspects of human and (model) rodent memory and attempt to suggest that memory relies on navigation by fragment fitting, roughly as envisaged in [1].

First of all, we build upon the mathematical model of CA3 after [56] and address a number of questions concerning spatial memory. It has been shown in [56] that the recurrent network of CA3 can store multiple maps in the form of quasi-continuous attractors. We suggest that by only changing the learning procedure to slower and interleaved, we can improve uniformly the storage and thus the decoding accuracy for maps of multiple sample environments. We find a capacity limit for the network storage that is broadly in line with analytical predictions.

Further, we show that intrinsically within the recurrent network we can model replay of exploration routes by adding neuronal adaptation to their firing dynamics. Finally, analyzing the continuous portions of correctly decoded trajectories, we observe that the closer they are to the trajectory used in learning, the higher is the chance of their correct classification. We suggest that this finding falls in line with the navigation by fragment hypothesis: given a very short exploration period the CA3 network may, in fact, not learn the *map* of the environment evenly, but rather just some fragments and, possibly, generalize in between.

Moving on, we suggest a human analogue of random exploration and retrieval – the free recall of spatial locations. We explore the human capacity of recall and its dynamics through spatial memory tasks consisting in recalling locations on a hexagonal grid, shown simultaneously or serially and with or without spatial structure (in separate experiments).

In the first experiment, we find that participants can freely recall $\propto \sqrt{L}$ out of L given random locations shown simultaneously. This result has been earlier shown to be a limit scaling law for recall of lists of words and facts [80, 82], but it is puzzling with respect to the numerous studies showing a single limit of e.g. 4 or 7 items in short-term memory (STM).

We suggest a Potts network, and specifically its *latching dynamics* as a model of recall of a number of STM items within all the LTM items (stored in the Potts network). We show that the unrestricted latching dynamics of Potts network also follows the scaling law found for words and for spatial locations.

Instead, in the more restricted experiment testing serial recall, where we did not allow any mistake in the order, we find indeed a capacity limit of short-term memory largely independent of the material used, that is lower the shorter the time of presentation. This again was a feature reproducible by the Potts network, when restricted to no return to the visited pattern. Notably, latching by the network has itself a semi-random nature, so its ability to reproduce some statistical aspects of human recall stimulates more detailed analyses of the latter.

Further analyzing the responses in a free recall task, we find that participants seem to remember a general fragmentary schema of locations on the screen and during recall they make mistakes that are spatially very close to correct answers – a feature difficult to quantify in a free recall of words.

To further explore the difference in recall limits we have designed a single experiment for spatial recall under different conditions – with simultaneous or serial presentation, differing recall instructions and two measures of correct recall. As expected, the results indicate that the overall number of recalled items is highest when the procedure is least restrictive: the items are shown simultaneously and mistakes are allowed.

Next, we chose to add words as stimuli for recall to the spatial task. Given evidence that mnemonic techniques using spatial anchoring of non-spatial memory items improve longer-term memory recall performance [101, 107], we wanted to test whether anchoring words to spatial locations would help their memorization. The main result of this experiment is that no, the spatial positioning does not seem to improve the recall capacity in words, nor the opposite effect seems to exist. The participants have been shown to memorize well neighbourhoods of shown items, not so well individual locations, while word recall seemed to stably hit the same limit of 4-5 items.

Putting together the observations gathered so far, we decided to test whether the above-mentioned fragmentary schemata are common across population. We have run experiment presenting exactly the same spatial configurations of stimuli to participants. We found that the participants choose common orders when recalling the positions and these orders are, on average, shorter than random and possess a number of other shared properties. We describe our attempt to characterize these schemata as configuration biases favoring the observed geometrical features (e.g., shortest distance), but we do not find this first approach to fully explain the observed unanimity in order of recall.

We further hypothesize that these orders are in fact biased by fragmentary schemata that are native to participants' memory. Considering these schemata as attractive states, we were able to model recall basing it on 'collective' experience and show that the biases could explain the correct and unanimous recall in the beginning of each trial that fades as the trial progresses.

Considering fading, we returned to the question of an overall blurry schema that the participants seems to have for a configuration of positions on the screen, that results in faulty recall dynamics, but similar across participants. Indeed, when we calculate the occurrence of subsequences in the recall order across participants, grouping responses by proximity across nodes, we find that the recall capacity increases with a decreasing entropy of choices within groups of common fragments (including errors), suggesting that the more common across participants the fragments of a sequence are the easier it is to remember the whole sequence.

These last two results give us two important pieces of the puzzle we have been putting together in this thesis: we find both common biases in the beginning of recall that fade with the trial progression and traces of an overall schema consisting of these common fragments. Furthermore, these results seem to suggest a common tendency in increasing entropy during recall within trial as well as increasing entropy with the trial number. We plan to further validate these findings with more different spatial patterns and more participants.

Now, taking a leap to a different memory-dependent process, namely, remembering poetry, we suggest that remembering fragments of poems is aided by poetic meter as schemata-driving elements. We show that a similar replay process to the one mentioned above helps in the memorization of non-words inserted in poetry: the different features of poetic meter are shown to improve the recall performance selectively, when needed.

Just before that, we propose an experimental setting in which we should be able to test to what extent fragmentary schemata operate our train of thought in mind wandering. Bridging the evidence suggesting that mind wandering depends on vmPFC [119] and schemata [122], we hope to be able to quantify the differential effect evoked by introducing novel episodic schemata to free thought process between a group of patients with lesion to vmPFC and a group of healthy controls.

We want to highlight the thread binding these results and observations together – the idea of fragmentary memory. In our recall experiments we find that people tend to remember better the fragments of information that fall in line with what they have experienced before and that traces of experience help memory by activating selectively: for example, as rhyme sometimes helps to recall verses in our poetry experiment, and sometimes it does not.

We observe similar trends in our simulations of models of the memory system: in our model of CA3 we find that the localization accuracy is higher when the remembered fragment is close to a learning trajectory – similarly to how human participants have shown to recall better the locations close to forming a familiar geometrical shape, say, a straight line. In our model of short-term memory, the Potts network, we find that congruent instructions lead to better recall – again, in human recall it is analogous to remembering better the fragments that resemble one's own experience. This also supports our findings on influence of recall instructions on performance.

The message of this thesis is that schemata that help our memory are sparse: in our model of CA3 they are parts of the learning trajectory that "attract" memory and in human recall the same function is associated to the fragments that "click": only when an observed configuration of dots is close to a remembered schema it will be recognized as such. In our Potts model of STM we hypothesize that a similar effect may be produced by the heteroassociative term: once the network retrieves the element in a short sequence that has been effectively stored and that can then be remembered, a whole fragment emerges following the initial order. We hypothesize that variables of poetic meter have a comparable role, e.g. the length of a phrase can fit a remembered verse schema and this helps retrieve the correct (non)word.

Theoretically and empirically we thus observe potential signatures of fragmentary memory effect on recall performance – in number of items and in accuracy, so we cannot help but wonder whether storing information by fragment could be more efficient computationally. It seems intuitively reasonable that memorizing all the incoming information in great detail can be computationally expensive and thus probably redundant over time. After all, we all learn the importance of generalization at school. But understanding the advantages of fragmentary memory at the level of navigation, for example, could help us learn more about memory function overall.

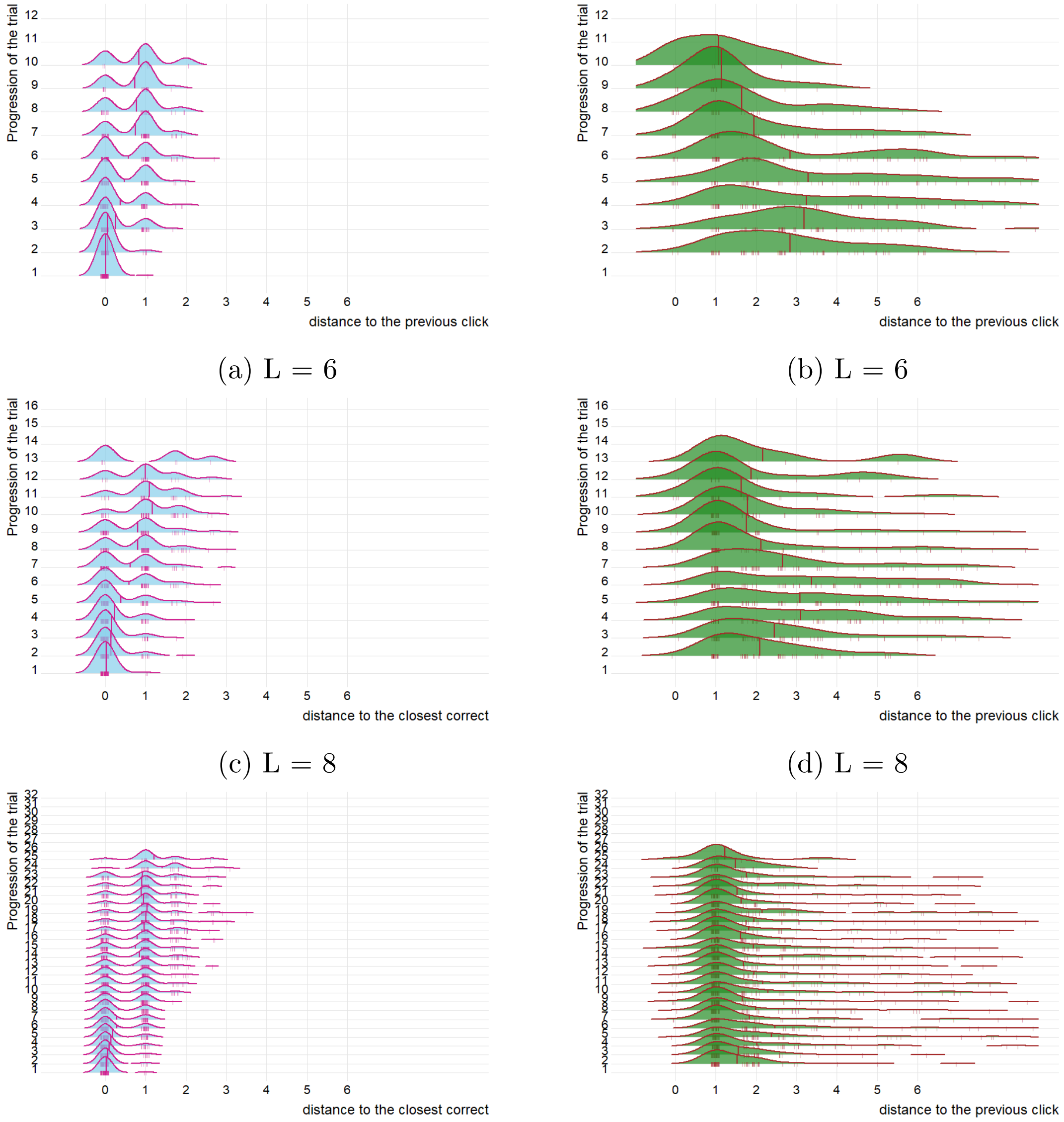
Further work is needed to determine how place cells could be involved in the fragmentation. Recent experimental evidence in rodents has been showing that the presence of objects in the foraging arena attracts activity of spatially selective cells

[64, 132], and it seems natural to suggest that fragments of space near important objects bear more significance for memorization, like, for example, routes to an exit or to food storage sites in an animal natural habitat. It would be interesting to design experiments in humans and rodents to test such a hypothesis and on the computational modeling could give further insight on of the problem.

Another direction to explore in this framework is the influence of fragmentary memory on different cognitive functions. One of the cognitive functions we started to study has been the free thought process. More work is needed in order to be able to characterize the role of fragments in memory on mind wandering. For the nearest future we envisage to explore at least two areas of interest within our paradigm – the imposed fragmentary schemata in mind wandering (experiment of Chapter 5.1) and those schemata that are already present in participants' experience (continuation for the word-location experiment of Chapter 4.2).

Appendix A

Dynamics of recall: additional figures



(a) L = 6

(b) L = 6

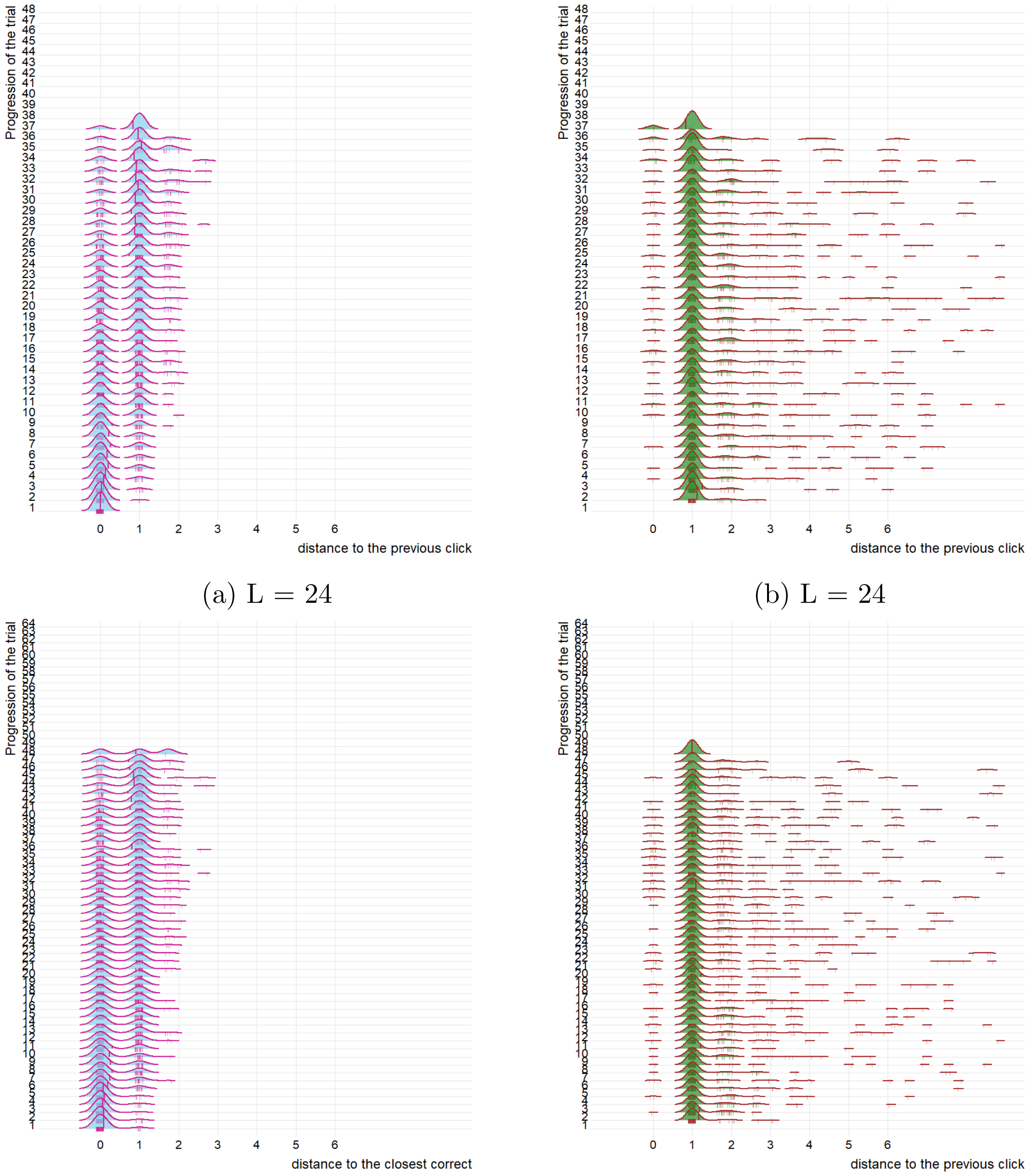
(c) L = 8

(d) L = 8

(e) L = 16

(f) L = 16

Figure A.1: The individual clicks by the participants as the trial progresses (y axis) for different trial types – of length 6, 8, 16. The figures in the left column show the distribution for all participants of the distance of the current click to the closest correct. The figures in the right column show the distance of the current click to the previous click. Distances are normalized to grid units (as in Fig. 3.3). Vertical lines show the average value, while the short vertical dashes show individual points (with jitter). Note: there was a maximum of $2L$ clicks available for each trial of length L, but the trial ended sooner, e.g. if the participant recalled all the correct locations before reaching $2L$.



(a) L = 24

(b) L = 24

(c) L = 32

(d) L = 32

Figure A.2: The individual clicks by the participants as the trial progresses (y axis) for two trial types – of length 24 and of length 32. The figures in the left column show the distribution for all participants of the distance of the current click to the closest correct. The figures in the right column show the distance of the current click to the previous click. Distances are normalized to grid units (as in Fig. 3.3). Vertical lines show the average value, while the short vertical dashes show individual points (with jitter). Note: there was a maximum of $2L$ clicks available for each trial of length L, but the trial ended sooner, e.g. if the participant recalled all the correct locations before reaching $2L$.

Bibliography

[1] Robert Worden. "Navigation by fragment fitting: a theory of hippocampal function". In: *Hippocampus* 2.2 (1992), pp. 165–187.

[2] John O'Keefe and Jonathan Dostrovsky. "The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat." In: *Brain research* (1971).

[3] Torkel Hafting et al. "Microstructure of a spatial map in the entorhinal cortex". In: *Nature* 436.7052 (2005), pp. 801–806.

[4] Rebecca D Burwell and Kara L Agster. "Anatomy of the Hippocampus and the Declarative Memory System, chapter 3.03". In: *Learning and memory-A comprehensive reference. Volume* 3 (2008).

[5] Song-Lin Ding. "Comparative anatomy of the prosubiculum, subiculum, presubiculum, postsubiculum, and parasubiculum in human, monkey, and rodent". In: *Journal of Comparative Neurology* 521.18 (2013), pp. 4145–4162.

[6] Erica L Stevenson and Heather K Caldwell. "Lesions to the CA 2 region of the hippocampus impair social memory in mice". In: *European Journal of Neuroscience* 40.9 (2014), pp. 3294–3301.

[7] Kazuki Okamoto and Yuji Ikegaya. "Recurrent connections between CA2 pyramidal cells". In: *Hippocampus* 29.4 (2019), pp. 305–312.

[8] Ute Häussler et al. "Mossy fiber sprouting and pyramidal cell dispersion in the hippocampal CA2 region in a mouse model of temporal lobe epilepsy". In: *Hippocampus* 26.5 (2016), pp. 577–588.

[9] Tamir Eliav et al. “Multiscale representation of very large environments in the hippocampus of flying bats”. In: *Science* 372.6545 (2021).

[10] C. Daniela Schwindel et al. “Reactivation of Rate Remapping in CA3”. In: *Journal of Neuroscience* 36.36 (2016), pp. 9342–9350.

[11] James B Ranck Jr. “Head direction cells in the deep layer of dorsal presubiculum in freely moving rats”. In: *Society of Neuroscience Abstract*. Vol. 10. 1984, p. 599.

[12] Klara Gerlei et al. “Grid cells encode local head direction”. In: *BioRxiv* (2020), p. 681312.

[13] Francesca Sargolini et al. “Conjunctive representation of position, direction, and velocity in entorhinal cortex”. In: *Science* 312.5774 (2006), pp. 758–762.

[14] Hanne Stensola et al. “The entorhinal grid map is discretized”. In: *Nature* 492.7427 (2012), pp. 72–78.

[15] Trygve Solstad et al. “Representation of geometric borders in the entorhinal cortex”. In: *Science* 322.5909 (2008), pp. 1865–1868.

[16] Emilio Kropff et al. “Speed cells in the medial entorhinal cortex”. In: *Nature* 523.7561 (2015), pp. 419–424.

[17] David B Omer et al. “Social place-cells in the bat hippocampus”. In: *Science* 359.6372 (2018), pp. 218–224.

[18] Sachin S Deshmukh and James J Knierim. “Representation of non-spatial and spatial information in the lateral entorhinal cortex”. In: *Frontiers in behavioral neuroscience* 5 (2011), p. 69.

[19] H Eichenbaum et al. “Cue-sampling and goal-approach correlates of hippocampal unit activity in rats performing an odor-discrimination task”. In: *Journal of Neuroscience* 7.3 (1987), pp. 716–732.

[20] Timothy V Bliss and Anthony R Gardner-Medwin. "Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetized rabbit following stimulation of the perforant path". In: *The Journal of Physiology* 232.2 (July 1973), pp. 357–374.

[21] Terje Lømo. "The discovery of long-term potentiation". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 358.1432 (Apr. 2003), pp. 617–620.

[22] David Marr. "Simple memory: a theory for archicortex". In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 262.841 (1971), pp. 23–81.

[23] David Marr. "A theory for cerebral neocortex". In: *Proceedings of the Royal Society of London. Series B. Biological Sciences* 176.1043 (Nov. 1970), pp. 161–234.

[24] Brenda Milner. "Intellectual function of the temporal lobes." In: *Psychological Bulletin* 51.1 (1954), pp. 42–62.

[25] Brenda Milner, Suzanne Corkin, and Hans-Lukas Teuber. "Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of H.M." In: *Neuropsychologia* 6.3 (Sept. 1968), pp. 215–234.

[26] Bruce McNaughton and Richard GM Morris. "Hippocampal synaptic enhancement and information storage within a distributed memory system." In: *Trends in Neurosciences* 10.10 (1987), pp. 408–415.

[27] John O'Keefe and Lynn Nadel. *The Hippocampus as a Cognitive Map.* Oxford: Clarendon Press, 1978.

[28] Donald O Hebb. *The organization of behavior.* na, 1949.

[29] John J Hopfield. "Neural networks and physical systems with emergent collective computational abilities." In: *Proceedings of the National Academy of Sciences* 79.8 (Apr. 1982), pp. 2554–2558.

[30] Daniel J Amit, Hanoch Gutfreund, and H Sompolinsky. “Statistical mechanics of neural networks near saturation”. In: *Annals of Physics* 173.1 (Jan. 1987), pp. 30–67.

[31] Jeffrey S Taube. “Head direction cells and the neurophysiological basis for a sense of direction.” In: *Progress in Neurobiology* 55.3 (1998), pp. 225–256.

[32] William E Skaggs et al. “A model of the neural basis of the rat’s sense of direction.” In: *Advances in Neural Information Processing Systems* 7 (1995), pp. 173–180.

[33] Robert U Mulner and John L Kubie. “The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells”. In: *The Journal of Neuroscience* 7.7 (1987), pp. 1951–1968.

[34] John L Kubie and Robert U Mulner. “Multiple representations in the hippocampus.” In: *Hippocampus* 1.3 (1991), pp. 240–242.

[35] Karel Jezek et al. “Theta-paced flickering between place-cell maps in the hippocampus”. In: *Nature* 478.7368 (Sept. 2011), pp. 246–249.

[36] Laura L Colgin et al. “Attractor-map versus Autoassociation based attractor dynamics in the hippocampal network”. In: *Journal of Neurophysiology* 104.1 (2010), pp. 35–50.

[37] Alexei Samsonovich and Bruce L McNaughton. “Path integration and cognitive mapping in a continuous attractor neural network model.” In: *The Journal of Neuroscience* 17.15 (1997), pp. 5900–5920.

[38] Francesco Battaglia and Alessandro Treves. “Attractor neural networks storing multiple space representations: A model for hippocampal place fields.” In: *Physical Review E* 58.6 (1998), pp. 7738–7753.

[39] Alessandro Treves. “The Dentate Gyrus: defining a new memory of David Marr”. In: *Computational Theories and Their Implementation in the Brain: The Legacy of David Marr*. Ed. by Lucia Vaina and Richard E Passingham. Springer, 2017, Ch 5.

[40] Rémi Monasson and Sophie Rosay. "Transitions between spatial attractors in place-cell models". In: *Physical review letters* 115.9 (2015), p. 098101.

[41] Yves Chauvin and David E Rumelhart. *Backpropagation: theory, architectures, and applications*. Psychology press, 1995.

[42] Daniel J Amit. *Modeling brain function: The world of attractor neural networks*. Cambridge university press, 1992.

[43] John C Eccles. "Synaptic and neuro-muscular transmission". In: *Physiological Reviews* 17.4 (1937), pp. 538–555.

[44] Alessandro Treves and Edmund T Rolls. "Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network". In: *Hippocampus* 2.2 (1992), pp. 189–199.

[45] Jean-Michel Lassalle, Thierry Bataille, and Hélène Halley. "Reversible inactivation of the hippocampal mossy fiber synapses in mice impairs spatial learning, but neither consolidation nor memory retrieval, in the Morris navigation task". In: *Neurobiology of learning and memory* 73.3 (2000), pp. 243–257.

[46] Inah Lee and Raymond P Kesner. "Encoding versus retrieval of spatial memory: double dissociation between the dentate gyrus and the perforant path inputs into CA3 in the dorsal hippocampus". In: *Hippocampus* 14.1 (2004), pp. 66–76.

[47] Stefan Leutgeb. "Distinct ensemble codes in hippocampal areas CA3 and CA1." In: *Science* 305.5788 (2004), pp. 1295–1298.

[48] Patricia E Sharp. "Computer simulation of hippocampal place cells." In: *Psychobiology* 19.2 (1991), pp. 103–115.

[49] Alessandro Treves, Orazio Miglino, and Domenico Parisi. "Rats, nets, maps, and the emergence of place cells". In: *Psychobiology* 20.1 (1992), pp. 1–8.

[50] Sylvia Wirth et al. "Single neurons in the monkey hippocampus and learning of new associations". In: *Science* 300.5625 (2003), pp. 1578–1581.

[51] Edmund T Rolls and Sylvia Wirth. “Spatial representations in the primate hippocampus, and their functions in memory and navigation”. In: *Progress in neurobiology* 171 (2018), pp. 90–113.

[52] Trygve Solstad, Edvard I Moser, and Gaute T Einevoll. “From grid cells to place cells: a mathematical model”. In: *Hippocampus* 16.12 (2006), pp. 1026–1031.

[53] Edmund T Rolls, Simon M Stringer, and Thomas Elliot. “Entorhinal cortex grid cells can map to hippocampal place cells by competitive learning”. In: *Network: Computation in Neural Systems* 17.4 (2006), pp. 447–465.

[54] Bailu Si and Alessandro Treves. “The role of competitive learning in the generation of DG fields from EC inputs”. In: *Cognitive neurodynamics* 3.2 (2009), pp. 177–187.

[55] Erika Cerasti and Alessandro Treves. “How informative are spatial CA3 representations established by the dentate gyrus?” In: *PLoS Computational Biology* 6.4 (2010).

[56] Erika Cerasti and Alessandro Treves. “The spatial representations acquired in CA3 by self-organizing recurrent connections”. In: *Frontiers in cellular neuroscience* 7 (2013), p. 112.

[57] Charlotte B Alme et al. “Place cells in the hippocampus: Eleven maps for eleven rooms”. In: *Proceedings of the National Academy of Sciences* 111.52 (2014), pp. 18428–18435.

[58] John O’Keefe and Michael L Recce. “Phase relationship between hippocampal place units and the EEG theta rhythm”. In: *Hippocampus* 3.3 (1993), pp. 317–330.

[59] Kenneth L Blum and Larry F Abbott. “A model of spatial map formation in the hippocampus of the rat.” In: *Neural Computation* 8.1 (1996), pp. 85–93.

[60] Kenway Louie and Matthew A Wilson. "Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep." In: *Neuron* 29(1) (2001), pp. 145–156.

[61] Brad I Pfeiffer and David J Foster. "Hippocampal place-cell sequences depict future paths to remembered goals." In: *Nature* 497.7447 (2013), pp. 74–79.

[62] Alessandro Treves. "Computational constraints between retrieving the past and predicting the future, and the CA3-CA1 differentiation." In: *Hippocampus* 14.5 (2004), pp. 539–556.

[63] Julija Krupic et al. "Grid cell symmetry is shaped by environmental geometry". In: *Nature* 518.7538 (2015), pp. 232–235.

[64] Charlotte N. Boccara et al. "The entorhinal cognitive map is attracted to goals". In: *Science* 363.6434 (2019), pp. 1443–1447. ISSN: 0036-8075.

[65] Benjamin Dunn et al. "Grid cells show field-to-field variability and this explains the aperiodic response of inhibitory interneurons". In: *arXiv preprint arXiv:1701.04893* (2017).

[66] Benjamin R Kanter et al. "A novel mechanism for the grid-to-place cell transformation revealed by transgenic depolarization of medial entorhinal cortex layer II". In: *Neuron* 93.6 (2017), pp. 1480–1492.

[67] John B Calhoun. *The ecology and sociology of the Norway rat.* US Department of Health, Education, and Welfare, Public Health Service, 1963.

[68] Alessandro Treves. "Graded-response neurons and information encodings in autoassociative memories". In: *Phys. Rev. A* 42 (4 Aug. 1990), pp. 2418–2430.

[69] Alessandro Treves and Edmund T Rolls. "What determines the capacity of autoassociative memories in the brain?" In: *Network: Computation in Neural Systems* 2.4 (1991), pp. 371–397.

[70] EunHye Park, Dino Dvorak, and André A. Fenton. "Ensemble Place Codes in Hippocampus: CA1, CA3, and Dentate Gyrus Place Cells Have Multiple Place Fields in Large Environments". In: *Plos ONE* 6.7 (July 2011), pp. 1–9.

[71] B. Harland et al. "Dorsal CA1 Hippocampal Place Cells Form a Multi-Scale Representation of Megaspace". In: *bioRxiv* (2021).

[72] Maya Geva-Sagiv et al. "Spatial cognition in bats and rats: from sensory acquisition to multiscale maps and navigation". In: *Nature Reviews Neuroscience* 16.2 (2015), pp. 94–108.

[73] Ruy Gómez-Ocádiz et al. "A synaptic novelty signal to switch hippocampal attractor networks from generalization to discrimination". In: *bioRxiv* (2021).

[74] Sandro Romani and Misha Tsodyks. "Short-term plasticity based network model of place cells dynamics". In: *Hippocampus* 25.1 (2015), pp. 94–105.

[75] Davide Spalla, Isabel Cornacchia, and Alessandro Treves. "Continuous attractors for dynamic memories". In: *eLife* 10 (Sept. 2021).

[76] Federico Stella et al. "Hippocampal Reactivation of Random Trajectories Resembling Brownian Diffusion". In: *Neuron* 102 (2019), pp. 450–461.

[77] Jonathan F. Miller et al. "Neural Activity in Human Hippocampal Formation Reveals the Spatial Context of Retrieved Memories". In: *Science* 342.6162 (2013), pp. 1111–1114. ISSN: 0036-8075.

[78] John J. Sakon and Michael J. Kahana. "Hippocampal ripples signal contextually-mediated episodic recall". In: *bioRxiv* (2021).

[79] Kwang Il Ryom et al. "Latching dynamics as a basis for short-term recall". In: *PLoS Computational Biology* (Feb. 2021).

[80] James Bennet B Murdock Jr. "The immediate retention of unrelated words". In: *Journal of Experimental Psychology* 60 (1960), pp. 222–234.

[81] Nancy C Waugh. "Presentation time and free recall". In: *Journal of Experimental Psychology* 1.73 (1967), pp. 39–44.

[82] Michelangelo Naim et al. "Fundamental Law of Memory Recall". In: *Physical Review Letters* 124 (1 2020). ISSN: 10797114.

[83] George A Miller. “The magical number seven, plus or minus two: some limits on our capacity for processing information.” In: *Psychological Review* 2.63 (1956), pp. 81–97.

[84] Nelson Cowan. “The magical number 4 in short-term memory: A reconsideration of mental storage capacity”. In: *Behavioral and Brain Sciences* 24.1 (2001), pp. 87–114.

[85] Eleonora Russo and Alessandro Treves. “Cortical free-association dynamics: Distinct phases of a latching network”. In: *Physical Review E* 85.5 (2012), p. 051920.

[86] Chol Jun Kang et al. “Life on the Edge: Latching Dynamics in a Potts Neural Network”. In: *Entropy* 19 (Sept. 2017), p. 468.

[87] Vezha Boboeva, Romain Brasselet, and Alessandro Treves. “The Capacity for Correlated Semantic Memories in the Cortex”. In: *Entropy* 20 (Oct. 2018), p. 824.

[88] Neta Haluts et al. “Professional or Amateur? The Phonological Output Buffer as a Working Memory Operator”. In: *Entropy* 22.6 (2020). issn: 1099-4300.

[89] Valentino Braitenberg. “Thoughts on the cerebral cortex”. In: *Journal of theoretical biology* 46.2 (1974), pp. 421–447.

[90] Ido Kanter. “Potts-glass models of neural networks”. In: *Phys. Rev. A* 37 (7 Apr. 1988), pp. 2739–2742.

[91] Robin Tremblay, Soohyun Lee, and Bernardo Ruby. “GABAergic interneurons in the neocortex: from cellular properties to circuits”. In: *Neuron* 91 (2016), pp. 260–292.

[92] Sandro Romani et al. “Scaling laws of associative memory retrieval”. In: *Neural computation* 25.10 (2013), pp. 2523–2544.

[93] Klaus Oberauer and Reinhold Kliegl. “A formal model of capacity limits in working memory”. In: *Journal of memory and language* 55.4 (2006), pp. 601–626.

[94] Jens Keilwagen, Ivo Grosse, and Jan Grau. “Area under Precision-Recall Curves for Weighted and Unweighted Data”. In: *Plos One* 9.3 (Mar. 2014), pp. 1–13.

[95] Klaus Oberauer et al. “Benchmarks for models of short-term and working memory.” In: *Psychological bulletin* 144.9 (2018), p. 885.

[96] Robert G Crowder. “Intraserial repetition effects in immediate memory”. In: *Journal of Verbal Learning and Verbal Behavior* 7.2 (1968), pp. 446–451.

[97] Wayne A Wickelgren. “Short-term memory for repeated and non-repeated items”. In: *Quarterly Journal of Experimental Psychology* 17.1 (1965), pp. 14–25.

[98] Richard NA Henson. “Item repetition in short-term memory: Ranschburg repeated.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24.5 (1998), p. 1162.

[99] Mark J Hurlstone, Graham J Hitch, and Alan D Baddeley. “Memory for serial order across domains: An overview of the literature and directions for future research.” In: *Psychological bulletin* 140.2 (2014), p. 339.

[100] Haim Sompolinsky and Ido Kanter. “Temporal association in asymmetric neural networks”. In: *Physical review letters* 57.22 (1986), p. 2861.

[101] Serena Di Santo et al. “Working Memory Training: Assessing the Efficiency of Mnemonic Strategies”. In: *Entropy* 22 (May 2020), p. 577.

[102] James Deese. “Influence of Inter-Item Associative Strength upon Immediate Free Recall”. In: *Psychological Reports* 5.3 (1959), pp. 305–312.

[103] Henry Roediger and Kathleen McDermott. “Creating False Memories: Remembering words not presented in lists”. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21 (July 1995), pp. 803–814.

[104] Vittorio M Iacullo and Francesco S Marucci. “Normative data for Italian Deese/Roediger-McDermott lists”. In: *Behavior research methods* 1.48 (2016), pp. 381–389.

[105] Michael A Stadler, Henry L Roediger, and Kathleen B McDermott. “Norms for word lists that create false memories”. In: *Memory & Cognition* 27 (1999), pp. 494–500.

[106] Emmanuelle Pardilla-Delgado and Jessica Payne. “The Deese-Roediger-McDermott (DRM) Task: A Simple Cognitive Paradigm to Investigate False Memories in the Laboratory.” In: *Journal of visualized experiments : JoVE* 119 (2017).

[107] David Reser et al. “Australian Aboriginal techniques for memorization: Translation into a medical and allied health education setting”. In: *PloS one* 16.5 (2021), e0251710.

[108] William F Brewer and David A Dupree. “Use of plan schemata in the recall and recognition of goal-directed actions.” In: *Journal of Experimental Psychology: Learning, Memory and Cognition* 9 (1983), pp. 117–129.

[109] Vanessa E. Ghosh and Asaf Gilboa. “What is a memory schema? A historical perspective on current neuroscience literature”. In: *Neuropsychologia* 53 (2014), pp. 104–114. ISSN: 0028-3932.

[110] Dorothy Tse et al. “Schemas and Memory Consolidation”. In: *Science* 316.5821 (2007), pp. 76–82. ISSN: 0036-8075.

[111] Donald Norman and Tim Shallice. “Attention to Action: Willed and Automatic Control of Behavior”. In: *Consciousness and Self-Regulation: Advances in Research and Theory IV*. Ed. by R. Davidson, R. Schwartz, and D. Shapiro. Plenum Press, 1986.

[112] Shelley E. Taylor and Jennifer Crocker. “Schematic bases of social information processing”. In: 1 (Jan. 1981), pp. 89–134.

[113] Marlieke Van Kesteren et al. “How schema and novelty augment memory formation”. In: *Trends in Neurosciences* 35 (2012), pp. 21–219.

[114] Sara Andreetta et al. “In Poetry, if Meter has toHelp Memory, it Takes its Time”. In: *Open Research Europe* (2021).

[115] Matthew A Killingsworth and Daniel T Gilbert. “A wandering mind is an unhappy mind”. In: *Science* 330.6006 (2010), pp. 932–932.

[116] Marcus E Raichle et al. “A default mode of brain function”. In: *Proceedings of the National Academy of Sciences* 98.2 (2001), pp. 676–682.

[117] Malia F Mason et al. “Wandering minds: the default network and stimulus-independent thought”. In: *Science* 315.5810 (2007), pp. 393–395.

[118] Jonathan Smallwood and Jonathan W Schooler. “The science of mind wandering: empirically navigating the stream of consciousness”. In: *Annual review of psychology* 66 (2015), pp. 487–518.

[119] Elena Bertossi and Elisa Ciaramelli. “Ventromedial prefrontal damage reduces mind-wandering and biases its temporal focus”. In: *Social cognitive and affective neuroscience* 11.11 (2016), pp. 1783–1791.

[120] Cornelia McCormick et al. “Mind-Wandering in People with Hippocampal Damage”. In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 11.38 (2018), pp. 2745–2754.

[121] Elisa Ciaramelli and Alessandro Treves. “A Mind Free to Wander: Neural and Computational Constraints on Spontaneous Thought”. In: *Frontiers in Psychology* 10 (Jan. 2019).

[122] Asaf Gilboa and Hannah Marlatte. “Neurobiology of schemas and schema-mediated memory”. In: *Trends in cognitive sciences* 21.8 (2017), pp. 618–631.

[123] Marco Marelli. “Word-Embeddings Italian Semantic Spaces: a semantic model for psycholinguistic research”. In: *Psihologija* 50 (Oct. 2017), pp. 503–520.

[124] Maria Montefinese et al. “The adaptation of the Affective Norms for English Words (ANEW) for Italian”. In: *Behavior Research Methods* 46 (Oct. 2013), pp. 887–903.

[125] Elisa Ciaramelli, Filomena Anelli, and Francesca Frassinetti. “An asymmetry in past and future mental time travel following vmPFC damage”. In: *Social cognitive and affective neuroscience* 16 (3 2021). ISSN: 17495024.

[126] M. Finkelberg. *The Homer Encyclopedia, 3 Volume Set*. Wiley, 2011. ISBN: 9781405177689.

[127] Almut Hintze. “On the literary structure of the Older Avesta”. In: *Bulletin of the School of Oriental and African Studies* 65.1 (2002), pp. 31–51.

[128] A Baddeley. “Working memory”. In: *Science* 255.5044 (1992), pp. 556–559. ISSN: 0036-8075.

[129] Gijsbert Stoet. “PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments”. In: *Teaching of Psychology* 44.1 (2017), pp. 24–31.

[130] Jonathan Peirce et al. “PsychoPy2: Experiments in behavior made easy”. In: *Behavior Research Methods* 51 (2019), pp. 195–203.

[131] Najme Dorri, Mahdi Pourmohammad, and Abdul Majid Ziaratzade. “Oral-Oriented Teaching of Meter through the Use of Music: Proposing a New Method of Teaching Meter in Poetry”. In: *Literary Arts* 12.4 (2020), pp. 73–96. ISSN: 20088027.

[132] Øyvind Arne Høydal et al. “Object-vector coding in the medial entorhinal cortex”. In: *Nature* 568.7752 (2019), pp. 400–404.