UNIVERSITY OF CALIFORNIA, SAN DIEGO

Vision and Imagery:
the Role of Cortical Attractor Dynamics

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Cognitive Science

by

Thomas P. Rebotier

Committee in charge:

Professor David J. Kirsh, Chair
Professor Martin I. Sereno, Co-Chair
Professor Jeffrey L. Elman
Professor Henry D.I. Abarbanel
Professor Patricia S. Churchland
Professor Donald I. MacLeod

1998

Copyright
Thomas P. Rebotier, 1998
All rights reserved.

The dissertation of Thomas P. Rebotier is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

1998

This dissertation is dedicated to my parents, Jean-Jacques and Jacqueline Rebotier.

TABLE OF CONTENTS

Signature Page iii

Dedication iv

Table of Contents v

List of Figures viii

List of Tables xvii

Vita, Publications, and Fields of Study xx

Abstract xxi

I General Introduction 1

II "Imageries not Imagery" 5
A. Introduction: Is white a color? 5
B. Vision and Imagery 6
1. The visual system as a hierarchy of representations 6
2. The altered story: interactive vision 11
3. Top-down effects in the visual system: Steve Kosslyn's account of imagery 13
C. Attractor classification, Bistability and Imageries 22
1. Attractor networks 23
2. Consequences of bistability and of pattern completion 43
3. What are imageries? – Imagery at different levels 45
4. The interplay of imagery and perception 47
5. Consequences and predictions 48
D. Imagery debates revisited 48
1. Variety of imagery types 48
2. Where in the brain ? : the neuropsychological debate 53
3. Imagery and reasoning : no longer a philosophical debate 62
E. Conclusion: traps and detours of introspection 68

III A Minimal Model of Mental Imagery by Recurrence-Induced Pattern Reconstruction 70
A. Introduction 70
1. Historical background 71
2. Our model 78
B. Methods 80
1. Architecture of the model: logical components 80
2. Architecture of the model: logical connections 82
3. Architecture of the model: actual implementation 82
4. Computation and averaging of the results 87
C. Results 92

1. Short term memory by reverberation 92
2. Imagery by pattern reconstruction 94
3. Non-linear transfer functions result in threshold implementation 96
4. Dependance of threshold and imagery on recurrence gain 97
5. Synergy between perception and imagery 101
6. Samples of imaged pictures 101
D. Discussion 104
1. Summary of results 104
2. Scope and limits of this model 107
3. What we believe the model will be robust to 117
4. What we believe the model won't be robust to 118
5. Merging older models to get pattern-specific dynamics 118
6. What this model suggests for the phenomena of mental imagery 119
7. Computational power of less-than-total connectivity: the storage-computation trade-off 120
8. Conclusion: Classification starts in V1 122
E. Appendices 123
1. Appendix 1: A family of transfer functions 124
2. Appendix 2: The one-unit closed system 124
3. Appendix 3: Detailed parameters of the model 131

IV Content-Specific Low Level Interaction between Imagery and Perception . . . 132
A. introduction 132
1. Historical Background 134
2. Goals of the experiment 151
B. Methods 152
1. Overview of the experiment 152
2. Preliminary: Reasons for the choice of task and stimuli 153
3. Subjects 155
4. Hardware, software, and physical setting 155
5. Stimuli 155
6. Task, response and dependent variable 158
7. Factors 159
8. Blocking and order of trials 160
9. Pretests 161
10. Interview 163
11. Timeline of experiment 164
12. Facilitation was the analyzed variable 165
C. Results 165
1. Preliminary results: pretests, base RTs, error rates 165
2. Main result: imagery facilitates the task even more than vision 173
3. Fatigued subjects have less facilitation 174
4. Gender effects 175
5. Response effects 175
6. Relation between effects and the base reaction times 185
7. Additional analysis of the two cases of 'NO' 185

D. Discussion and Conclusion 191
1. Summary of Results 191
2. Discarding the after-image hypothesis 191
3. Just seeing the congruent cue induces no priming 192
4. Imagery actively interacts with vision 192
5. What is the balance between facilitation and inhibition? 194
6. Hemispheric difference 197
7. Conclusion 198
E. Appendices 199
1. Appendix 1: Clock pretest 199
2. Appendix 2: Direction pretest 203
3. Appendix 3: Instructions for the Letter pre-test 205
4. Appendix 4: Instructions for the practice 205
5. Appendix 5: Instructions for the VISION block if it comes first 206
6. Appendix 6: Instructions for the IMAGERY block if it comes second . 207
7. Appendix 7: Instructions for the INTERLEAVED block 207
8. Appendix 8: Instructions for the IMAGERY block if it comes first . . . 208
9. Appendix 9: Instructions for the VISION block if it comes second . . . 208
10. Appendix 10: Instructions for the BASELINE 209
11. Appendix 11: Why D' must not be computed from averaged rates . . . 209

V General Conclusion 211

VI Résumé en français – Summary in French 217
A. Première partie: Théorie 219
1. Vision et imagerie 219
2. Attracteurs, complétion et bistabilité 219
3. Les imageries mentales 221
B. Seconde partie: Modèle 222
1. Le modèle 223
2. Résultats 224
3. Discussion et conclusion 225
C. Troisième partie: Expérience 226
1. Abrégé des méthodes 228
2. Abrégé des résultats et discussion 229
D. Conclusion 231

Bibliography 233

LIST OF FIGURES

II.1 Determination of connection type according to origin and termination . . 7
II.2 Hierarchichal organisation of the macaque visual system 8
II.3 Outline of Kosslyn's analysis of the visual system 14
II.4 Details of Kosslyn's functional analysis of the visual system 19
II.5 Phase space of attractor networks 23
II.6 "Happy" v.s. "frustrated" connections 24
II.7 Dynamics of Hopfield nets 25
II.8 Stability analysis of Wilson and Cowan's first model 29
II.9 Hysteresis in Wilson and Cowan's first model 30
II.10 The timecourse of an active transient 32
II.11 The evolution of edge enhancement 33
II.12 Propagation of activity waves 34
II.13 Stimuli used by Shepard and Metzler 65
II.14 Stimuli used by Lynn Cooper 66
II.15 The Treasure Island map of Steve Kosslyn 67

III.1 Architecture of the model 81
III.2 Profile of the Gabor wavelet functions used 83
III.3 Filter used in the preliminary work 84
III.4 Mechanism of filtering 85
III.5 Images used in the simulations 89
III.6 Static memory 93
III.7 Imagery 95
III.8 Perceptual threshold 98
III.9 Imagery Threshold 99
III.10 Thresholds decrease with lateral connection strength 100
III.11 Imagery and Perception of the same image combine 102
III.12 Imagery and perception of different images do not combine 103
III.13 Imaging the Mona Lisa 105
III.14 Imaging the line images 106
III.15 A type of illusion that attractor dynamics can explain 110
III.16 Transfer functions 123
III.17 System made of a single neuron with feedback 125
III.18 Construction of the analysis 126
III.19 Theoretical threshold as a function of gain 127

IV.1 Illusions replicated with an imaged context 143
IV.2 Timeline of a trial 152
IV.3 CUE 156
IV.4 STIMULI 157
IV.5 Factors of the experiment 161
IV.6 Different cases of trials 162
IV.7 Experimental schedule 164
IV.8 Effect of the cue type 174

IV.9 Effect of fatigue 176
IV.10 Effect of fatigue split by cue type 177
IV.11 Interaction of gender, cue type and hand used 178
IV.12 Effect of response 179
IV.13 Main effect of hand used 180
IV.14 Interaction between hand used and response 181
IV.15 Three-way interaction between cue type, response and hand used 183
IV.16 Base reaction times by cue type, response and hand used 184
IV.17 Absence of correlation between base RT and facilitation 186
IV.18 Absence of within-subject correlation between base RT and facilitation . 187
IV.19 Facilitation per detailed categories of response 189
IV.20 Response X hand used X cue type interaction with detailed categories of response 190

LIST OF TABLES

III.1 Constants and Parameters used for the simulations 131

IV.1 Advantage of being a good tester for overall facilitation 166
IV.2 Advantage of good testers specifically for imagery 166
IV.3 Relation between pretests scores 168
IV.4 Facilitation scores for good and bad scorers 169
IV.5 Effects observed on basic reaction times 170
IV.6 Error rates 171

ACKNOWLEDGEMENTS

This dissertation was made possible by the help and support of a great many people. In particular, I would like to thank David Kirsh, for his friendship, his advice, his intelligence and his patience. I am also very indebted to other members of my committee. Marty Sereno's advice has always borne the stamp of his very sharp understanding of deep related issues. Jeff Elman's help was always very to-the-point and strategically helpful. Don MacLeod has helped me with several experimentally relatted issues. Pat Churchland has been supportive, and her own ideas have helped nurture the fertile ground in which this dissertation has grown. Henry Abarbanel deserves special recognition for his courage and good will to plunge deeply into a topic far from his own research.

I am also deeply ingratiated to other scientists in San Diego and in France. Laraine McDonough has dedicated considerable time to helping me with the experimental design and issues. Likewise, Aaron Cicourel has left discreet but decisive footsteps in this dissertation – he will know where to find them. The Fellows from the McDonnell-Pew center for Cognitive Neuroscience have over and over given me ample feedback on my ideas, models and experiments – Thank you all! Jacques Droulez and Andrei Gorea, in France, have provided very instructive feedback on several occasions. Daniel Andler, Michel Imbert and Alain Berthoz have also helped me greatly in my scientific evolution, and on occasion with comments about this work. Several former students from the "DEA de Sciences Cognitives" have also been supportive and helpful, among others, Christel Agonie, Fréderic Lavigne, Catherine Marlot, Gérard Roszavogyi, Hugo Berroyer, Jacques Lautrais.

I am grateful to the McDonnell-Pew foundation for the fellowship that made this dissertation possible, and for animating a great scientific melting pot in San Diego. The McDonnell-Pew foundation, the board, the fellows, and our organizer Rosemary Miller have all participated to this wonderful environment.

My family and friends have provided unconditional support and love throughout

these six years of graduate school. My parents, Jean-Jacques and Jacqueline, to whom this dissertation is dedicated, have been wonderfully patient. I love you mom! My brothers Blaise and Maxime have been very supportive. Numerous friends in France have stayed in touch and encouraged me. To my many surfing friends and roommates, Mark, Peter, Brett, Sabo, Josh-bro, the two Jacks, all the Cabo bros, Poole Street and Shores crews, the stormy Cohn twins, Scott and the Christian crew, all the girls, Charlene, Jen and many others, thank you! There is a lot of warmth and talent on the San Diego waves... All grad students in my class, now slowly, dispersing, Deborah, Mike, Marissa, Scott, Lucy, Dan, Rita, Jon, thank you too! Finally, thanks to all in the cogsci department, faculty, students and staff – it has been a wonderful six years.

VITA

July 2, 1964	Born, Boulogne-Billancourt, France
1981	Baccalauréat, série C, mention bien, Académie de Versailles
1981-1983	Ecole Sainte Geneviève, Versailles
1986	Ingénieur civil des Mines, Ecole des Mines de Paris
1986–1987	Ensign, French Navy
1987–1990	Software engineer and consultant, senior consultant, McDonnell-Douglas Information Systems, Paris
1990–1992	Software consultant, Banque Indosuez, Paris
1991–1992	D.E.A. de Sciences Cognitives, University of Paris VI, EHESS.
1992–1993	Teaching Assistant, University of California, San Diego
1993–1996	Research Assistant, University of California, San Diego
1996–1998	McDonnell-Pew Fellow, University of California, San Diego
1998	Doctor of Philosophy University of California, San Diego

PUBLICATIONS

"Implementation of Dynamic Memory Models and Multiplicative Laws with Spiking Neurons", T Rebotier, J Droulez and A.Berthoz, submitted to Biological Cybernetics.

"Explorations with the Dynamics Wave Model" T Rebotier and J Elman, published at NIPS 95

"Multiplicatory spiking Neurons in a Dynamic Memory Net: a two-level Model" T Rebotier, J Droulez and A.Berthoz, presented at the symposium "Dynamics of Neural Processing" (Washington DC 1994).

"Sigma vs Pi Properties of Spiking Neurons" T Rebotier and J Droulez, in the Connectionnist Modeling Summer School proceedings, 1993.

ABSTRACT OF THE DISSERTATION

Vision and Imagery:
the Role of Cortical Attractor Dynamics by

Thomas P. Rebotier
Doctor of Philosophy in Cognitive Science
University of California, San Diego, 1998
Professor David J. Kirsh, Chair
Professor Martin I. Sereno, Co-Chair

Mental imagery presents individual differences and degrees of vividness which appear discrete. We suggest that visual brain areas follow an attractor dynamics giving them two properties: pattern completion and bistability. Bistability implies that visual brain areas are entirely active or inactive. Vision and imagery use the same system, vision being bottom-up and imagery top-down input. In imagery, different configurations of active visual areas account for different imagery types. The more vivid the imagery, the more occipital areas are recruited. The dissertation comports three parts. The first part details this theory. The second part presents a computer model where a layer of neurons represent an image by wavelet (Gabor) components. Lateral connections make that layer an attractor network. The properties observed support the statements made above about attractor networks: The model is bistable, has an activation threshold, can improve the shape of a noised pattern (restoration by pattern completion), and can combine top-down and bottom-up sub-thresholds inputs to reach activation. The third part is a psychological experiment testing this prediction on how imagery interacts with vision: when imagery is going on, low-level visual processes are more efficient if the visual image is similar to the mental image. 63 subjects completed 160 trials each. In 80 trials, they are cued by a visualization, and must afterwards make a visual judgement on a clock image: they decide if both hands of the clock are in a shaded plane which covers half of the image. When visualizing the right clock, subjects are faster than when visualizing the wrong clock. A control condition (80 trials) used a visual rather

than imaged cue (Previous studies lacked controls for acquisition strategies, such as eye movement planning or attention allocation). The effect is smaller with the visual cue, and restricted to a particular sub-case where one expects acquisition strategies. This experiment demonstrates that the influence of imagery on perception depends on the content of the image. It is an interaction between mental images and visual images, rather than an interaction between "systems". Besides this main finding, lateralization results confirm the growing hypothesis that composite imagery is strongly lateralized.

Chapter I

General Introduction

"Lovers and Madmen have such seething brains,
Such shaping fantasies, that apprehend
More than the cool reason ever comprehends.
The lunatic, the lover, and the poet
Are of Imagination all compact.
One sees more devils than the vast hell can hold.
That is the madman. The lover, all as frantic,
Sees Helen's beauty in a brow of Egypt.
The poet's eye, in a fine frenzy rolling,
Doth glance from heaven to earth, from earth to heaven.
And as imagination bodies forth
The form of things unknown, the poet's pen
Turns them to shapes and gives to airy nothing
A local habitation and a name.
Such tricks hath strong imagination..."

William Shakespeare, *A Midsummer Night's Dream*, V.i.

Imagery dwells deep within us. It is such a permanent feature of our mental landscape that we often forget it is there. Imagery plays a major role in arts, in rhetoric and, for some of us, in problem solving. It can be tactile, kinaesthetic, auditive or visual, gustatory or olfactive. We can image pleasure as well as pain. Imagery is the medium of fantasy, of phantasms, and of meditation. Imagery can change history. Countless great characters of history or fiction, from Saint Paul to Lady Macbeth to Albert Einstein,

owe their fame –or fate– to a vision, sometimes sudden and vivid, or, to the contrary, slowly assembled and meditated upon. Truly, the ability to shape a mental image is part of our species' best heritage.

Thus, it is no surprise that Imagery played a foremost rôle in early psychology. Wilhelm Wundt, Francis Galton, William James, and many others, devoted an important part of their opus to imagery. For introspectionists, imagery felt like vision even though mental images were distinct in quality from visual images. This paradox made the topic even more spicy. In addition to the ambiguous relation of visual and mental images, strong individual differences in the vividness of mental images soon became undeniable (Galton, 1907). From person to person, mental images can be as sharp as photographs, or as vague as abstract concepts. Likewise, from the pre-dawn dream period to the late morning hustle, mental image quality and accessibility can briskly decrease. All these variations could be reported, but not explained. Mental Imagery, explored and on its way to elucidation, could be used as a cornerstone for theories of how the mind operates.

William James, an indefatigable empiricist, was quite happy with the classification of different imagery types. He considered that "a blurred picture is just as much a single mental fact as a sharp picture is". Thus, vague images could still be solid building blocks of mental life. Individual differences showed the existence of several distinct levels, if not distinct faculties, of imagery. There were several "imageries" not one "Imagery".

Donald Hebb (1968), in our view, took the next step forward. He suggested that the different imageries corresponded to re-activation at different levels in the cortex, vivid images being produced by the more occipital re-activations. Written well before the various maps of the visual cortex were described, his description was perforce more abstract than can be done today. Yet, his notion of re-activation of *cell assemblies* also carried as a seed the concept of attractor network. Hebb's visionary concept lacked the connection to neuropsychology on one end, and to mathematics on the other, but nevertheless was sitting at the right spot: different "imageries" could be implemented by the same system.

Our vision is truly an extension of Hebb's, but before we introduce it, we must present two more strands to the braid of history.

One strand is the outstanding development of brain science, and especially of

visual system science, that has occurred since Hebb's time. Although we are still far from the big picture, extensive inroads have mapped visual areas of the brain, and are well on the way to listing the functional properties of the neurons and interneurons of at least the most occipital areas. Functional Magnetic Resonance Imaging is now available to compare the organization of living human cortex to that of more easily studied macaques. Available techniques for all kinds of neuroscience have advanced by orders of magnitude. All this progress makes it possible to present partial theories of the visual system.

The second strand is the occurrence, in the seventies and eighties, of an *imagery debate* which polarized the issues around the confrontation of two mighty opponents, Steve Kosslyn and Zenon Pylyshyn. Pylyshyn had detected a trap in the introspectionist reasoning: if a vague mental image was a sharp mental fact, what prevented any mental fact to be called a vague mental image? He argued that mental images were really not images, but symbolic descriptions. Kosslyn defended the position that mental images had a spatial layout, implemented in the brain by pictorial visual areas. The final word belongs to Kosslyn, but owes much to Pylyshyn also.

An important consequence of this debate and of the progress in neurosciences is that Kosslyn (1994) has merged in one theory mental imagery and top-down effects of the visual system. With the details to be filled in, or possibly changed, he has said **what** imagery is: top-down effects, *à vide,*in the visual system, and **where** it is implemented: in some areas of occipital cortex. Kosslyn's theory needs only to be supplemented on two points: **how** re-activation takes place, and **why** there is such a range of individual differences.

By formalizing Hebb's theory with modern instruments, we hope to suggest an answer to these questions. A mechanism for imagery, a reason for individual differences: both of these are provided by two basic properties of attractor networks: bistability, and pattern completion.

In our first chapter, purely theoretical, we look at the consequences of having bistable brain areas. If brain areas had a continuous range of activation, objects could be at any level of activity between full representation and no representation. In contrast, bistability rules out semi-representation. Each dimension of representation, such as color, details of shape, texture, etc., must be wholly included or wholly excluded from

the current configuration of the representation. Thus the configuration varies. Like a swiss army knife, visual representations can present different shapes. Depending on the individual, on his or her state of ebriety or drowsiness, on the intensity of top-down feedback, and other factors, some areas will be engaged and some not. Different types of imagery, or of vision, result from these different configurations. Thus bistability accounts for the variety of imagery.

As a bonus, the attractor network dynamics which implements bistability also gives us pattern completion, now accounting for the "how" of imagery.

In the second chapter, we present a computer model illustrating these ideas. An archetypal "low-level brain area" is simulated by a layer of "neurons" representing an image by its wavelet (Gabor) components. Because of lateral connections, that layer is an attractor network. It demonstrates several desirable properties, such as having a threshold of activation, being bi-stable, and improving the shape of noisy initial activity patterns. The key elements of the model are discussed and their realism makes it likely that the cortex engages in similar behavior .

In the third chapter, we test experimentally a strong prediction of the theory. Imagery and perception should interact at a low level, and this interaction should depend on the similarity between the sensorial image and the mental image. We indeed find that similar contents of mental and sensorial images lead to faster response than dissimilar contents. As a control condition for imagery, we use previous perception. This control discards many alternative explanations of our conclusions that other experiments had neglected. In addition to the low-level effect, we also find a high level effect which is strongly lateralized – a cute finding, considering that imagery of composite objects such as our stimuli is supposed to be strongly lateralized. This successful experiment concludes our work.

Chapter II

"Imageries not Imagery"

II.A Introduction: Is white a color?

Rarely do authors admit to a lie in their first sentence, but we have to, because this paper is actually a theory of the visual system. However, the most portent consequences of this theory concern mental imagery. Hence, past the most general considerations, we will focus on mental imagery rather than vision.

We propose that whole visual areas in the brain are bistable. They can be turned on or off, depending on the brain state and on their input. This makes the visual system a kind of swiss army knife: it can have many different configurations of active tools. Let us consider, as an example, the *color* tool. It is active most of the time, and overused if one visits a painting museum. But, should we take a walk under the moonlight, or watch a Charlie Chaplin movie, this "color tool" is "folded" in, unused, for we are then no longer aware of color. To the old Zen koan "Is white a color?" we can finally answer "It depends which blade is pulled"! If our color subsystem is active, it will represent white spots of the image as spots colored in white. If it is inactive, color no longer exists for us. As for color, so for details of shape, shape itself, movement, texture, and probably many other attributes of a scene. Bistability gives the visual system a variable geometry, where some of these attributes are used, and some are not.

The flexibility of the visual system shows its full measure in mental imagery. Depending on the individual and on its state of consciousness, mental images will be formed with very different attributes because different brain areas are recruited. In

particular, the occipital areas, which encode the spatiotopic attributes of images, appear to be much more fickle than the temporal areas where representations are non-pictorial. Thus, there is a variety of imageries, ranging from Luria's (1968) patient, Viennamin, whose imagery was so vivid that he could not tell it apart from reality, to the famous writer Aldous Huxley (1957), who reports that even when trying mescaline, he was impervious to the very vivid imagery that the drug usually causes. For imagery even more than for vision, the metaphor of the swiss army knife describes accurately the variety of possible phenomena.

In this chapter, we explore this theory in three main sections. The first section is dedicated to a presentation of the visual system and of Kosslyn's (1994) view of imagery, now the most commonly shared in cognitive neuroscience. The second part presents dynamical systems, in particular fixed point systems, with their properties, including pattern completion and bistability. Applying the bistability property to visual areas, it delineates the principal tenets of our theory. The third part considers several fields of debate or of enquiry related to imagery, and works out the contribution that our theoretical stance brings to these questions.

II.B Vision and Imagery

II.B.1 The visual system as a hierarchy of representations

Before we proceed towards further analysis, it is advisable to briefly review principles of organization in the primate visual cortex. Most of what is known at the physiological level comes from studies on monkeys, especially on the macaque –but seems to be extensible to man. In this section, we will draw heavily on the review by Daniel Felleman and David van Essen (1991). We will also outline the relevant subcortical pathways.

Cortical areas boundaries are defined by a variety of means, both anatomical and physiological. The main anatomical method is the study of cell types and cell type distributions – the "cytectonic architecture" of cortex. The famous Brodmann classification is based on this method. Anatomical studies can also focus on the different patterns of innervation. Physiological recordings can delimit topographic maps by looking at

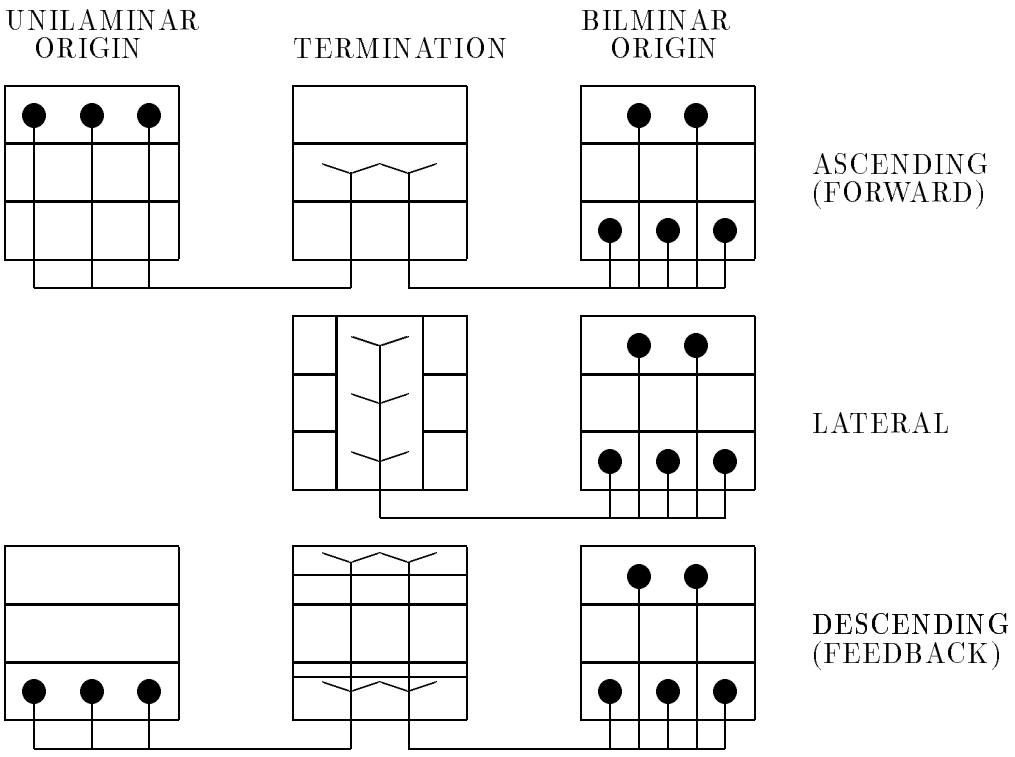

Figure II.1: **Determination of connection type according to origin and termination:** ascending connections terminate in layer 4, descending connections avoid specifically layer 4, and lateral connections innervate homogenously the whole microcolumn. (Adapted from Felleman and van Essen, 1991)

radical changes in receptive field locations of nearby neurons. Collecting many results, Felleman and van Essen (1991) differentiate 32 visual areas of which 25 are exclusively or predominantly visual. Extrapolating on present positive and negative evidence, Felleman and van Essen estimate that 40 % of the possible connections between these areas are actually implemented. They then proceed to analyze the relation between these areas.

Felleman and van Essen organize the visual cortex hierarchically by identifying ascendant, lateral and descendant pathways between cortical areas (see Figure II.1). Pathway types depend on their termination patterns in the six cortical layers. Ascendant pathways typically terminate in layer 4, just like the sensorial input in primary cortex;

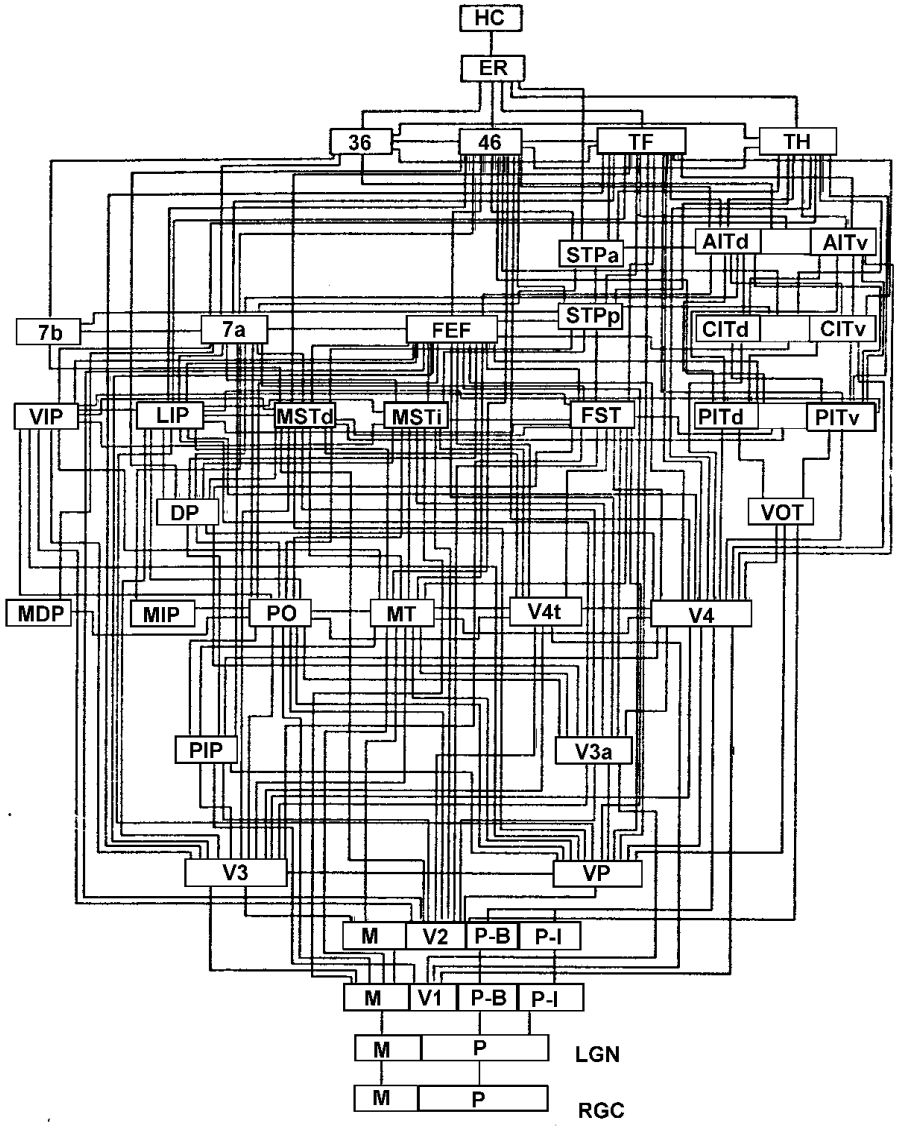

Figure II.2: **Hierarchichal organisation of the macaque visual system:** This hierarchy of the visual system, obtained mostly from anatomical methods, raises two kind of questions: first, what is the encoding in each area; second, how much of the processing is parallel and how much is serial? (From Felleman and van Essen, 1991).

they originate mostly from layers 2/3 or from both 2/3 and 5/6. Lateral pathways terminate diffusely through a whole microcolumn of cortex, synapsing into layers 2-6 including layer 4; they tend to originate from both superficial and deep layers (2/3 and 5/6). Descendant pathways avoid layer 4, usually terminating in both 2/3 and 5/6, and can originate either in deep layers (5/6) or in both deep and superficial layers (2/3 and 5/6). Not all links are perfectly definable. For example, the projection of MT to V4 has some patches with multilaminar terminations ("descendant") and some patches of columnar termination ("lateral"). Most connections are reciprocal, but there are a few documented cases of unidirectional connection, or of much weaker feedback connections, such as that from V4 to V1. Out of 80 links for which all information is available, 65 have clear ascendant/descendant antagonism, 5 are lateral/lateral links, and only 10 appear to be irregular, violating the strict hierarchical principle. Thus the hierarchical organization of visual cortex presented in Figure II.2 seems to be robust, although not systematical.

Concerning the physiological and functional properties of cells, evidence is far less detailed. It is clear that occipital visual areas (V1, V2, V3, V4, MT) are retinotopically organized and have relatively small receptive fields, whereas higher areas, whether ventral or dorsal, have much bigger receptive fields. Upstream of the visual cortex, the Lateral Geniculate Nucleus cells have radial receptive field and present no evidence of orientation (Imbert & Buser, 1987). V1 includes cells sensitive to absolute wavelength (Zeki, 1993), to segments in different orientations, to segment terminations (Hubel & Wiesel, 1977), and to motion gradient in different orientations. V2 includes cells which are sensitive to illusory contours (von der Heydt, Peterhans & Baumgartner, 1984). V4 cells can discard illumination to become sensitive to relative color; their physiological response is modulated by attention (Moran & Desimone (Moran & Desimone, 1985; Haenny & Schiller, 1988; Spitzer, Desimone & Moran, 1988) . MT cells solve the aperture problem for movement, still present in V1. The size of receptive fields augments from 0.5-0.9 degrees in V1 to 2-4 degrees in V4 to at least 25 degrees in IT (Moran & Desimone, 1985). Likewise, in the dorsal stream, receptive fields of area MT are about 2-4 degrees, and those of area 7a are over 20 degrees. Temporal areas neurons are also modulated by attention (Desimone & Duncan, 1995); they are relatively invariant with

size, orientation and retinal position, and respond preferentially to very complex stimuli (Miyashita, 1993). Thus, whereas it appears that different areas do implement different representational spaces, less and less retinotopic as one goes along the ventral pathways, these representational spaces are still to be defined.

Finally, the cortex should not be considered in isolation, especially when discussing the functions of the dorsal stream. In the brain stem, the Superior Colliculus is involved in spatial orientation and eye movement, in close connection with parietal areas and premotor frontal cortex. There is a strong feedback loop between the thalamus (Striatum) and the cortex, and selective spatial attention could involve selection by the reticulate nucleus which wraps around the thalamus (Desimone & Ungerleiter, 1990); such mechanisms could play a role in spatial components of imagery.

This synthetic organization of cortical studies is more easily made for the macaque than for the human, but recent techniques have allowed similar delineation of low-level human cortical areas. While classical studies of cortex were limited to post-mortem anatomical studies or lesion studies[1], modern imaging techniques are providing more and more non-invasive ways of studying visual cortex in action. PET studies started the charge[2], and were soon supplemented by FMRI studies using new ways of determining area boundaries[3]. Until this ongoing exploration work is completed, it is difficult to give definitive conclusions, but at this it appears that most differences are in the relative size of areas[4] rather than in their functional rôle[5] and organization. Because the theory presented in this paper does not depend on details which appear to change from macaque to human, we do not detail these differences any further.

From our point of view, the system proposed by Felleman and van Essen has

[1] In vivo stimulation during brain surgery hardly ever reaches occipital lobes because most epileptic focii are located in the temporal lobe, and occipital cortex is not exposed during the surgery. See Horowitz (1978), chapter 11.

[2] See, for example Zeki (1993)

[3] Adjacent cortical areas often feature alternated mirror symmetry of the retina. This fact, already used in mass single recording organizing techniques (Sereno, McDonald & JM, 1994), can be used with FMRI because the field sign can be determined from the cortical response to rotating visual stimuli and dilatating/contracting stimuli (Sereno, Dale, Reppas, Kwong, Belliveau, Brady, Rosen & Tootell, 1995).

[4] See Tootell, Dale, Sereno and Malach (1996) for a review of area delimitation and organization.

[5] For example, ?), shows strong resemblance between human and Monkey V1.

one essential property: the functional differentiation of visual areas. Other properties, and especially the degree of seriality in the hierarchical organisation are irrelevant to our theory. We will see in the next section that these properties are indeed questioned. But the functional division, by which different areas encode different properties, features, or "representational dimensions" of the stimulus, is essential for our theory. As we will see later, that division allows imagery to take different configuration. For example, when V5/MT is used, imagery will be in color, because V5/MT is the area that encodes color of objects.

II.B.2 The altered story: interactive vision

The orthodox view supported by Felleman and van Essen's synthesis of cortical organization is more and more called into question. We will review here more recent viewpoints moderating the picture sketched above, and in particular other anatomical studies (Bullier & Nowak, 1995; Hilgetag, O'Neil & MP, 1996), and a synthetic viewpoint (Churchland, Ramachandran & Sejnowski, 1994).

Felleman and van Essen's conclusions have been challenged, on anatomical grounds, over two points. First, the exact hierarchical tree is a compromise. Much like the optimal path of a traveling salesman, the tree result from a choice amongst very numerous possibilities. Hilgetag, O'Neil and MP (1996), using the constraints listed by Felleman and van Essen, find that several different trees can be computed with an equal fitness. van Essen and Felleman (1996) comment that most likely, the cortex is a "quasi-hierarchy". Second, the hierarchy hints that information is sequentially processed from the lowest to the highest levels. Recent studies, summarized by Bullier and Nowak (1995), suggest that the processing is a lot more parallel than previously thought. For example, firing latencies (the minimal time for a cell to fire after stimulus presentation) present an important overlap between "consecutive" areas. Also, lower areas operate figure-ground segmentation on the basis of information that they are not supposed to encode, for example, the global direction of motion (Reppas, Niyogi, Dale, Sereno & RBH, 1997). So, at this point, the global picture of the visual cortex is not yet sharp.

On a more general take, Churchland, Ramachandran and Sejnowski (1994) point out the fallacies of the orthodox model of "pure vision". The conception that

there is a "pure vision" is built on three main assumptions:

- that we maintain a full representation of the visual scene,
- that visual subsystems are hierarchically ordered,
- and that processing is feedforward between these subsystems.

These assumptions are, incidentally, optimal specifications for a general purpose computer vision system. Such a system would be easy to exploit, because conveniently encapsulated, easy to parcel and develop, because intrinsically modular, and easy to implement in systolic dedicated hardware. In fact, one of the most influential study on the visual system, David Marr's *Vision* (1982) subscribes very much to this model, and presents the methods of conception and analysis necessary to organize a visual system. The orthodox model of pure vision has been very popular. Yet, all the underlying assumptions have now been knocked down. Studies using images generated on the fly show that disappearances or replacements in the peripheral visual field are unlikely to be detected. Firing latencies between different areas of the visual brain have considerable overlap. Presumably low-level processes such as stereopsis or segmentation can depend on higher processes of identification. All-in-all, there is now no other reason to hold on to the orthodox model than its convenience.

In lieu of the old orthodox model, Churchland, Ramachandran and Sejnowski (1994) propose a model of interactive vision where representations are more labile, and where interaction between brain subsystems plays a greater role. In their prospectus for "interactive vision", they suggest that:

- the cortex encodes only relevant information, rather than a complete representation of the scene,
- vision is exploratory and predictive,
- low-level visual properties can influence behavior, and in particular acquisition of more visual information,
- cortical feedback is instrumental in even low-level processing, such as figure-ground segmentation and stereopsis,

- memories influence which percept is seen.

Thus, in their view, information acquisition and maintenance is kept to a bare minimum. Vision *checks* the validity of ongoing assumptions, and it *grasps* information immediately necessary. The importance of feedback at the cortical level is related to the *checking* function of vision: inasmuch as possible, incoming visual information is incorporated in the current representation. The importance of visio-motor connections is related to the *grasping* function: sensors are controlled to maximize the acquisition of information looked for. Overall, the conception of the visual system proposed by Churchland, Ramachandran and Sejnowski (1994) is that of a more specialized and less encapsulated system. The notion of hard hierarchy must be replaced by a notion of "soft hierarchy" or, better, the notion of organic hierarchy must be replaced by that of **functional hierarchy.**

The theory of "Interactive Vision" once more delivers us from an introspective trap, and it in this that it is most relevant for mental imagery. Whereas the orthodox model suggests that the pinnacle of visual information processing is the conscious experience of a complete visual scene, the interactive model is much more subtle. It articulates two principles: first, that a lot of non-conscious visual processing still influences motor behavior, and second, that our conscious experience –or to be accurate, those representations of which we are conscious– biases and tricks even pre-conscious perceptual processes. Thus, like Ouroboros,[6] the visual system bites its own tail. Let us now follow Steve Kosslyn, who suggests that top-down influence in the visual system can be so strong as to create percepts out of thin air, "mental images".

II.B.3 Top-down effects in the visual system: Steve Kosslyn's account of imagery

Given the visual system as a soft hierarchy where top-down effects highly contrive low-level processes, imagery can be the result of these top-down down effects in the absence of input. This is Kosslyn's theory, expounded in his 1994 book, "Image and

[6]Orobouros, the great snake of pre-classic Greek mythology, was wrapped around the Earth, under the seas, and held the world together by biting its own tail. Because its body wrapped around the earth, the Serpent did not know it was its own tail that it was biting.

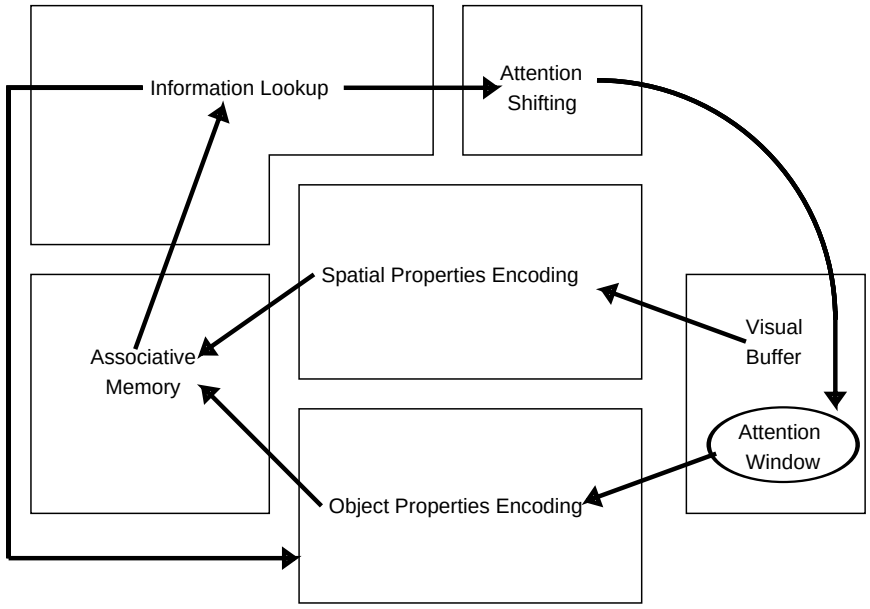

Figure II.3: **Outline of Kosslyn's analysis of the visual system:** Kosslyn decomposes the cortical visual system into six grand subsystems.

Brain, the resolution of the imagery debate". Kosslyn presents an integrated picture of the mechanisms of vision and defends the thesis that pure imagery is a by-product of these mechanisms (Kosslyn, 1994). Because psychological data alone underdetermines the problem of describing the mechanisms of vision, Kosslyn attempted to use information from cognitive psychology (Reaction Times and Error Rate studies, divided field studies) and from cognitive neuroscience (brain damage deficit assessment, EEG, ERP, MEG, SPECT, PET, FMRI), as well as from what neuroscience teaches us about the visual system of other primates (lesion studies, single cell recording). Kosslyn's detailed analysis of the visual system, which we review hereafter, allows a classification of top-down effects. Kosslyn himself uses the phrase "pattern completion", but he remains voluntarily vague as to which mechanism can implement it. For that reason, Kosslyn's view of imagery can be considered as a natural continuation of the theory of pure vision.

In his model of visual perception, Kosslyn identified seven subsystems (see Figure II.3). Although they are seen as black boxes, Kosslyn suggested that each of them is a neural net. The whole system is "nearly decomposable", which means that the subsystems have a weak modularity, and are functionally interdependent. Moreover, they can be localized anatomically, but can also overlap, and the same fine-grained networks can be members of more than one subsystem. Finally, these subsystems process information concurrently, as a cascade model (McClelland, 1979), and use cooperative computation, so that a set of subsystems can settle in a mutually consistent pattern of activity.

The gross organization of the model is shown in Figure II.3. The first subsystem is the visual buffer. It corresponds to the low-level visual areas (V1-V5), and represents information by encoding in a spatial retinotopic layout various features such as lines, line segments, line terminations, texture grain and orientation, and color. An attentional window can select a region from this buffer for further processing along two pathways: the ventral subsystem, which encodes object properties such as shape and color, and the dorsal subsystem, which specifies spatial properties, and location and movement information related to the guidance of action. The outputs of these two subsystems come together in an associative memory subsystem where they are matched to stored information. When the input does not retrieve a particular object, the stored information

which is partially activated guides an active look-up in associative memory, and plans spatial attention shifts and active attentional priming of the properties looked for.

In this model, imagery is activation of the visual buffer not caused by sensory input. The role of the feedback processes is to enhance the activity patterns corresponding to a perceived object, and to facilitate matches between stored images and partially seen objects; imagery is a consequence of these top-down mechanisms. Images tend to fade rapidly because the primary goal of the imagery apparatus is not duration but perceptual enhancement. Imagery is a useful adjunct of memory because it allows to "unpack" visual information. Operations on images are successive image reconstructions rather than transformations within the visual buffer. Once imagery images are formed in the visual buffer, they can be inspected much like sensory images.

Before it even considers object identification in difficult situations, such as identifying occluded or distorted objects, a theory of visual perception must account for recognizing an object in full sight. Recognition from different views, regardless of location and size, involves a trade-off between the number of stored representations and the amount of processing to do before matching the input to a stored representation. Both the size of the anatomical pathways dedicated to shape detection and neural network studies underline the difficulty of the preprocessing. It is therefore reasonable to think that totally different perspectives on the same object are stored separately. Kosslyn suggests that the figure-ground separation is made in the visual buffer by selecting the figure, or part of it, in the attentional window. Psychophysics studies demonstrating a scope-resolution trade-off, as well as computational models such as Marr's (1982), suggest that the visual buffer stores information at different spatial scales. Position and size invariance could be accounted for by changes in the location and size of the attentional window, or, for very eccentric positions, by image shifts in the visual buffer. The attentional window shifts are the same in imagery and perception, except that stimulus-prompted attention shift ("visual grasp reflex") is restricted to perception and perhaps REM dream.[7] That imagery uses the visual buffer and the attention window just like perception can account for similar effects in perception and imagery concerning the

[7] This restriction may account for the partial recovery of some lesion patients, who recover from perceptual neglect, but not from the corresponding imagery neglect.

scope-resolution trade-off and the maximum visual angle. The scope-resolution trade-off further suggests that associative memory stores information at different spatial scales.

Besides attentional rescaling, the visual system can compute an intermediary stage: that of sets of features invariant with changes in point of view. Two kinds of such features could be involved (Lowe, 1985) : non-accidental properties (NAPs) and signal properties (SPs). Non-accidental properties are invariant organizations of edges and lines, for example parallel lines in three dimensions stay parallel in any projection; Biederman's experiments (1987) on identification of partially erased line drawings show how some edges are more useful than others, thus illustrating the idea that these edges participate in computing NAPs. Signal properties are properties of surfaces, such as color and texture. Extraction of these properties would take place in the visual buffer at the interface with the ventral stream; V4 is the suggested candidate. These properties are used in a pattern activation subsystem where they are matched to stored shapes. The shapes encoded in the pattern activation subsystem tend to be viewer-centered; object-centered representations are not at that level, but rather consist of disambiguating visual routines (Ullman, 1984). The pattern association subsystem would be located mostly in the infero-temporal cortex; Kosslyn actually suggests a lateral specialization, the right subsystem focusing on shapes of exemplars and the left on prototypes. The pattern activation subsystem implements the viewpoint consistency constraint because the NAPs and SPs can only retrieve the view of an object where they are arranged properly, and are fully consistent with a unique viewpoint.

Concurrently with this bottom-up process, the most active representation in the pattern activation subsystem sends feedback to the visual system. If appropriate, this feedback augments the input image. If not, the corresponding representation is inhibited, and the mechanism repeats with the next most active representation. "In this sense," says Kosslyn, " imagery serves to augment the input to the visual system, allowing the representation to be more consistent with the representation downstream". This implies that the feedback connections must be more diffuse than the feedforward connections, and that feedback must skip some stages so that the system does not simply consist of a chain of subsystems with reciprocal connections, but, rather, has direct feedback over several stages. Anatomically, one can differentiate two sorts of feedback

connections: cortico-cortical – which will implement pattern completion–, and cortico-thalamic-cortical – which modulates the shape of the attentional window. In this view, perception needs cooperative computation, and imagery is the feedback part of this mechanism. This view explains the brevity of imagery images: the perception mechanism must be altogether data-driven, therefore the neural substrate must not sustain feedback information too long. Also, the properties of the pattern association system justify the fact that, like perception, imagery is viewer-centered. Finally, the dual nature of the pattern activation code allows us to image categorical objects as well as particular exemplars: for example, we can image "a dog" very much unspecified, or specifically "Auntie's dog" (with all due respect to Jerry Fodor, this one truly looks like a Pittbull to me).

In occluded or degraded situations, other attributes such as motion and the relationship of an object to its parts become important. Motion is encoded by a separate subsystem – the motion information subsystem –, before joining the main data stream in the pattern association subsystem. The motion information subsystem could be implemented in MT (V5); thus MT plays for motion the same role as V4 does for shape: that of the ultimate preprocessing before pattern association. The fact that motion is processed like other attributes by the pattern association mechanism explains the non-trivial fact that we can image pictures "tainted" with motion. For example, if we visualize Carl Lewis running, the image can include motion percepts even without any movie-like temporal extension.

Kosslyn argues that there is no "alphabet of shapes", like the Geons suggested by Biederman (1987). Rather, parts and objects made of these parts are encoded at the same level. Identifying an object by some of its parts relies on spatial encoding. In the dorsal stream (7a and, to a lesser extent, 7b, LIP, and MST), a distinct spatiotopic system computes the location of objects in body-centered or other-object-centered coordinates. It gives a map of the location, size and orientation of objects. This map is then processed to encode spatial relations. Once again, there is evidence of hemispheric lateralization, with the right hemisphere encoding continuous spatial coordinate relationships and the left encoding categorical spatial relationships – that is, relations of discrete nature such as "is at the right of" or "on top of". These relations are used to identify exemplars

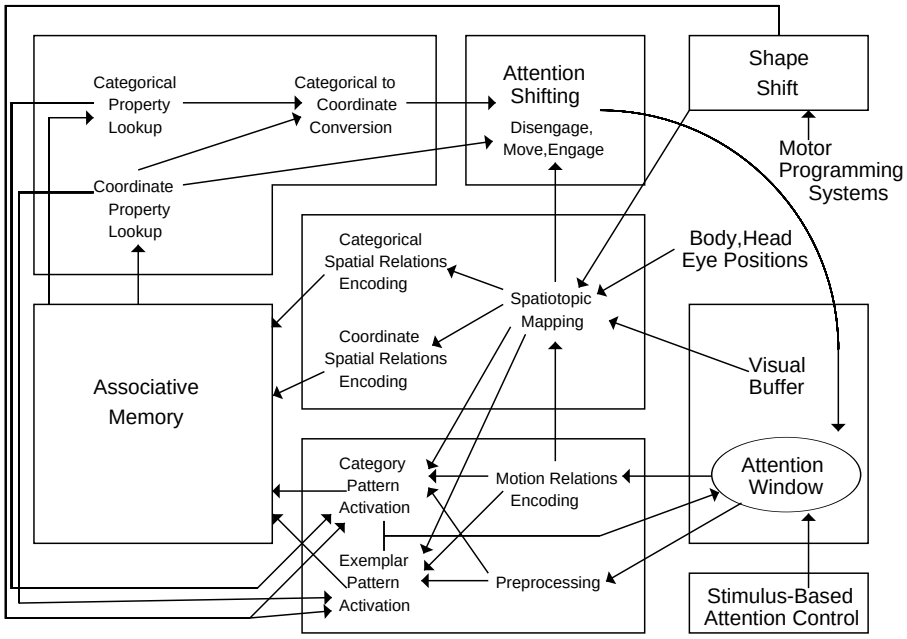

Figure II.4: **Details of Kosslyn's functional analysis of the visual system:** The subsystems previously introduced are now detailed. Note the dichotomy between categorical and particular properties, presumably encoded respectively in the left and right hemispheres. (From Kosslyn 1994).

or prototypes of shapes made of identified parts. They allow propositional description of object composition. Moreover, the flexibility of the categorical encoding makes it possible to specify deformable objects, such as animal or human bodies, or machines. This system also allows imagery to create objects from different parts connected by categorical spatial relations: for example, a lion's head, a goat's body, and a snake's tail, to make the Chimera.

When the input is unsufficient to enable first-shot identification, the system engages in active look-up. This necessitates a conversion into specific coordinate spatial relations of the categorical relations specified in the propositional description of an object. These coordinate spatial relations are in turn converted into retinal coordinates, to provide specific targets for attentional shifts. Not only does the lookup mechanism control spatial search; it also primes the pattern activation subsystems to facilitate encoding of the expected part or characteristic. With the coming into play of this look-up mechanism, perception and imagery are even more blended than in the straightforward case because perception relies more and more on top-down processes.

With this system in mind (detailed chart in Figure II.4), Kosslyn now addresses imagery phenomena separate from perception, and firstly how images are generated and sustained. Single part images are generated when associative memory representations prime the pattern activation systems and the visual buffer. In multi-part images, a global image is generated first, then details are added according to their spatial relation to the whole. These images can be made from prototypes or from particular exemplars. The whole issue of imagery lateralization seems to be connected to this dichotomy: the right hemisphere tends to participate in generating images of particulars, and to compute spatial coordinate relations; the left hemisphere processes primarily categories, generates images of prototypes, and uses categorical spatial relationships. A strong claim of this model, which seems to be experimentally verified, is that the higher the number of components, the longer it takes to image a complex shape, because components are added one after the other, rather than all at once, even when they are details of the same size. There is a distinction between images induced directly from visual memory and images induced by attention, for example, by enhancing a choice of segments in a perceived lattice. Ventral imagery can recreate a single part image in parallel, but

dorsal imagery cannot. Finally, it is possible to do "attentive mental tracing" by moving the attentional window across the visual buffer and keeping an activated trace of its trajectory. Image maintenance is mostly image re-generation, but there is no need to integrate all the parts again because the top-level representations are driving the image generation. Imagery seems to be limited to as many objects as we can maintain chunks in short term memory. Kosslyn argues that chunking is primarily a visual phenomenon. He also posits that imagery is an important part of working memory, by allowing us to implement "mental models" (Johnson Laird, 1983).

Next, Kosslyn addresses image inspection. Imagery images are inspected just like perception. All mechanisms of vision, with the exception of stimulus-driven attentional shifts, proceed from the visual buffer and, therefore, perform the same processing on sensory images and imagery images. Two patients with intact imagery but impaired object recognition have been reported recently by researchers attacking this view: C.K., by Behrman, Winocur and Moscovitch (1992), and M.D., by Jankowiak, Kinsbourne, Shalev and Bachman (1992). Kosslyn suggests that lack of processing resource could impair the visual mechanisms so that top-down processes would not be affected even though bottom-up processes would. The congenitally blind are an exception to this rule; visual angle studies demonstrate that imagery is not visual for them. It seems that in many tasks, image generation followed by inspection is a slow process but one which allows parallel, "pop-out" search. In the opposite, propositional computation is a serial process which starts sooner. The two processes appear to run easily in parallel. This could account for complex results in tasks where either imagery or propositional computation can be faster.

Finally, Kosslyn examines image transformation mechanisms. First, images make accessible the geometric properties of objects: spatial decompositions can especially be worked on by addition or subtraction of parts. Next, images can be manipulated by motor processes that the dorsal system ordinarily uses to compensate for expected movement. Even without considering images of body parts, for which there is strong evidence of kinesthetic imagery during "visual" rotation, image transformation seems to be implemented by computations in the dorsal subsystem aimed at anticipating the result of a movement. Thus, zooming towards an image uses the same high-level motor

processes that allow our perception of an object to stay steady when we walk towards it. These processes magnify image size so that the image adapts to the closer point of view. Likewise, mental rotation uses the same processes which give us good continuous perception of an object we manipulate. Such transformation yields a sequence of images, which are very different from a single image where motion is encoded. Rotation is independent from attention, but in some tasks the subjects have a choice between image inspection (by attentional shift) or image transformation (by motor processes). Finally, other subsystems enable other kinds of transformations, such as adding or changing color.

In summary, Kosslyn's view of vision is a complex one, but it accounts for realistic situations of object and scene recognition. Kosslyn's point is that these realistic situations often provide very scarce information on some objects, and that, given a store of possible objects, the visual system has pattern completion and lookup mechanisms to make the best of the available information. These mechanisms are themselves complex, and range from static image enhancement to reconstruction of occluded parts in deformable moving objects. Functioning of these mechanism without perceptual input results in imagery, in the same spatial medium (the visual buffer) which is used for perception.

II.C Attractor classification, Bistability and Imageries

Now that we have set the stage – the visual system, it is time to introduce our main actor: attractor dynamics. In our theory, attractor dynamics has the first two roles! First, it makes brain regions bistable, allowing a near all-or-none regime. Second, it makes the activity pattern evolve towards a particular type of shape: one of the local attractors. Thus the attractor dynamics operates a classification. We will see that these two roles are critical for mental imagery. Because areas are bistable, each of them can be used or not for mental imagery – this account for the variety of imageries. Because the dynamics makes the pattern evolve, images roughly formed by top-down effects can be perfected by a local mechanism. Thus does attractor dynamics kill two birds with one stone.

In this section we will move from the principles to their application. First we will

present typical attractor networks, the Hopfield net, the Brain-State-in-a-Box net, and the Cowan and Wilson models. Next, we will review some of Walter Freeman's studies of attractor dynamics in real cortex. Finally, we will incorporate attractor dynamics into the classical view of the visual system –seen in section 1– and present in particular the consequences for mental imagery.

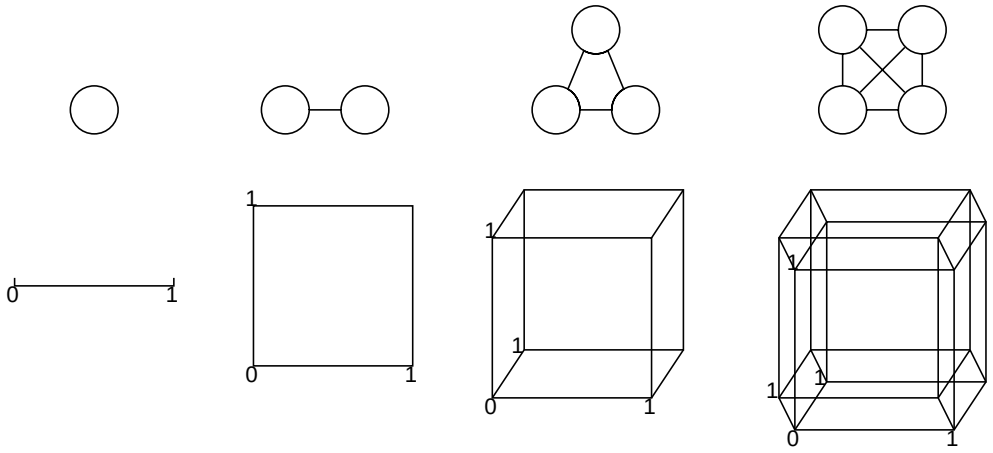

Figure II.5: **Phase space of attractor networks:** This Figure illustrates the phase spaces for networks of 1, 2, 3 and 4 neurons, shown in the top row. In a Hopfield net, each neuron can individually take value 0 or 1, in a Brain-State-in-a-Box, any value between 0 and 1. Hence, when N neurons can take these independent values, the phase state is a hypercube of dimension N. Precisely, for a BSB, the phase space is the hypercube itself, but for a Hopfield net, it is restricted to the corners of the hypercube. (To simplify the presentation of the algebra, [0,1] can be replaced by [-1,+1] without any loss of generality, and we do so in the text).

II.C.1 Attractor networks

Hopfield networks and Bain-State-in-the-Box networks

The most famous, although not earliest, dynamical system model is that of John Hopfield (1982). The Hopfield model is a discrete time-scale network of totally connected neurons updated one after the other with a simple update rule. These neurons (Pitts &

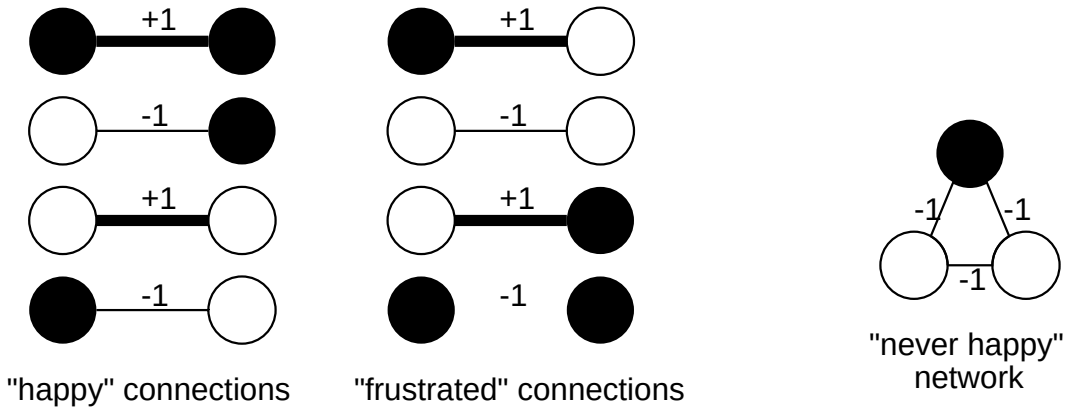

Figure II.6: **"Happy" v.s. "frustrated" connections:** Because the update rule is that neurons take the sign of their cumulated input, connections tend to push the network state to a state where the relationship between connected neurons is harmonious. Therefore, connections of positive sign will be "happy" if they relate neurons of the same sign. Positive connections between neurons of opposite sign are "frustrated". Conversely, connections of negative sign will be "happy" if the connected neurons are of opposite signs. The network evolves to maximize happiness (or minimize frustration!). Sometimes, as shown in the 3-neuron network on the right, the weights cannot be all satisfied. In this 3-neuron particular case the worst states are those with all neurons of the same sign; any state distributed 2-1 is better, all of which are equivalent, but none optimal.

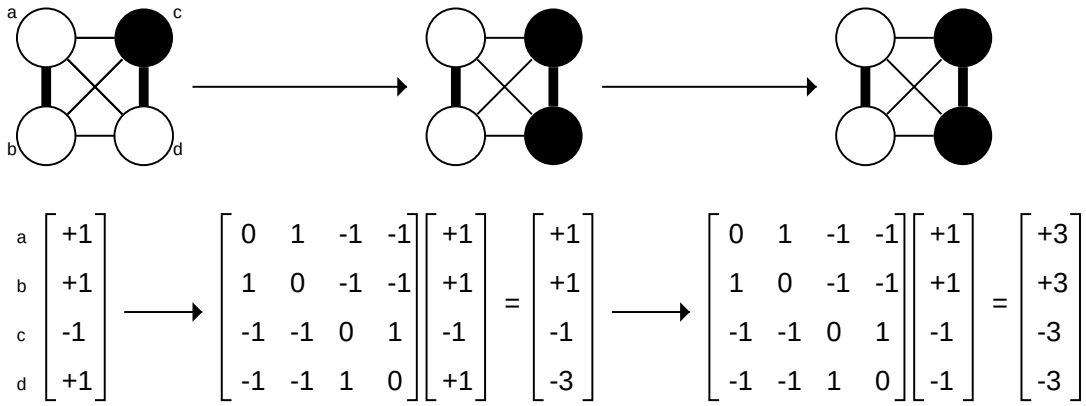

Figure II.7: **Dynamics of Hopfield nets:** Neurons are updated to take the sign of their cumulated input. This figure illustrates the normal evolution of the network (here with batch updating): the state of individual neurons becomes what the rest of the network tell them to be. Here, one neuron flip is enough to make all connections "happy". In large networks, several update cycles may be necessary, and a fair minority of the connections usually remain "frustrated". It is also customary to update networks one neuron at a time, rather than all at the same time, because the latter procedure can create limit cycles of period greater than two (usually *around* rather than *at* the local minimum of the frustration function). Filled neurons have negative activations, hollow neurons have positive activation, wide lines indicate connections of weight +1, narrow lines connections of weight -1.

McCulloch, 1947) take state values 0 or +1 depending on the sign of their input, which is a weighted sum of the states of all other neurons at the previous time step. With little loss of generality[8], we can consider the simpler case where the activation values are +1 and -1. With this simplification, the update rule for the state X_i of neuron i is:

$$net_i = \Sigma_j w_{ij}.X_j \tag{II.1}$$

$$X_i = sign(net_i) \tag{II.2}$$

The foremost characteristic of orthodox Hopfield nets is that the weights are symmetrical.[9] [10] Symmetrical weights tend to create mostly fixed point attractors, generally excluding cycles or chaotic attractors[11] . This simple dynamics allows classification and content-indexed retrieval, both of which are very desirable computational properties.

The key concept in Hopfield nets is that of frustration: each connection "pushes" the state of the net towards its own preference. The net, as a whole, changes its state in function of the aggregate of these little pushes. This is illustrated by Figures II.7 and II.6. If one considers a net of only two neurons connected by a positive weight, these neurons have positive feedback, and they tend to take the same state. Indeed, if neuron 1 is excited, input to neuron 2 will be positive, and neuron 2 will flip to the excited state, so that the connection will be "happy". If, due to the contribution of input or other neurons in the net, these neurons have to take opposite states, that connection will be "frustrated" and will increase the total frustration of the network, by a penalty equal to the connection weight. As seen in the case of two neurons, the network left to itself always evolves towards a state of minimal frustration (this property generalizes to n neurons easily if one takes care of having asynchronous activation, to avoid limit cycles

[8] By using 0/1 states, Hopfield shifts the energy function, which then takes its minima near the 0000...0 corner of the hypercube. This disrupts the symmetry between opposite corners, and makes 0000...0 a default attractor. With -1/+1 values, the energy function is centered on 0000...0 (now the hypercube center rather than a particular corner) and the dynamics is symmetrical.

[9] This makes their analysis isomorphic to that of spin glasses in physics, a state of matter where sparse strong-spin crystals interact symmetrically. By simply remarking this, Hopfield converted for a generation of physicists to neuro-computation.

[10] Hopfield also reported simulations with non-symmetrical weights, but the theory about these is not as complete.

[11] In a nutshell, because the eigenvectors of the weight matrix are all real. Asymmetric weight matrices have complex eigenvectors and can, for example, make a patch of activity rotate around a network

"orbiting" around these minima). So, the total frustration acts for this discrete step network like an energy or Lyapunov function that the system tries to minimize, and here physicists feel at home. [12]

Network modelers, though, want something more: learning. It happens that simple Hebbian learning (Hebb, 1949) has remarkable properties in a Hopfield net. Hebbian learning consists of presenting a target state and adjusting the connections depending on that state; the update rule is:

$$W_{ij} += K X_i X_j \qquad \text{(II.3)}$$

where K is a positive constant[13] reflecting the importance given to that particular target state. In an initially blank net, hebbian learning of any state makes it a fixed point[14]. Additional states can be added provided they are orthogonal to the all states already stored.[15] One sees that after learning the first target, other states which are exactly orthogonal to (X_i) are equally far from $(-X_i)$, and are on the "separatrix" between the basin of (X_i) and that of $(-X_i)$.[16] The theoretical maximum of distinct attractors is

[12] To be accurate, this function is $-\Sigma_{i\neq j} X_i W_{ij} X_j$. The derivative with respect to neuron i is $\frac{d(-\Sigma_{i\neq j} X_i W_{ij} X_j)}{dX_i} = -\Sigma_j w_{ij}.X_j$ so that updating neuron i by changing its sign to that of $\Sigma_j w_{ij}.X_j$ will take a step in the direction of a decrease of frustration. It is a natural Newton method. A geometer's eye will notice that the frustration function is a quadratic form, so that the equipotentials are quadrics; moreover, using +1/-1 activity values and no biases, these quadrics are centered on (0,0,...,0). This lets us see that the stable points will be those corners of the hypercube which are actually "cornered" by equipotentials of the frustration.

[13] Negative constants can be used for "unlearning" a state.

[14] Let us consider a blank net and let us teach it only one pattern: Hebbian learning states that the weight will become $+K$ if both neurons have the same state in the pattern, and $-K$ otherwise. For the sake of simplicity, let us note X_i the state of neuron i, taking values $+1$ for the active state and -1 for the passive one, the family (X_i) as the state of the whole net. Learning it yields the weight matrix $(W_{ij}) = K.(X_i X_j)$ and $W_{ii} = 0$. Now, let us consider the frustration associated with the pattern just learned:
$F = -\Sigma_{i\neq j}(W_{ij} X_i X_j) = -\Sigma(K.X_i Y_j X_i Y_j) = -K.(N^2 - N)$
because $X_i X_i = +1$ always, and so does $X_j X_j$. In summary, the frustration will then be the absolute minimal because, by construction, all the connections are "happy". In dynamical systems lingo, we have thus created a fixed point attractor whose basin encompasses the whole space –or almost: it is a double fixed point, because $(-X_i)$ is as much an attractor as $(+X_i)$, as can be seen by the fact that Hebbian learning of $(-X_i)$ or $(+X_i)$ yields exactly the same weight matrix (the minus sign cancels out in $-X_i.-X_j$). Thus, each of the states (X_i) and $(-X_i)$ are now fixed points attracting half the state space.

[15] Non-orthogonal states can also be added, but the more aligned they are with already stored states, the more likely it is that an interference will merge the two "memories" into a single chimera.

[16] States which are almost orthogonal to (X_i) are close to the separatrix and withstand very weak

N; in practice if the states are not perfectly orthogonal, the performance degrades in proportion to the correlation (or cosine) of the states [17]. As long as not too many states are so stored, the net will converge towards one of them from any likely initial condition. If too many states are stored some of them will be fused together to form intermediate attractors, named "chimeras".

The Brain-State-in-a-Box network (Anderson, Silverstein, Ritz & Jones, 1977) is analogous to the Hopfield net, except that the neurons are allowed to take intermediate values rather than just the two extremes. Their update rule is therefore:

$$net_i = \Sigma_j w_{ij}.X_j(t) \tag{II.4}$$

$$X_i(t+1) = X_i(t) + Sigmoid(net_i) \tag{II.5}$$

Where the function $Sigmoid(x) = \frac{1}{1+e^{-x}}$ can be replaced by any other non-decreasing bounded function. The dynamics for the BSB is very analogous to that of the Hopfield net. If there is any non-zero attractor, they are all located at the corners of the phase space, so that they are in practice identical to the attractors of a Hopfield net set with the same weights. Because the dynamics allows continuous values, a BSB can work with initial states having rational coordinates. Hopfield (1984) proved formally that both types of networks have the same computational properties.

A few items must be remembered from the Hopfield nets:

- They implement mostly simple fixed points, not fancy dynamics like limit cycles or, fancier, like chaos. The key condition for that behavior is the symmetry of the connection weights.

- This fixed point behavior allows spatial pattern completion: an initial state will decay into one of the learned states.

attraction. In fact, in the net where only (X_i) has been learned, the frustration associated with another pattern (Y_i) can be expressed very simply in function of the angle T between X and Y: $Frustration = -K.N^2[cosine(T)]^2$. Thus, the frustration for the one target net is a quadratic function of $cos(T)$, so that it will stay much closer to zero than $cos(T)$. Thus, the influence of a learned target state onto an "almost-orthogonal" state is very small (a second order of the deviation from perfect orthogonality).

[17] Hopfield and later analyzes report between 0.10 N or 0.20 N attractors.

- The neurons may be replaced by more complicated automata, as long as their response is monotonic of the summed input and that the network is totally connected with symmetrical weights, one will have such fixed point behavior.

- Simple Hebbian learning in a totally connected net always creates such symmetrical weights, and hence a network which has this fixed point dynamics.

- If one uses a"true Hopfield" net, that is, with 0/1 activation values, or a BSB with a non-linear transfer function, the point 0000...0 is always stable; the net can then be considered as being bistable, taking either the 0000...0 default value, or moving into any other non-zero fixed-point – then a meaningful class.

Thus, we see that already the simplest variety of attractor networks already present the two properties of bistability and pattern completion.

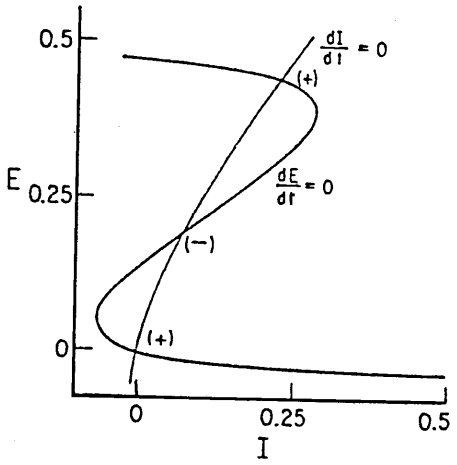

Figure II.8: **Stability analysis of Wilson and Cowan's first model:** The two curves indicate the (E,I) values for which excitation E or inhibition I has a null first order rate of change. Fixed points of the system are expected where both curves intersect because at these points both E and I won't change. There are three such points, of which two are stable (marked '+') and one is unstable (marked '-'). (From Wilson and Cowan, 1972).

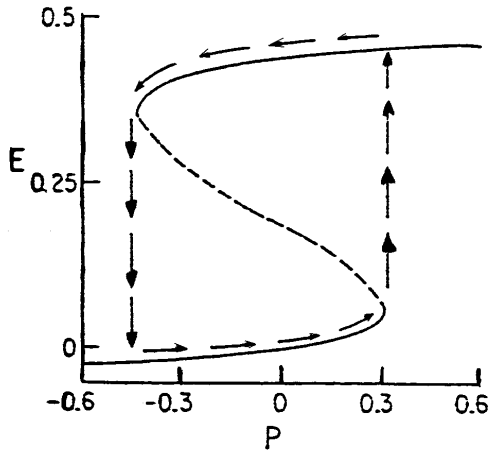

Figure II.9: **Hysteresis in Wilson and Cowan's first model:** This plot shows values of the input P and the excitation E which are solution of the system's equations. For values of P between -0.4 and +0.3 two stable states are possible (plus one unstable state in dotted line). The actual state of the system is one or the other of these two, depending on the history of the system. (From Wilson and Cowan, 1972).

The Wilson and Cowan model of neural sheets

In a different approach, Jack Cowan and Hugh Wilson (1972; 1973) propose a model of a simple cortical sheet made of two populations: one of excitatory neurons, and the other of inhibitory neurons. The most interesting feature of this model is that the weights are antisymmetrical rather than symmetrical. This allows more complex behavior than simple fixed points. Let us look at both versions of this model.

In a first version (Wilson & Cowan, 1972), they detail the dynamical behavior of aggregates without spatial organization. Each population receives input from itself, from the other, and from external connections. These inputs are integrated over time with exponential decay, so that all past states influence the next state. The population activity is a sigmoid[18] of its input, scaled by the proportion of neurons which are not in their refractory period [19]. The resulting equations are integral equations governing the excitatory (E) and inhibitory (I) populations. These equations can be solved analytically with one approximation.[20] The behavior of the system under constant or slowly varying input can be understood by drawing the phase plane (Figure II.8): the nullclines represent the loci where $\frac{dI}{dt} = 0$ and $\frac{dE}{dt} = 0$; their intersections are fixed points. Here, there are two attractors ('+') and one repeller ('−'), as is the case when feedback of the excitatory population dominates the dynamics. Other parameters, which make the loop I$\longrightarrow$E$\longrightarrow$I predominant, will yield five fixed points including three attractors. These multi-attractor settings cause a hysteresis behavior as the input changes. Figure II.9 summarizes the hysteresis behavior: as the input increases from very low values, the system stays as long as possible on the low-activation steady state. When this state disappears, the system's activation jumps to the higher steady state. If the input is again decreased, the system will now stay in the higher activation state, again as long as possible. In that manner, the state of the system does not only depend on its input,

[18] The sigmoid is justified as resulting from having in the population a Gaussian distribution of hard thresholds.

[19] Neurons which have not fired in the past 5ms

[20] The states E and I of the excitatory and inhibitory populations are replaced by their time-average over a short time scale (the membrane time constant). This technique, known as temporal coarse-grained, is justified here by numerical computations on the original equations which show that its only effect is to suppress high frequency oscillations which would dampen in less time than the membrane time constant.

but also on its past history. Some parameter settings amongst the hysteresis-causing parameter settings –especially those with very strong excitatory and very weak inhibitory feedback– can also cause limit cycles. The oscillation starts above some input threshold, and its frequency increases with the stimulation; but the amplitude of oscillation decreases at the same time, so that above another threshold the system is in a fixed point again. Thus, this system allows stimulus intensity to be coded by output frequency. Such coding has been found in monkey thalamic somato-sensory neurons, which encode wrist flexion angle by a frequency monotonically increasing with the angle. Finally, it must be pointed out that neural aggregates composed of more than two types of neurons may have a bipolar distribution of thresholds; this would result in a transfer function more complex than the sigmoid, and therefore in many more possible steady states, yielding more complex hysteresis behaviors. In summary, this zero-dimensional model exhibits complex hysteresis and limit cycles behaviors.

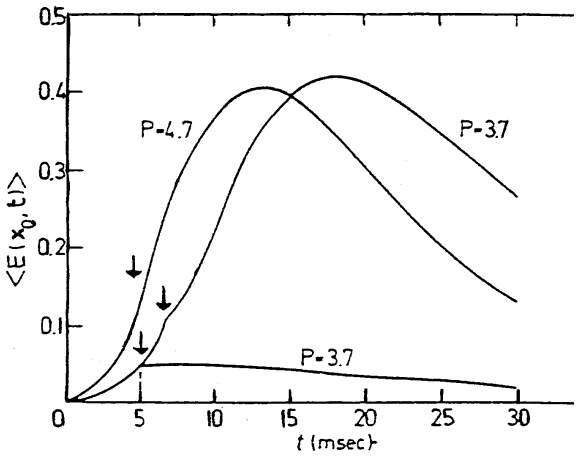

Figure II.10: **The timecourse of an active transient:** Stimulation lasts from time=0 to the arrow. As can be seen, strong stimulation (P=4.7) can be stopped early and still trigger a transient with early latency. Weak stimulation (P=3.7) will not reach threshold and will decay slowly, unless it is kept on a little longer. (From Wilson and Cowan, 1973).

In a second version (1973), Wilson and Cowan give the previous model spatial extent, by studying a one-dimensional line of neural tissue made of excitatory and in-

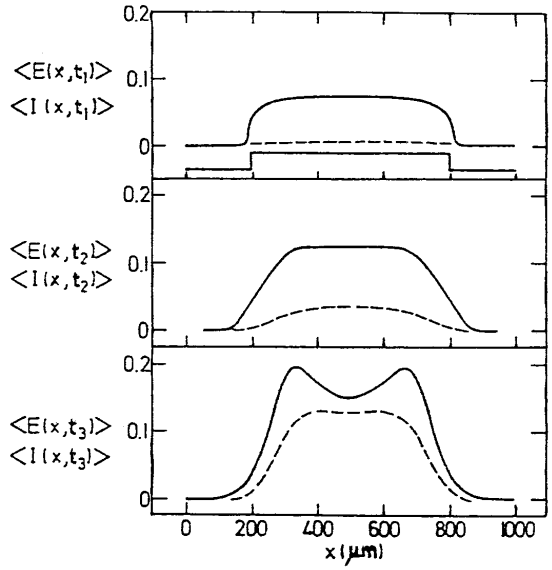

Figure II.11: **The evolution of edge enhancement:** The three plots show the spatial distribution of Excitation (solid lines) and Inhibition (Dotted lines) at three successive times (from top to bottom). One sees how an initial square stimulation evolves first into a blurred lump of activity and then into a lump with stronger activation at its edges. (From Wilson and Cowan, 1973).

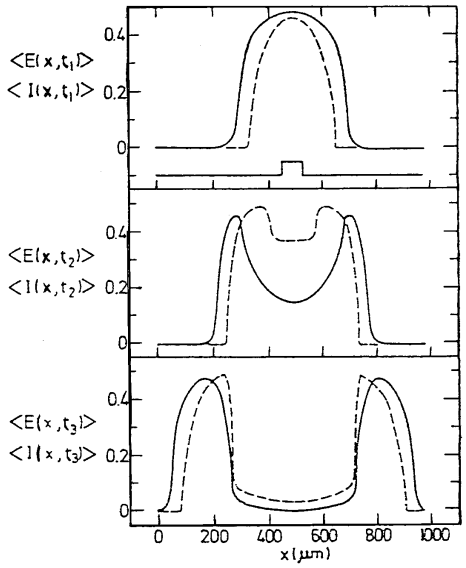

Figure II.12: **Propagation of activity waves:** In the proper situation, edge enhancement will be so strong as to generate two decoupled waves of activation traveling in opposite direction from the spot of initial stimulation. (From Wilson and Cowan, 1973).

hibitory populations. The model is adapted by having spatial connectivity patterns and input pattern, instead of scalar connection strengths and input. The connection patterns used are exponentially decreasing with distance, but the qualitative results only depend on their being monotonically decreasing, so that the conclusions can be extrapolated to more plausible connectivity. Again, the equations are transformed using coarse time-graining. Analysis shows three types of behavior associated with different connectivity settings: active transients, spatially localized limit cycles, and spatially localized steady states. Active transients are short term reactions to quasi-instantaneous stimulation. They develop when stimulation is stronger than some threshold and wider than some minimal width (about twice the connectivity decay constant), peak some 10 to 15ms after stimulation ends, and decay afterwards. Figure II.10 shows the time course of an active transient. For wide enough stimuli, the transient develops an edge enhancement in its late phase (Figure II.11). More intense stimulations result in shorter peak latencies and in more external edge enhancement (making the decaying trace wider). Spatially localized limit cycles occur when the inhibitory feedback is much weaker, making the

E$\longrightarrow$I$\longrightarrow$E loop dominant of the dynamics. Again, they only occur for stimuli of some intensity and width, and the oscillation lasts only as long as stimulation is on. They demonstrate edge enhancement for sufficiently large stimuli. As in the zero-dimensional model, the frequency increases with stimulus intensity: it goes from 14 Hz to 23 Hz for a four-fold increase in stimulus intensity. The same parameter settings allow the neural tissue to respond to periodic stimulation with a divided frequency; ratios of 1/2, 1/3, 1/4, etc..., can be achieved depending on the frequency and intensity of the stimulation. With diffuse inhibition of the inhibitory population, the same tissue can also be the locus of propagating waves (Figure II.12). Spatially inhomogeneous steady states occur for a third kind of settings. Contrarily to the previous two, these states remain activated in the absence of sustained stimulation. Narrow initial stimulation causes one residual peak, whereas broad initial stimulation causes two peaks, one at each edge of the initial wave. The authors suggest that the three different types of behavior that the model can demonstrate could correspond, in the brain, to sensorial tissue for the active transient mode, thalamic tissue for the oscillatory mode, and frontal tissue –implementing short term memory– for the steady state mode.

To further support their suggestion about the transient mode, Wilson and Cowan replicate three psychophysical results: the sensorial transfer function, metacontrast, and stimulus fusion hysteresis. The visual apparatus is known to perceive contrast more easily at small spatial frequencies than at very small or very large ones. Using a brief sinusoidal stimulation, and measuring the maximum peak to through amplitude of the transient, also gives a transfer function attenuating very small and large frequencies. Metacontrast is a masking phenomenon with the mask presented from 10 to 20 ms after the stimulus, flanking but not overlapping the stimulus location. The model shows that the decaying period of the transient is strongly dampened by the masking, whereas its peak amplitude is relatively unimpaired. Fusion hysteresis has been demonstrated by Fender and Julesz (1967) in stereopsis. When two stimuli are fused, the visual system tolerates more disparity before it separates them than it admits to fuse them, if they are initially separated. In the Wilson and Cowan model, two sharply peaked Gaussians were used as stimuli. Their means were originally confounded, causing a single peak of activation in the tissue. Afterwards, they were slowly moved away from each other, up

to some critical disparity where the central peak quickly faded to be replaced by two separate peaks. The critical disparity to separate the inputs was much larger than the critical disparity to fuse them (obtained by driving initially far-away stimuli towards each other). In conclusion, these three psychophysical effects can be explained by very local properties of dual-type neural sheets. Ether parameters of the same model make it possible to replicate oscillatory behavior, such as is found in the thalamus, and steady state properties which the authors hypothesize could implement short term memory.

It is interesting to note that, out of the twelve parameters in the model, the minute change of only one –the weight of the excitatory feedback connections (E–E), from 1.5 to 2.0– leads to the change of the active transient behavior into the steady-state mode which manifests hysteresis. This suggests that the steady-state behavior is the extension of the active transient behavior. Indeed, a bifurcation analysis of the 0-dimensional model (Borisyuk & Kirillov, 1992) shows that the critical input thresholds (values for which activation jumps due to hysteresis phenomena) change with another parameter. In the 1-D model with the active transient parameter setting, one can see easily that sustained input would supplement positive feedback and have over the late course of activation the same effect as stronger positive feedback without sustained input. Thus, even though Wilson and Cowan do not present this result, we can see that the steady state is actually a mode which occurs as soon as positive input is sustained over an amount depending on the weight of excitatory feedback. The steady settings correspond to values for which the necessary sustained input is less than zero. Finally, one can predict that even stronger values of excitatory feedback will only make the steady state harder to reset by inhibitory input.

After this review of the Wilson and Cowan models, we want to consider again the basic two properties that we were looking for: bistability and pattern completion. Both of models studied by Wilson and Cowan show bistable behavior associated with a threshold of stimulation. Interestingly enough, the threshold depends both on intensity and duration of the stimulation –this is also the case for visual perceptive thresholds. Pattern completion is reduced to its simplest expression and expressed only in the second model –because the first model has no spatial extent. Two narrow and close peaks of stimulation initiate a single lump of activation. Considering that even the second model

has only one spatial dimension and that its connections are strictly repetitive, implementing no particular substructure, this fusion of activity peaks is the best that could be expected and represents truly pattern completion in this system. The many other interesting properties and regimes shown by these models open the way for interesting theories, some of which have been pushed forward by Jack Cowan, in particular a theory of low-level visual hallucinations. To conclude, besides interesting properties peculiar to these models, we find in them bistability and pattern completion, just as we did in simpler models.

Previous work on cortical dynamical systems

Where the two previous sections have introduced theoretical or modeled neural nets, there remains to show that attractor behavior can be a feature of real cortex. For this, we turn to Walter Freeman. Freeman's work presents a dynamical system interpretation of biological data. Freeman works mostly on the olfactory system, with field potential measures, recording the EEG at the surface of the olfactory bulb or of the cortex with arrays of closely packed electrodes. This technique provides a spatial resolution of the order of 0.1 to 1 mm and a 1 millisecond temporal resolution. In addition, he derives and analyzes theoretical models (Freeman, 1975; Yao & Freeman, 1990) based on the classical dynamical systems theory, which describes the system dynamics by differential equations.

Freeman on the olfactory system There are several interesting aspects to the olfactory bulb. First, it has a more simple structure than other sensorial relays. There are two main layers: one is made of the mitral cells, which are excitatory, have lateral connections, and send their output to a thalamic relay and limbic structures; the other is made of granule cells, which make a densely connected layer of inhibitory interneurons. Input from the olfactory nerves stimulates the mitral cells in small plexii of synapses: the glomeruli, where many synapses are densely packed. This two-layer architecture resembles that of the Wilson and Cowan model. Second, the olfactory bulb has a burst dynamics, rather than a continuous dynamics. The olfactory bulb is sensitized by each inspiration, and presents a burst of electrical activity with 2 to 5 peaks. Finally, the

activity is spatially inhomogeneous.

Freeman's early work (Freeman, 1978) showed that, contrary to expectations, the overall spatial pattern did not change when sniffing new odors. Only a few foci on the bulb –usually lateral or ventromedial–, depending on the animal but not on the odor, were able to generate high intensity bursts. These higher-intensity loci had highly synchronized activity during bursts. The burst itself is presented as a localized limit cycle, resulting from feedback with strong gain (Freeman, 1975). Further work (Freeman, 1983) using classical conditioning, demonstrated that this locus of high activity was amenable to change by learning. This, along with the fact that the appearance of the pattern was strongly dependent upon the motivational state of the rabbit, led Freeman to postulate that the pattern represented odor expectancy rather than perception. He hypothesized that the connections between mitral cells (all excitatory) can be strengthened by learning in a Hebbian manner, and that some subsets of the mitral population became more strongly interconnected, thus providing the substrate for particular attractors in the dynamics.

These hypotheses were explored further in a series of three papers (Freeman & Viana Di Prisco, 1986; Freeman & Baird, 1987; Freeman & Grajski, 1987). The same data, recorded on an 8x8 electrode array from the olfactory bulb of 5 rabbits, was used in three different analyses: temporal, spatial, and factorial. The rabbits were conditioned to lick or sniff in reaction to different odors, so that the EEG segments could be correlated with behavioral performance. The temporal analysis revealed that there were two kind of bursts: first, disorderly bursts with a main frequency below 55 Hz, a wide spread of oscillation energy throughout the frequency spectrum, and no correlation of the spatial patterns between bursts obtained in similar conditions; second, orderly bursts of main frequency over 55 Hz, with a much more concentrated spectrum, and for which the spatial patterns of amplitude were more reproducible across bursts of the same class. This temporal analysis suggests that most previous work on the olfactory system was scrambled by including the "pure noise" low frequency bursts. In consequence, the low-frequency bursts were removed from the data for the other two analyses. The spatial analysis also included extra data recorded on a 1-D array with smaller (0.2 mm versus 0.57 mm) center-to-center distance. To test for possible use of spatial high frequencies

in encoding information, the data was deconvoluted to restore the activation pattern at the depth of the source dipoles (the granule cells). Both the raw and the deconvoluted data pointed to a low-frequency encoding of odor information (the passband containing information is from 0.15 to 0.25 cycles per millimeter). In addition, all the channels were relevant for information encoding, which suggests that the whole bulb, and not only the loci of high activity, is relevant for information encoding. The third analysis of this data was a factor analysis of the amplitude patterns. For all but one rabbit, loadings on the 4 to 7 main factors made it possible to classify correctly over 65 % of the bursts (chance level being 33 %). When the channels were normalized to remove the effect of the individual "signature" (the foci of high intensity), the invariance of patterns within conditions was visible. These results imply that stable spatial patterns in response to the conditioning odor emerged under reinforcement.

From the three studies, Freeman asserts that "the main conclusion supported by these data is that the bulb maintains a collection of preferred states of activity; each state is identified with a reproducible spatial pattern of EEG oscillation and with an odor condition. The bifurcation or switching of the bulb from one preferred state to another requires a change within the neural system, not merely in the input." As for the causes of such changes, Freeman suggests increases in basal activity, corticofugal input, and reinforcement of mitro-mitral connections by learning. He views the main spatial pattern characteristic of the individual as non-significant, and suggests that it is the olfactory version of a fovea. He proposes that subsets of mitral cells with reinforced connections implement an attractor in the bulbar dynamics, and that this stronger feedback makes the bulb extremely sensitive to some input patterns. Thereby, when the corresponding odor is recognized, activity in that subset governs the whole bulbar dynamics. Even though the attractor is created by the reinforcement of the excitatory connections between a subset of the mitral cells, the entire bulb participates in the spatial pattern. He argues that the temporal shape of the bursts is also irrelevant, and that activity becomes synchronized by coupling of oscillators, each of which is the group of mitral cells and granule cells synapsing into the same glomerulus.

Comments on Freeman's dynamical model of the olfactory system Before we study Freeman's brief incursion into the visual system, his views on the olfactive system call for a few comments. First, his research elicits a remarkable picture of dynamics in the olfactory system. In his view, learning implements a bias towards certain dynamical states in the olfactory system, much like it does in a Hopfield net. This bias comes from reinforcement in some of the mitro-mitral cell connections (Gray, Freeman & Skinner, 1986). In the exhalation phase, the olfactive system is in a quiescent mode, where chaotic low-activity takes the network state close to different basins of attraction in turn. When the inspiration comes, overall arousal increases the activity, thereby placing the system in a zone of its transfer function where the gain is much higher. This has two consequences: first, the state of the neural tissue will be oscillatory; second, the increased gain magnifies the attraction basins. If the input fits one of the learned odors, and the state of the system is anywhere near the corresponding basin, its dynamics converges into it, and the burst takes on one of the stereotyped forms. Otherwise, odor identification waits for a later sniff.

A few consequences of Freeman's model which he does not emphasize are still worth considering. One point which Freeman fails to make is why the significant bursts have higher and more concentrated frequency: this can be explained qualitatively by the property of intensity to frequency translation exhibited by the Wilson and Cowan model: when the odor is recognized, the system "frustration" (in the Hopfield sense) diminishes considerably after the convergence; this should result in higher activity. However, as shown by the Wilson and Cowan model, activity is limited by the negative feedback loop, so that higher activity actually translates into higher frequency, which is the experimental observation. In this view, one also realizes that there is a functional significance for the inspiration-expiration modulation of basal activity: it acts as a reset mechanism to get the system out of the attractor. An important functional aspect of this view also pops to the eye, although Freeman does not insist on it. Since all the olfactory bulb is involved in the coding of one odor, and that each individual odor is the center of an attraction basin, the olfactory bulb can only encode one odor at a time - even if states not corresponding to an odor were steady, they would not be a linear combination of odors but, like a chimera in a Hopfield net, they would present a new irreducible structure.

In turn, some of Freeman's assertions are questionable. The idea of "olfactory fovea", for example, is not consistent with the homogeneity of the relevant spatial frequency which Freeman and Baird demonstrate. It might be more appropriate to suggest that this "individual signature" is a side-effect of the strengthening of mitro-mitral connections throughout the rabbit's lifetime. The non-linearity could result from increased positive feedback acting to break the spatial near-homogeneity that derives from initial random odor encoding. Another point, the 50 % rate of useless bursts could well be not an experimental data processing artifact, but a consequence of the storage-accessibility tradeoff and of limitations of attractor networks: basins of attractions cannot be made all-encompassing. The ratio of the number of odors (sometimes several thousands in humans) to that of glomeruli (2000 in the rabbit, 3200 in the cat) suggests that some odors cannot have full-size attraction basins, especially since the attractor has only one side to work with –the low baseline does not allow high negative activity rates, and in this theory, Hebbian learning only affects excitatory connections. Finally, since his early work could not find significant spatial encoding for odors, Freeman settled on conditioning the rabbits. Such experimental design bears an ambiguity in the results as to whether the bursts are elicited because the rabbits are strongly expecting one odor or the other, or just because they know the odor. This confound should be disambiguated by Freeman's early experiments. Yet, his later theory states that one should discard the individual basic amplitude pattern by normalization before considering the spatial patterns – something not done initially. This suggests that early results do not have the finesse of analysis necessary to disambiguate this issue. Freeman's theory also presents a problem regarding odor-learning: when no known odor has been recognized, that is, when the system is sending one of the low-frequency bursts, the system cannot know whether the state corresponds to a new odor to be learnt or to an undefined computational state due to failed recognition. Moreover, even when a new odor is actually present, the state is not merely the raw input, but rather a transform of the sensorium corresponding to this new odor, where all the previous biases tend to overactivate the mitral cells who are already more strongly connected. In summary, beyond the need for further experimental validation, which Freeman himself points out, his potent picture still contains a host of theoretical points to be addressed.

Freeman on the visual system Freeman has done little experimental work on the visual system. With van Dijk (1987), he studied recordings in one monkey visual cortex with similar methods as the olfactory bulb recordings. The monkey was fixating a square with a flickering pattern and had to press a lever when the stimulus stopped flickering. Electrodes covered two windows: a 2.4x1.4 cm patch of striate cortex and a 3.6x2.8 cm patch of prestriate cortex. Statistical analysis allowed to find spatial patterns differentiating the two conditions (67 % correct classification in the early part of each trial, the chance level being 50 %). Like in the olfactory bulb data, these patterns were indifferent to the temporal frequency and had a high degree of variability. The main difference between the visual and the olfactive system was that the visual EEG had no burst, and, consequently, no stable peak in the gamma (20-90 Hz) range. The authors suggest that the proper correlate of bursting, which is respiration in the olfactory system, may be found in eye movement for the visual system. This work is not as useful as that in the olfactory bulb, especially because of the lack of temporal structure in the signal.

In the theoretical realm, Freeman has suggested that some aspects of his theory of the olfactory system also apply to the visual system. In particular (Freeman, 1994a), the role of the gamma-band oscillation as a carrier wave rather than an encoding wave has lead him to propose that the much famed "40 Hz" oscillation measured by Gray and Singer (1989) in the visual cortex is not, as they suggest, a binding activity of feature detectors; rather, it is the construction of holistic perceptual images by the cortex. Freeman also generalizes the attractor approach to all sensory cortices (Freeman, 1994b). The irreproducibility of the temporal dynamics is not seen as random, but rather as a chaotic dynamics. Since the temporal wave is irrelevant to the meaning, this irreproducibility is no hindrance to perception; rather, it allows the dynamics to sample the phase space in many dimensions, thereby visiting the vicinity of many sub-attractors, which Freeman calls "wings", so that any of them can be activated if the sensorial input or the corticofugal input biases the system to do so. Freeman suggests that this picture applies equally well to all sensory systems, including the visual system.

This last view of Freeman again calls for a major comment: in the olfactory system, there is no notion of space. As we have seen, one consequence of his view is that the attractor "percept-detector" can only identify one thing at a time. On this

ground, it would be more reasonable to suggest that, if Freeman's theory does apply anywhere in the visual system, it only does so in the ventral stream, at a level where the representation is no longer pictorial.

In a double conclusion on the review of Freeman's findings and on the overall presentation of attractor networks, it is good to insist on the variety of possible dynamics. The two properties which are essential for our theory are pattern completion and bistability. They are both presented by many systems with very different dynamics. In particular we have seen them illustrated in simple models such as the Brain-State-in-a-Box or the Hopfield net, by the more complex model of Wilson and Cowan, and by the olfactory cortex of the rabbit described by Freeman. Having made the point that these properties are almost universal amongst attractor networks, we now turn to their theoretical consequences for the visual system, and in particular for mental imagery.

II.C.2 Consequences of bistability and of pattern completion

Now that we have amply argued that these two properties, bistability and pattern completion, are widely spread, let us consider the theoretical consequences.

Bistability and the dual regime: activation or priming

What is the consequence of having bistable cortical regions? Bistability clearly defines two widely separated regimes of neural activity. When activity is in the low range, the cortical region is only "primed"; when in the high range, we want to call it "activated".

We cannot at this point commit to the nature of the "activated" regime, but the alternative possibilities share the same computational features. Whether it be a transient excitation –as Wilson and Cowan suggest–, a steady-state with hysteresis properties, or a dampened oscillation –as proposes Freeman–, the basic phenomenon remains that non-linearities implement two qualitatively distinct responses to input under or above some threshold. All these candidates also share the property that this threshold is determined by feedback connections, and that its value can be modulated –along with other quantitative properties of the response, by the gain of this feedback.

What might be the functional utility of such a mechanism ? It implements a perceptual threshold. There is an advantage to having a perceptual threshold. Non-linearity implements it by making above-threshold input result in output a full order of magnitude stronger. In that order of things, the explanational power of the dual-mode view encompasses that of all activation and threshold models in classical cognitive psychology. Therefore, the question applies more to the necessity of having the same mechanism over and over again at different levels of sensorial processing. We believe that the answer lies in the computational relevance of extracting features which are the basis for computation at a higher level. If shape identification relies on Non Accidental Properties, such as parallel lines, then it is relevant to extract these and amplify them non-linearly, even though the same kind of non-linear mechanism will again be used later at object identification, this time to implement the perceptual threshold. Since the task of the visual system is to discard the overwhelming flow of sensorial visual information, it makes sense to have radically non-linear behavior at every stage.

Finally, on a speculative note, bistability allows for one pleasant property: the all-or-none quality of conscious experience. If consciousness is somehow dependent upon the level of activity in the brain area encoding the conscious representation, then bistability insures that this activity, and hence the apprehension we have of this representation, is all-or-none.

Pattern completion and well-formed representations

The tendency to complete imperfect patterns is of course useful for imagery. Notwithstanding, it is also useful during perception.

Pattern completion during imagery can help with two defects of the signal fed back by higher visual centers. First, it can disambiguate imprecision. Long ranging feedback connections in the visual cortex are more diffuse than feedforward connections. This means that the representation suggested by imagery is blurred in whatever representational space the target layer implements. For example, feedback to orientation selective microcolumns will overlap into neighboring columns, making each feature of the mental image somewhat imprecise in orientation. In that case, the pattern-completion mechanism will select the most likely orientation for each feature and shut down the

others. Second, pattern completion can fill in parts of the representation which are not restored by top-down input. One must realize that higher representations most likely do not incorporate an explicit description of objects. Rather, they prioritize elements important for recognition. For example, a cube will be represented in a distorted way, where the corners will have more importance than the edges, and the edges than the faces. When visualizing a cube, it will be easy to restore the corners, but harder to summon the edges. The role of pattern completion is then to fill-in the representation with items discarded during the recognition and memorization processes. Thus, pattern completion makes mental images both more complete and sharper.

During vision, pattern completion can serve to help recognize degraded or occluded visual stimuli. It will restitute whole parts of objects or, when necessary, will improve the representation of a scene seen with insufficient illumination or focus. It is, in particular, likely that pattern completion is essential for scotopic vision (night vision).

Thus, in both imagery and perception, pattern completion reshapes the activity pattern into a better-formed pattern. The argument may appear circular, but it is not – let us restate it. The "weight matrix", that is, the connectivity [21], determines a family of particular states (fixed points, cycles, or more complex attractors). These are the "well-formed representations". Attractor dynamics makes a *any* representation converge towards one of these. Thus, even if there are myriads of such attractors, the dynamics pulls apart representations of different things and classify together representations of the same things. At a low level, the classification may differentiate between objects with a given orientation and location, and there are likely many possible attractors, even for very similar views of the same object. At a higher level, classification is more conceptual. Thus, pattern completion implements classification at every level, reshaping the activity pattern into a well-formed representation.

II.C.3 What are imageries? – Imagery at different levels

We suggest that imagery is **activation**, not merely **priming**, of neural layers in the visual processing stream. Thus, because of bistability, each visual region of the

[21] We assume here that the connectivity stays basically inchanged at the time scales of the few seconds used for scene perception.

brain is either involved or not in the present configuration of active regions. The set of **primed** regions determines the type of mental imagery.

Obviously, not all possible degrees of freedom are frequently used. Some configurations are highly unlikely, especially those where there are strong connections from an active to an inactive region. For example, it is hard to conceive a non-pathological regime of activity where V1 would be activated and not V2. Thus, as a first approximation, we propose that for the imagery of objects, all regions "above" the lowest activated region are also activated. In this view, there are a few different kinds of imagery, and these discrete categories arise from activation at different depths of the visual pathways. At this stage, at least three levels clearly emerge from the data: eidetic imagery, ordinary pictorial imagery, and non-pictorial imagery.

The reticence to attribute these different levels of imagery to qualitatively different processes should disappear if one considers that, in a dual-mode functioning, minute changes in input or in connectivity gain can decide that some areas will have activation instead of priming. For example, Freeman showed by looking at bursts of activity in the rabbit olfactory bulb that a mere change in arousal (that is, in non-specific input) can result in a dramatic change in dynamical regime. Other factors, such as norepinephrine release can change the gain of local connections (Waterhouse, Ausim Azizi, Burne & Woodward, 1990; Woodward, Moises, Waterhouse, Yeh & Cheun, 1991). Finally, it is not excluded that attentional focus on different properties prompts different lookup mechanisms, or simply arouses primarily different areas, resulting in some being activated or not depending on the task. Moreover, it appears that the firing threshold of cortical cells lowers if the cell is not stimulated for several minutes (Gilbert & Wiesel, 1989; Gilbert & Wiesel, 1990; Das & Gilbert, 1995b; Das & Gilbert, 1995a). Thus, details of personal history can determine the occurrence or the failure of imagery vividness in a particular situation.

In this view, it becomes clear that not only are there imageries rather than one imagery, but also most of us can experience in turn different varieties of imagery. Thus, the apparent graduation between good imagers and bad imagers is the result of considering the average type of imagery, which corresponds to no real type of imagery. Just as in quantum physics, where a particle state is the superposition of eigenstates

with different probability values, but where only one of these eigenstates can be observed, personal imagery is the superposition of a range of imagey types with different occurrence frequencies. None of us dreams in "dirty watercolors". Some of us dream sometimes in color, and sometimes in black and white – not in between!

II.C.4 The interplay of imagery and perception

Along with Kosslyn, we think that imagery is part of perception in two ways. First, there are permanently on-line reciprocal or feedback connections assigned to cooperative computation. These reciprocal connections between two areas cause some activity in the lower area whenever there is some in the higher. Second, there are active lookup mechanisms which, in the absence of input, provoke imagery. Contrarily to on-line feedback, lookup mechanisms can be engaged or disengaged, and can therefore occasionally be under volitive control. It seems likely that the permanent feedback by itself only induces **priming;** it biases the dynamics of the lower layer towards states that make sense only at higher stages of processing. On the other hand, the effects of active lookup always cumulate with those of background feedback, and, as Kosslyn and Ullman (1995) suggest, these active lookup processes are probably the mechanisms of ordinary imagery.

In our view, non-pictorial object representations are easily **activated** from memory, but if the sensorial information is degraded, they are only **primed** by visual imput. In that case, it is the role of associative memory to suggest interpretation of the sensorial data. The ultimate example of such functioning is the Rorsach test, where sharp an unambiguous sensorial information gives considerable leeway to the high level visual representations. Reciprocally, low-level representations are data-driven, that is, almost systematically activated by the incoming sensorium, but only primed by memory.

The distinction between automatic and engaged mechanisms also bears on the issue of whether vivid imagery is an epiphenomenon in psychological tasks. After Moscovitch, Behrman and Winocur (1994), one may suggest that on-line perception-enhancement mechanisms create activity in pictorial areas which is epiphenomenal and superfluous in the accomplishment of imagery. The answer to that objection is that these on-line mechanisms probably cause only priming, not activation; ergo, if activation is observed, it has been caused by more than these permanent mechanisms, and is not

epiphenomenal, or at least, not by their mediation.

II.C.5 Consequences and predictions

First, we would like to suggest behavioral tasks which may help differentiate whether V1 or V4 is the locus of some type of imagery. There are, in monkeys, a number of properties known to be implemented by extrastriate, not striate cortex. Amongst these are the resolution of the aperture problem, and the discarding of color illumination. These properties suggest tasks where inspection of an image in V1 or V4 might give different results. For example, if the inspection were made on a V1 representation, visualizing a pattern of diagonal lines moving smoothly from left to right would result in seeing them moving perpendicularly to their orientation, e.g., rightwards and upwards. Of course, it might be argued that the V1 representation is disambiguated by keeping the notion of horizontal movement somewhere else, for example in MT. This interpretation would be consistent with the lesion studies which show that integration is not as necessary in imagery as in perception. However, if one subscribes to this alternative interpretation, one must admit that the V1 activation during imagery is an epiphenomenon.

Tests of the dual-mode processing itself can be based on the fact that cumulated priming leads to activation. Thus, one could test the same population for eidetic imagery twice, by projecting a image to be studied on a screen, then asking the subject to visualize it on the screen. On the second test, subliminal images of the scene would be projected during the recall period. In that situation, one might expect more subjects to demonstrate eidetic imagery.

II.D Imagery debates revisited

II.D.1 Variety of imagery types

Our theory is fully compatible with the variety of imagery types observed. Because the representational space of a mental image can be made of almost any combination of the possible attributes of picture, there is no need for a procrustean bed. To convince the reader of this variety, we present in this section a short review of different imagery types. (Galton, 1907) was the first to center a study of imagery onto individ-

ual differences. Following a classification owed to Richardson (1969), we will review the various kinds of imagery, before focusing on the most common one. Richardson distinguishes four categories of imagery: after-imagery, where images can be directly related to a previous sensorial stimulus, and where subjects have no control over them; eidetic imagery, a type of extremely vivid imagery; "memory" imagery —our garden-variety of imagery— which allows us to remember pictures, or build pictures from known parts; "altered states" imagery, which depart somewhat from the previous categories because it is artificially induced[22] In addition to these gross differences, our theory accounts for the erratic occurrence of color in dreams, and other minor variations of imagery type [23].

Types of very vivid imagery

After exposure to a bright or prolonged fixed image, a positive after-image develops, soon followed by a negative one. After-images are very vivid, comparable to the actual image, have short duration, move with the eyes, and appear "projected" onto the visual background. Their apparent size follows Emmert's law, which states that the visual angle of after-images stays constant, so that the apparent size grows linearly with the distance to the background. A control experiment (Oswald, 1957) demonstrates that this effect cannot be faked because the accuracy of this projection effect is greater than that of distance-to-background estimation. After-images can be obtained from retinal stimulation only (Craik, 1940), or from pure imagery (Oswald, 1957). Craik demonstrated that stimulation with a temporarily paralyzed optic nerve created a negative after-image after restoration of optic nerve function). Oswald found negative after-images to imaged stimulus.

Another kind of very vivid image is the recurrent or perseverative image, which can occur after prolonged attention to a variety of stimuli. For example, when relaxing

[22] In addition to these types, there are pathological situations where special cases of imagery are created by brain lesions (Bisiach & Luzzatti, 1979; Farah, 1988; Farah, 1989c). We do not detail these results here, because they are reviewed in another section (**Imagery in the brain**).

[23] A personal communication by Pr. Ramachandran suggests the existence of an imagery of "pure movement". It can be experienced by closing one's eyes and sweeping with one's hand the visual field just in front of the eyes, and thinking about the same hand, vaguely enough to avoid visualization. Despite the absence of a mental image of the hand, one can experience a mental image of the hand's movement direction and speed.

in bed after driving all day, one can see the road scroll and very vivid moving images of driving-like situations pass. Like the after-images, these images are out of the subject's control, but unlike the after-imagery, this imagery is idealized, and the images evoked are not necessarily those of actual earlier sensations.

The third kind of natural very vivid imagery is eidetic imagery. Eidetic imagery is the ability to form a percept-like picture of a remembered or fantasized object or scene (Miller & Peacock, 1982; Haber, 1988). The most objective test of eidetic imagery is retaining the visual image of a presented stimulus for several minutes; the main differences between an eidetic and an after-image are the location invariance with eye movements (eidetic images can be scanned) and the retention duration for a positive image. Eidetic images can be erased by blinking, are hindered or prevented by naming the objects during presentation, and fade away after a few minutes with entire objects or people fading away one after the other. Eidetic imagery has long been considered to be primarily a transient phenomenon, mostly developed in children under 12 (maybe in up to 10 % of children), but very rarely in adults (Richardson & Harris, 1986), although it may reappear in elderly, which is still a point of dissension (Giray & al., 1985; Zelhart, Markley & Bieker, 1985). Eidetic imagery seems to be anti-correlated with literacy between Western and African cultures (Doob, 1965), but not within Western culture (Paivio & Cohen, 1979). The theory that eidetic imagery was a primitive function, bound to disappear with development, has not stood the test of longitudinal studies (Haber, 1988). Rather, the rarity of reported cases among adults must be attributed to increased naming of examined objects, which blocks eidetic imagery. Eidetic imagery can be monocular (Haber, 1988), and also can cause electro-retinogram responses (Kunzendorf, 1984). These two facts suggest occasional centrifugally-induced retinal involvement in eidetic imagery. Success at projecting an eidetic image may depend on the motivational state of the subject, and on the plausibility of the imaged object. There appears to be among eidetikers differences in performance which modulate the complexity, the duration of, and the control over the projection. Thus, besides "ordinary eidetikers", there are a few reported cases of extremely vivid and controlled eidetik imagery, for example (Luria, 1968; Purdy, 1936). Regardless of these subdivisions within eidetic imagery, it appears as a discretely distinct phenomenon rather than a degree of vividness along the continuum of ordinary imagery.

Very vivid imagery can also be induced by various artificial means. Hallucinogenic drugs, such as mescaline, can produce eidetic imagery and synesthesia in normally non-eidetic subjects. The same drugs can also induce geometric visual patterns with life-like vividness, where the subjects sometimes see meaningful images (much like a Rorsach test), in which case the reported image replaces the pattern. Falling asleep sometimes produces a type of very vivid imagery called hypnagogic imagery. Hypnagogic images tend to be autonomous and of short duration. Pulsed stimulation, either electric or, more frequently, from a stroboscope, induces geometric images in about one third of subjects. Finally, sleep deprivation, meditation, and perceptual isolation also induce very vivid imagery in a proportion of subjects ranging from one third to one half. In all these situations, the line between imagery and hallucination is thin, and may be defined as the boundary between being conscious or not conscious of the unreality of the image. Especially, habitual good imagers or eidetikers may have a better experience of very vivid imagery and, therefore, identify as "imagery" phenomenal experiences which would be hallucinations for poor imagers.

Ordinary imagery

To ordinary people, imagery does not mean anything as fancy as eidetic or mescaline-induced life-like images. Rather, it is a more or less visual reconstruction of images of scenes or objects remembered or made up of known parts. It is markedly less vivid than perception, and there are wide individual differences in imagery vividness.

A major difficulty in imagery studies is the assessment of vividness. For long, this was only done by self-assessment questionnaires, the most famous being that of Betts (1909). The objectivity of Bett's test was asserted because its results correlated with accuracy of recall of images - hence a confound with memory. Sheehan (1966) introduced a test where recognition replaces recall: a projector with controllable settings creates color squares of variable luminance, size, and border sharpness. The subject first sees a pattern made by the experimenter, forms a mental image of it, and then is left free to play with the controls until he thinks he has recreated the original square. The proximity of the first image to the second can be measured by Euclidian distance in the space of the parameters. Sheehan showed that the results of this test correlated with an abbreviated

form of Bretts' test, now known as Sheehan's test. Hebb (1968) asks subjects to spell backwards an orally presented word, something that should be easier when visualizing the written word. All these measures are only an estimate; the self-tests depend heavily on previous explanations given by the experimenter about what is a mental image, so that the test is mostly valid within a population submitted homogeneously to the same test with the same experimenter. In brief, there is no metric measure of imagery vividness, but rather a relative measure sorting subjects at different levels of vividness.

Individual differences are another major focus of vividness studies. Apart from eidetikers, who form a class all of their own, it appears that about 10 % of subjects are habitual visualizers, who rely heavily on visual imagery for many mental tasks. Also 10 % of subjects are habitual verbalizers, who use verbal and propositional approaches as much as possible. The vast majority of subjects switch from one mode to the other according to the task. These subpopulations are extremes on a continuum rather than discrete classes. Besides the self-report tests discussed above, position on the visualizer/verbalizer axis can be inferred from EEG studies by measuring the degree of alpha blocking in various tasks (Golla, Hutton & Walter, 1943). The alpha wave is a low frequency wave of big amplitude that dominates the EEG in awake subjects relaxing with their eyes closed. It is blocked by eye opening or, even with eyes closed, by strong arousal and by vivid imagery. Both the total duration of alpha-blocking during imagery and the relative smaller amplitude of alpha EEG over eyes-open EEG (beta waves) can be taken as indexes of vividness, which demonstrates wide differences between subjects. Although there was a confound between the effect of imagery and general arousal, it has since been disambiguated by Davidson and Schwartz (1977), who conclude that the ratio of occipital to general alpha indeed decreases during imagery tasks. Another objective measure is the regularity of the breathing pattern: verbal problem-solving tends to disturb the regular breathing pattern (supposedly due to subvocalic breath modulation), so that habitual verbalizers tend to have irregular breathing patterns, whereas visualizers have very regular ones (Golla & Antonovitch, 1929). The distinction does not appear to be correlated with IQ, but there might be some advantage for visualizers to solve concrete problems and for non-visualizers to solve more abstract ones. It seems that word meaning has less variance for habitual verbalizers, but there could be a confound with the effects of vocabulary size.

Finally, being a visualizer correlates with having higher hypnotizability and more field-dependence (i.e., depending more on context in problem-solving and other tasks). In summary, there objectively appear to be wide differences among individuals, and these differences correspond to differences in reported phenomenal experience. No absolute scale of imagery vividness has been constructed yet.

II.D.2 Where in the brain ? : the neuropsychological debate

One of the focal points of the discussion is whether the medium of imagery is or not pictorial, that is, laid out in space. Pylyshyn argues against the hypothesis that imagery is pictorial. Kosslyn insists that the visual buffer is the locus of imagery. In terms of anatomy, the question becomes whether imagery necessitates activation of the retinotopic areas located in the occipital lobe. Until recently, there were few means of conducting direct investigation of the human brain in the act of functioning. However, this last decade has seen the advent of many methods of neuro-imaging, which have been used to study the human brain during imagery tasks. The direct data brought by imaging methods can be related to indirect data from case studies of brain lesioned patients, in an approach championed by Martha Farah. In addition, a recent debate (July 1994) in "Trends in NeuroSciences", confronting opponents and supporters of pictorial representation, summarized the state-of-the art results using brain imaging methods. Finally, because imagery is conscious, one must consider the arguments raised by Francis Crick and Christof Koch about which visual areas support activity correlated to awareness. We will conclude with a "best bet" rather than a definitive conclusion, for the locus of ordinary pictorial imagery. The best bet is V4.

Martha Farah's arguments

The first question addressed by Martha Farah (1988) is the same raised by Pylyshyn: are the representations underlying mental imagery shared with perception ? She first reviews the main results in Cognitive Psychology (Finke, 1985): mental rotation; equivalence of mental images and visual percepts for many tasks such as visio-motor adaptation, McCollough effect, and spatial resolution of mental image. Pylyshyn's arguments about these findings can be summarized in three points: use of tacit knowledge by

the subjects, experimenter expectancy, and tasks demands. Another argument against the interpretation of these results in favor of shared representations is that some of the results have been replicated with congenitally blind subjects.

Martha Farah states that the classical approach to imagery, using behavioral results on healthy subjects, is not sufficient to disambiguate the question, and proposes to bring neuropsychological arguments into the debate. Roland and Friberg (Roland & Friberg, 1985) PET study showed activation of several visual areas when subjects imaged a walk in the streets. Goldenberg, Podreka, Steiner and Willmes (1987) , in a SPECT study, ask subjects to memorize words with either a verbal or a visual strategy. The words could be meaningless, abstract or concrete. They found that when using imagery on the concrete words, subjects had activation of the medial occipital and inferior temporal lobes. Davidson and Schwartz (1977) studied the localization of alpha waves of the EEG in subjects performing either visual imagery, kinesthetic imagery, or both. They compared the EEGs at a left occipital electrode and at a left sensorimotor electrode. It was found that visual imagery decreased the alpha level by 10% selectively over the left occipital area, and that sensorimotor imagery decreased the alpha level by 6% selectively over the left sensorimotor area. These effects were additive. Farah, Peronnet, Gonon and Girard (1988) measured ERP during letter-detection in three situations: when the subjects were imaging the right letter –that which they had to detect–; when they were imaging the wrong letter; and when they were not imaging. The "right letter" imagery condition enhanced an early visual component of the ERP, N1, which has a latency of 170 ms and could be generated by prestriate cortex (Maier, Dgnelie, Spekreijse & van Dijk, 1987; Gomez Gonzales, Clark, Fan, Luck & Hillyard, 1994). Both imagery situations enhanced the next ERP component, P2. Spatially, the enhancement was at the occipital pole for N1 –just like the basic wave– and more temporo-occipital for P2. These results strongly suggest that the ordinary processes of perception are boosted by imagery. Thus, electrophysiology and blood flow studies agree that imagery engages visual cortex.

Moreover, brain lesion studies strengthen this point of view. A very salient fact is that cerebral color blindness is associated with loss of color imagery. Lesions to the ventral or the dorsal stream produce parallel effects on vision and imagery. Levine, Warach and Farah (1985) studied two patients, one with bilateral parieto-occipital dam-

age, and the other with bilateral temporo-occipital damage. The former had impaired localization of objects in space, and was likewise unable to point at targets on a mental image; however, he could describe objects or faces mentally imaged. The second patient had object agnosia and was unable to draw or describe objects from memory, even though he could draw and describe in detail the locations of cities on a map. Reviewing the literature, Farah found that 14 cases were reported to have parallel agnosia and imagery impairment against only 3 having intact imagery despite agnosia; 9 cases of visual disorientation were paralleled in imagery, against 3 which were not. In finer cases of disability, such as selective agnosia for living versus non-living things, or specific face agnosia (prosopagnosia), imagery deficits also parallel the perceptual deficit. Left visual neglect, an impairment usually following right parietal lesion, yields neglect of the left side of imaged scenes (Bisiach & Luzzatti, 1979). In brief, neuropsychology shows that there is salient parallelism between perceptual and imagery impairments and preservation for a range of different brain lesions, supporting the argument that the systems which implement imagery are shared with perception.

Back to Fodor's arguments, whereas one can admit that daily perception gives subjects a detailed tacit knowledge of the functions and subtleties of their visual systems, it cannot inform them at all about underlying anatomy and physiology. Therefore, it is very unlikely that tacit knowledge could account for these blood flow, electroencephalography, and lesion results. Task demands should, if anything, prompt patients to behave just like normal subjects; and involuntary simulation without real image should in no case generate the same blood flow increments as actual vision. Experimenter expectancy arguments, although unlikely, could apply to the neuropsychology data, but quite not to the blood flow and EEG measures. Finally, imagery in congenitally blind patients need not be visual to be spatial; it could pertain to another modality. In conclusion, the first question of whether imagery is visual or not is answered positively by PET, EEG, and neuropsychology studies.

A second question is to find out how the present neuropsychological data constraints the possible location of imagery systems. Most evidence points towards a relative (Kosslyn & Ochsner, 1994) or absolute (Farah, 1989b) dominance of the left hemisphere for generating images. Farah (1985a) projected uppercase letters to a patient with cal-

losal commissurectomy ("split brain") and asked him to classify them according to the height of their lowercase form. She found that the left, but not the right, hemisphere could perform the task. Kosslyn, Holtzman, Farah and Gazzaniga (1985) studied split-brain patients and found that both hemispheres were able to image single part images (such as holistic skeletal pictures of animals), but only the left hemisphere could image composite parts (such as an animal's head plus the ears, or letters). Besides the hemispheric specialization, some evidence points towards the use of visual topographic areas in imagery, probably in the prestriate cortex. Arditi, Holtzman and Kosslyn (1988) tested congenitally blind patients and found that their "imagery" did not respect simple geometric properties. For example, pointing towards the right and left boundaries of the same object imaged at different distances did not yield the geometric relation of angle with distance, as it does in seeing people. Farah, Soso and Dasheiff (1992) tested visual angle for a patient who had one occipital lobe removed. The visual angle diminished by one half after the operation. Thus, neuropsychology suggests that some occipital cortex is involved in imagery.

The TINS July 1994 debate

Trends In NeuroSciences held in its July 1994 issue a debate opposing proponents and adversaries of the involvement of occipital cortex in imagery.

In the opening paper, Per Roland and Balasz Gulyas argue that imagery is only implemented by the higher visual perceptual areas. In their view, the visual areas subserving imagery are the association areas, which do not have pictorial coding. They quote several PET studies where the striate cortex is not activated, and where, usually, no other occipital area is either. In Roland, Eriksson, Stone-Elander and Widen (1987), subjects had to image a walk from their home in their neighborhood. The baseline task was relaxed awake rest with closed eyes. The authors reported activation in several frontal areas, and in various parietal and temporal areas which can be assumed to be visual association areas. The closest activated structures to the occipital pole are the superior occipital cortex on the lateral side of the occipital pole, and the most posterior parietal cortex (next to the calcarine cortex). In Decety, Kawashima, Gulyas and Roland (1992), subjects kept an image of several targets for 15-90 seconds before reaching; PET measure

was taken during this image sustain phase. Several areas were activated, including prefrontal cortex, ventro-lateral thalamus, cerebellar vermis, and several parieto-occipital visual association areas, but no occipital area. In a very similar study, Kawashita, Roland and O'Sullivan (1994, Kawashita, Roland and O'Sullivan (1995) found occipital activation on the right superior occipital gyrus and the left cuneus, but not in V1 or V2. Other studies by the same authors also found no evidence of activation in the striate cortex. Roland & Gulyas (1995) studied visual learning, recognition, and recall of complex geometrical patterns, and find no activation of the topographic cortical maps. In conclusion, the authors argue that imagery involves supero-latero frontal activity –which does not occur in vision–, and activation of parieto-occipital and parieto-temporal visual association areas, which are not retinotopic, and are also involved in visual tasks.

In the second paper, Kuniyoshi Sakai and Yasushi Miyashita support the theory that visual imagery is implemented by the same neural mechanism that subserves memory retrieval in vision. They remind us of the influence of attention on early visual areas such as V4 (Moran & Desimone, 1985). They point out the technical difficulties of differentiating activation of V1 from that of V2 with neuroimaging techniques, and the inherent variability of PET data due to the small number of trials. They argue that, although it is not the most activated representation, V1 is activated during some studies (Le Bihan, Turner, Zeffiro, Cuenod, Jezzard & Bonnerot, 1993) for most subjects. They suggest that the activation or lack thereof of V1 depends on the task: in all imagery tasks imagery uses visual memory retrieval, which results in activation of the associative areas Roland and Gulyas speak of. In some tasks, scrutiny of the details places attentional demands and is achieved with the participation of V1.

The third paper is a comment by Kosslyn and Ochsner, who summarize evidence supporting activation of the occipital visual areas, and suggest three different causes for the inconsistent results from Roland's laboratory. The first reason is the difference in baseline task. Most of Roland's studies use rest as a baseline task (that whose resultant brain activation is subtracted from that of the experimental task). In contrast, studies like Kosslyn's (1993) use additive factors, in this case comparing imagery plus perception to perception alone. The second reason, a Pylyshyn-like argument, is that Roland's tasks may not have involved imagery. Kosslyn and Ochsner argue that Roland did not measure

direct "behavioral signatures" of imagery: their experiments are only PET experiments with no other measure; the assumption that subjects use imagery is made by Roland & al from subject report and other measures usually correlated with imagery, such as EEG alpha-blocking (Davidson & Schwartz, 1977) and breath rhythm - but these are subject to confounds, for example between effects of imagery and arousal. By contrast, Kosslyn's experiments are classical cognitive psychology experiments, where one factor is manipulated, reaction times recorded, and PET measures taken simultaneously. The third reason is that the results from Roland's laboratory did not isolate individuals and could have averaged out individually significant activation of V1. Given the importance of individual differences in imagery according to the classical literature (Richardson, 1969, reviewed above), the more recent SPECT studies(Goldenberg, Podreka, Steiner & Willmes, 1987; Goldenberg, Podreka & Steiner, 1990), and single-subject FMRI studies demonstrating greater blood flow in V1 for subjects reporting more vivid imagery, the composition of the subject pool may account for discrepant results. Kosslyn and Ochsner conclude that at least some of the time, imagery relies on descriptive representation in the occipital lobe, but that perhaps tasks which do not require reconstructing the details of a shape do not activate topographic cortical maps.

The fourth paper, by Moscovitch, Behrman and Winocur, is a plea in favor of using neuroimaging techniques combined with lesioned patients studies. The authors first revisit Kosslyn and Ochsner's arguments, and agree that "some, but perhaps not all, tests involving visual imagery activate areas 17 and 18". However, they insist that this activation is no proof that these areas are necessary for mental imagery. To do so, one must study patients with occipital cortex damage and patients with intact imagery and impaired perception. Well-documented current studies present ambiguous evidence. For example, Farah, Soso and Dasheiff (1992) studied an epileptic patient (MGS) before and after hemi-occipitolobotomy, and observed that visual angle was reduced to about one half by the operation. In contrast, Behrman, Winocur and Moscovitch (1992) studied the patient MK who exhibited preserved imagery despite a perceptual deficit in object identification. To account for these results, Moscovitch & al, like Kosslyn in "Image and Brain", suggest that integration is not necessary in imagery. They also insist that the way to resolve the question is to study such patients under PET. Finally, they point

out that the tasks used by Roland and Gulyas do not make as high demands on the topographic organization of the image as those of the tasks used by Kosslyn or Farah. To lift the ambiguity, they propose that task-analysis be combined with PET activation studies of normal and brain-lesioned people.

The fifth and last paper in the debate is Roland & Gulyas's reply to their colleague's comments. After agreeing with Sakai's and Miyashita's point on long term memory, they move on to the defense of their theory. They argue that the tasks do involve imagery, and that the alternative strategies suggested by Kosslyn are implausible. They convincingly defend the validity of their rest task and that of their indirect physiological measures of imagery. Finally, they bring up a new point: that spatially mapped representations can bring no computational advantage, or can even bring computational disadvantage, to spatial operations such as rotation or zooming. Their final word emphasizes their point of view: "We suggest that neuronal populations in remote temporo-occipital and parieto-occipital areas represent complex objects and scenes during imagery as well as during operations on imaged material, and we expect that these areas are organized non-retinotopically in man".

Crick and Koch's arguments on the locus of visual awareness

A recent review by Francis Crick and Christof Koch (1998) delineates several hypotheses about visual consciousness.

One of the key ideas that they put forward is that we would not be conscious of activity in striate cortex. Crick and Koch (1998) and Koch and Braun (1996) draw arguments from blindsight, and monkey experiments, but also from behavioral experiments. In particular, it appears that the chance of detecting an anomalous region in disparate right-eye and left-eye stimuli which, superposed, give a homogenous visual field, is much above chance, whereas the conscious experience is that of the homogenous field. Crick and Koch also argue that regions of conscious representations must somehow project directly to the frontal cortex, which is **not** the case of V1.

Besides this, there are hints that processing in the dorsal stream is in part or even mostly unconscious. These should also be related to temporal considerations. The ventral pathways are concerned with extracting (relatively) time-invariant properties

of the scene. Thus, they are likely wired to discard the disorganizing effect of eye movements and head movements. In contrast, the dorsal pathways have to control a variety of coordinating and orientation behaviors which change at a very fast time scale. However, it difficult to determine if there is truly no consciousness of location when there is none of shape, or rather if it becomes hard to express consciousness of location when thee is none of shape. Movement just like location is represented parietally, and Ramachandran's experience of perceiving "pure movement"[24] suggests that it can be conscious.

Thus, at this point it is unclear which areas of the brain can sustain representations that we can be conscious of.

Our best bet

Because we recognize at least three basic types of imageries: eidetic, pictorial and non-pictorial, we want to nuance Kosslyn's opinion.

Non-pictorial object imagery is probably mediated in the temporal lobe. To Kosslyn's arguments, one might add the suggestions of Sakai and Miyashita (1994) that recall of objects is made by activation of the temporal lobe visual areas. Because the temporal lobe has lost most spatial information, its relationship to visual data is very analogous to that of the olfactory bulb to the olfactory data. Though the temporal lobe is not the locus of irregular activity bursts like the olfactory system, in monkeys it appears to sustain regular oscillations when an image lasts (Nakamura, Mikami & Kubota, 1992). The frequency of temporal epilepsies in humans suggest that oscillatory behavior is the primary regime of the temporal areas. All this indicates that non-pictorial imagery is a spatial pattern completion in the temporal areas analogous to the construction of smell expectation that Freeman has found in the olfactory system.

Pictorial imagery is the core of the debate. Although the evidence is by no means conclusive, convergent results hint that it is implemented in the human equivalent of area V4. As summarized by Kosslyn, behavioral data points to several sources of pictorial imagery: object reconstruction, scene reconstruction, spatial organization of

[24] See above note. Experienced, for example, by moving one's arm in front of one's closed eyes.

parts not seen together, and "mental plotting", using a moving mental punctual target to draw lines. This suggests that the area implementing such imagery must be connected to many others; the richness of mental images, which include color and movement, also point to a highly connected area. As can be seen on Figure II.2, V4 fully fills in the role. The ambiguous PET data can be best explained by supposing that extrastriate rather than striate areas are activated by the imagery tasks. Single unit recordings in monkeys show that, unlike V1 or V2, V4 can be influenced directly by attention (Haenny & Schiller, 1988; Spitzer, Desimone & Moran, 1988), and are corroborated in humans by PET data (Corbetta, Miezin, Dobmeyer, Shulman & Petersen, 1991). Thus, we would like to say that *the most occipital area recruited by ordinary pictorial imagery* is *V4*.

Eidetic imagery could be implemented at any level from V4 downwards. There is very little physiological data about eidetic imagery, and the behavioral data suggests that there are again several stages of vividness. Interaction of eidetic imagery and after-images suggests some participation of very low-level structures. "Ordinary" eidetic imagery has properties such as the fading of the image part per part, something which suggest it could still be strongly interactive with scene analysis and recognition, and hence extrastriate, perhaps in V4 again. One tentative conjecture is that eidetic imagery involves several rather than only one occipital area, in a more interlocked dynamics which allows vivid images to stay in the "steady-state" rather than in the "active transient" regime.

Concerning localization, it is worth pointing out that V1 gets very little feedback compared to V4. According to Felleman and van Essen (1991), no temporal area projects back directly to V1. The retroprojection of V4 to V1 is very weak. Adding this to the PET data, where no study claim categorically to have found activation in V1 during imagery tasks, suggests that V1 is actually a poor candidate for localization, although without doubt there are individual cases of extremely vivid eidetic imagery where V1, and maybe even upstream structures, are involved in image sustain.

To come back to the points raised by Crick and Koch, clearly, our theory states that there must as many occipital levels of which we are conscious as there are varieties of vivid imagery. Thus if one subscribes only to two types of vivid imagery, ordinary and eidetic, one could argue that V4 and V2 must be areas of which we are conscious. If one

thinks that eidetic imagery includes at least two different subtypes, as it appears from the data, one is left with the uncomfortable task of locating ordinary imagery in posterior temporal cortex or extreme eidetic imagery –say, for example, Vienamin's (Luria, 1968)– in striate cortex. On the other hand, because the cases of extreme eidetic imagery are so rare, and because their coupling with synesthesia suggests pathological wiring, it is possible that for these very few persons, V1 projects also to the frontal cortex.

In conclusion, albeit our theory does not assign unambiguously brain regions to imagery types, it gives a framework for doing so when more experimental evidence will become available.

II.D.3 Imagery and reasoning : no longer a philosophical debate

The "imagery debate" which raged in the eighties superimposed at least two main lines of problems. The first line was whether imagery was pictorial, and by extension, how much of it shared the visual representations. The second line was whether imagery was used in reasoning. Classical introspectionist philosophy held that mental images were manipulated during reasoning. Even lately, Johnson-Laird's theory (1983) of *mental models* subscribed to that point of view. In this section we will first review some of the main opponent of mental image reasoning, Zenon Pylyshyn. We will then argue from the point of view of our theory.

Pylyshyn's approach

The first salvo of this battle, still being fought, was shot by Zenon Pylyshyn in his 1973 paper "What the mind's eye tells the mind's brain: a critique of mental imagery" (Pylyshyn, 1973). Pylyshyn did not deny the phenomenal experience of visual imagery; rather, he was opposed to the use of imagery as an explanatory construct in cognitive psychology. He pointed out that words like 'see' or 'hear' refer both to information processing and to phenomenal experience. These two have no other reason to be connected than the vague introspection we have that we somehow 'see' our mental images. On this ground, he suggested that these words are inadequate as primitive constructs, and that the connection between perception and imagery is a metaphor which has been taken too literally. Pylyshyn suggested that knowledge does not necessarily have to be encoded in

either pictures or sentences, but that there are a host of possible representations: descriptive symbol structure; propositional structure (a proposition being a truth-valuable symbolic structure); computer-like data structure; procedural structure. He argued that the amount of storage required for memorizing actual images would be prohibitive, and that communicating and acting on the world requires a finite set of conceptual categories. Pylyshyn insisted on the difference between symbolic structures and sentences. He finally suggested a symbolic information processing model of "imagery" tasks.

In later papers (1986), Pylyshyn attacked more specifically the experimental imagery tasks, especially tasks of mental rotation and mental scanning. His main argument was that these tasks reflect tacit knowledge of the subjects –who take the task to be a simulation of the act of witnessing– rather than intrinsic knowledge-independent properties of the medium of representation. He suggested that imagery could be an epiphenomenon prompted by habits left over by an early stage of mental development. He argued, a point well taken, that Kosslyn's scanning tasks give the illusion of bringing the explanatory power of physical laws such as $Time = \frac{Distance}{Speed}$ without spelling out the ontological claim that images are laid in space in the brain. He brought several demonstrations of how the belief system pervades imagery, he emphasized the difference of task demands in imagery experiments, and argued about the influence of experimenter expectancy. Pylyshyn's conclusion was that nothing proves that imagery occurs in a spatially out-laid medium, and he suggested instead that imagery is the experience which we have when we are actually manipulating symbolic descriptive structures in a way consistent with the experimental task demands.

Functional advantages of pictorial imagery

Can there be a functional advantage to having pictorial imagery ? As Roland and Gulyas put it, "if the brain already possesses a computed representation, then why do the job over again" ? Two sub-issues arise here: the putative role of imagery to "unpack" visually stored information (a long term memory role), and the role of visual images as temporary data representations that are easy to work on (a working memory role).

Concerning the relation of memory and imagery, one can argue from an infor-

mation processing point of view that low-level representations do not store information, and are therefore useless for recall. One might even add that fortunately V1 or V4 does not store information, or our recall would block perception... Yet, there is a fault in that argument. As Don Norman (private communication) points out in his theory of dreaming, constraints that are systematic in the world need not be taken into account by our perception apparatus. Indeed, these constraints create noise, not information, because they are always there. Of course, they must be taken into account by our actions, but this is the other end of the system. In the visual stream, it makes sense that constraints such as object continuity or viewpoint consistency may be discarded by the system, as information moves downstream. One important functional role of imagery could then be to ensure that mental constructs are consistent with these constraints of the world. This function is especially important in visual memory, because it allows to store visual memories more parsimoniously, since all the properties which could be deduced from constraints in the world will be restored by visualizing them. In this case, the low-level processing is able to do appropriate pattern completion; even though it knows nothing specific about the pattern it is completing, it just forces it to follow the general rule. In a sense, we may consider that the low-level processing stages do also encode knowledge, but it is species-specific rather than individual specific knowledge; it has been learned by evolution, and is about invariants in the world rather than about particular events or objects in an individual's life. Imagery is the mechanism whereby this knowledge is incorporated in recall.

The role of imagery as working memory parallels the role of vision as an interface to this external memory that is the world. Johnson Laird mental models theory (1983) drives the point home. Perception and action systems in the brain evolved together to help us deal with the world. It therefore makes sense that the best model we can make of the world is what perception gives us at some level. By the same above argument, some representations of the world encode more of the constraints of the world. Hence, in some of the tasks relying on these constraints, pictorial rather than propositional representations allow immediate access to the relevant information, and are privileged for solving the problem at hand.

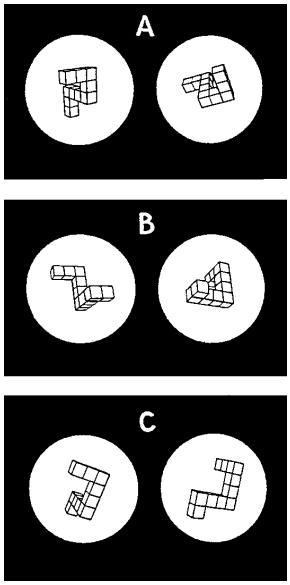

Figure II.13: **Stimuli used by Shepard and Metzler:** Subjects had to decide if both images in a pair were of the same 3-D object or if they were mirror image. The decision time depended on the angle of rotation. In A, the images differ by a rotation in the plane of the figure; in B, by a 3-D rotation; in C, by a rotation combined with a symmetry. (From Shepard, 1971).

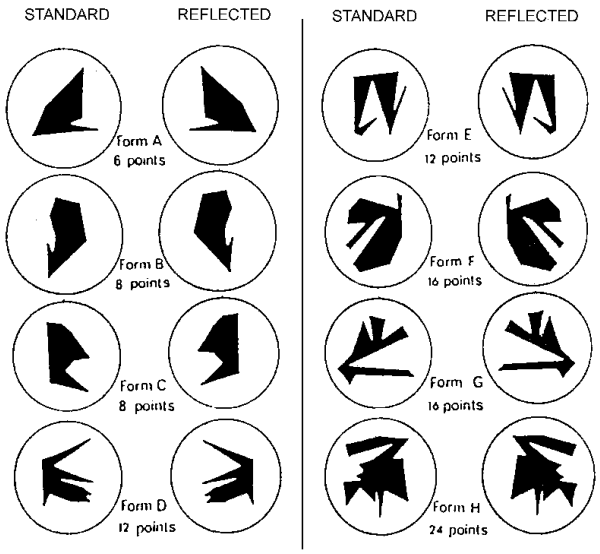

Figure II.14: **Stimuli used by Lynn Cooper:** the time to decide if two shapes are the same or mirror images depends solely on the angle of rotation, not on the complexty of the shapes. (From Cooper, 1975).

The many levels of thought

Very generally, it is well admitted that thinking can be conduced at many different levels of abstraction. This is particularly evident of mathematical reasoning. Now, the multiplicity of possible levels of abstraction forbids that their hierarchy correspond to a hierarchy or cortical areas or even sub-areas. Rather, it is by language that we create and multiply these levels of abstraction. We have *a priori* no reason to suppose a correspondence between abstraction and some particular dimension of representation.

However, in view of our theory of imagery, it appears that memories of objects at least can be formed with different levels of vividness. Thus, we would argue that not withstanding the scale of abstraction which language provides, there is another, hard-wired, scale of abstraction, at a lower level, and with a finite number of ladders. Most evidence from cognitive psychology concurs to argue that representations in that scale are being used for solving particular problems. The most famous tasks are Shepard's mental rotation (Shepard & Metzler, 1971; Cooper, 1975) and Kosslyn's mental scanning

Figure II.15: **The Treasure Island map of Steve Kosslyn:** subjects were asked to mentally travel from one landmark to another. Their time was found proportional to the distance between the two landmarks. (From Kosslyn, 1973).

task (Kosslyn, 1973).

Shepard and Metzler (1971) presented subjects with pairs of three-dimensional objects seen in perspective (see Figure II.13). The objects were made of ten small cubes arrayed in a bar with three elbows at right angle. The objects in each pair differed either by a rotation in the plane of the figure (Figure II.13.A), by a rotation in three dimensions (Figure II.13.B), or by a rotation in three dimensions combined with a mirror symmetry (Figure II.13.C). The reaction time to judge whether both objects were the same in three dimensions (as is the case in Figure 1.a and 1.b) or were different (as in Figure II.13.C) was a linear function of the rotation angle between both objects. Moreover, the same RT was obtained whether the rotation was in the plane or in 3D.

Cooper (1975) used random flat shapes as stimuli (Figure II.14). In her experiments, subjects learned the basic shapes and had to determine whether a presented shape was the original or its mirror-image. Again, she found a linear increase in RT with the rotation angle. She found no significant effect of shape complexity. It seems that all shapes were rotated approximately at the same pace. Finally, when subjects were

cued about the identity and orientation of the coming stimulus, they required a time proportional to the orientation to prepare themselves, but their reaction time afterwards became independent of orientation.

Kosslyn (1973) made subjects learn a fictional map (Figure II.15) with seven landmarks. Afterwards, subjects were given verbally a landmark name and had to say whether this landmark was on the map or not. In a first experiment, subjects had implicit instructions to image the map and picture a little black speck moving from the landmark they were seeing to the one they just heard the name of, and to press the button when they reached that point; another button had to be pressed if the named landmark was not on the map. Reaction times were proportional to the distance between the previous landmark and the one named. In a second experiment, subjects only had to answer as quickly as possible, without necessarily using the map. In that case, RTs were the same regardless of distance.

These three tasks illustrate the use of spatially arranged representations for problem-solving. Is, then, such simple problem-solving a form of reasoning?

Our answer would be definitely "Yes". Inasmuch as reasoning is the articulation of a sequence of thoughts, the articulation of a sequence of pictorial representations is reasoning. Or else, one has to restrict the meaning of "thought" to "verbal thoughts", and in that case the argument becomes circular. "Thoughts" should ad minima include whatever conscious representations are arranged sequentially in voluntary behavior. Mental images appear to qualify by that definition. We would even add that research on native reasoning abilities, based for example on Wason's task, suggests that correct logical reasoning is alien to the human mind, and can be acquired only at great effort. In contrast, imaged-based reasoning can be easier. Whether these images are pictorial or not, they nevertheless keep a much stronger relationship to the original stimuli they represent that the arbitrary relationship of a symbol to its reference.

II.E Conclusion: traps and detours of introspection

After this long tour, it is good to outline briefly our assertions, and to compare them with what intuition tells us.

We suggest that lateral connectivity makes each brain region a dynamical system. For that reason, brain regions have attractors, and the dynamics pulls the activity pattern towards those. This implements pattern completion. Simultaneously, attractor dynamics is the basis for a bistable regime, so that the participation of visual areas in vision or imagery is grossly all-or-none. In this view, vision and even more imagery are like a swiss-army knife: many configurations are possible, but only one is use at any time.

In particular, we suggest that the three grand categories of *imageries*: eidetic, pictorial, and abstract, differ by involving less and less occipital areas. Other, more subtle differences, such as dissociations of imaging color and shape, can also be predicted on the basis of our theory.

In retrospect, it seems that introspection has once more played a grand humorous trick on the psychologists. Because imagery *feels* like vision, introspectionnists –and, let us admit it, many recent psychologists too– have jumped onto the hypothesis of shared representations. For a few, the similarity between imagery and vision must have been strong, but for many, it was most likely a vague feeling, which was coupled with the traps of language. In reaction, some philosophers have denied any visual likeness to imagery. But between the procrustean bed of philosophic Imagery and the diehard pragmatism of a William James, for whom the variety of human experience is forever unending, there is room for a middle ground. Hopefully, this theory will have shown that the moderate complexity of the grossest levels of cortical organization can be matched by a moderate complexity in the types of imageries.

Chapter III

A Minimal Model of Mental Imagery by Recurrence-Induced Pattern Reconstruction

III.A Introduction

In the preceding chapter, we introduced the theory that visual areas can be bistable in virtue of local feedback. Lateral connections recirculate high activity, possibly maintaining the activity pattern. Low states of activity are recirculated but gradually decay. The difference between the sustain of high activity and the decay of low activity is a property of the non-linearity of the neuron's response. That difference is the foundation for the bistability: high activity states remain differentiated and encode a meaningful representation, low activity states correspond to the impoverishment of the representational space. Areas with too low an activity are not even part of the current set of representations. Thus the brain implements a representational space of variable geometry. We suggest that this mechanism is in particular true of the visual pathways during mental imagery. The variability of the representational space is the basis for the variability of mental imagery phenomena. In this way, bistability provides a theory to buttress Williams James' suggestion that there are "imaginations not Imagination".

In this chapter, we present a computer model to demonstrate that the theory

of bi-stable areas is sound. The properties announced can be mathematically proven for isolated neurons with a non-linear transfer function. When considering large ensembles, however, they are asserted by generalizing. Generalization from one to many units is rigorous when the units are linear, but with non-linear neurons, it is at best intuitive. By setting forth an actual model, we show the range and limitations of the theory, and hope to provide quantitative guidelines for the intuitive generalization.

Furthermore, the model is built using a bank of spatial wavelets (Gabor filters) as the main layer, replicating a representational space likely to exist in the occipital cortex. The representations encoded in the model can be projected back into an image, giving thereby a visual illustration of the phenomena going on in the model. Thus, beyond the general point of validating bi-stable theory, this model gives more concrete details, regarding very vivid or eidetic imagery phenomena.

III.A.1 Historical background

Mental imagery

Mental imagery was at the core of early psychology. Francis Galton, Wilhelm Wundt, Edward Titchener, and several others considered it was central to thinking. William James (1890) wrote, "In Locke's words, already quoted, 'the mind can frame unto itself no one new simple idea'. The originals of them all must have been given from without. Fantasy, or Imagination, are the names given to the faculty of reproducing copies of originals once felt. The imagination is called 'reproductive' when the copies are literal; 'productive' when elements from different originals are recombined so as to make new wholes." And Titchener, considering imagery central to mental life, asserts (1907) "There is no fundamental difference between the perception and the idea". In this early phase of experimental psychology, the salient fact that stood out was individual differences. Fechner first pointed this out, and his argument was supported by Galton (1907), who interviewed many subjects, and found in particular, a negative correlation between scholarship and imagery vividness: "To my astonishment I found that the great majority of the men of science to whom I first applied protested that mental imagery was unknown to them, and they looked on me as fanciful and fantastic in supposing that the words 'mental imagery' really expressed what I believed everybody supposed them to

mean. [...] "On the other hand, when I spoke to persons whom I met in general society, I found an entirely different disposition to prevail. Many men, and yet a larger number of women, and many boys and girls, declared that they habitually saw mental imagery, and that it was very distinct to them and full of color". Commenting on individual differences, James concludes "There are imaginations, not 'the Imagination', and they must be studied in detail." James, again, commenting on Taine, points out that vague mental images are not abstract ideas, but rather a different kind of representation which is still visual rather than just an abstract idea: "a blurred picture is just as much a single mental fact as a sharp picture is; and the use of either picture by the mind to symbolize a whole class of individuals is a new mental function". Thus, the two key ideas of the sensory nature of imagery and of individual differences in imagery were already put forward in the era of introspective psychology.

After these early progresses, mental imagery was neglected by behaviorists. Spearheaded by John Watson, they rejected every mentalist approach, and, in consequence, avoided the topic of imagery. In "Psychology from the standpoint of a behaviorist" (1929), Watson alludes to imagery briefly when discussing after-images, and later in a footnote, stating that "the upholders of 'imagery' and 'imageless thought' should show the need of such 'processes' and demonstrate objectively their presence.". He goes on by suggesting that a conditioned reflex making use of centrifugal neural elements innerving the retina could account for a visual stimulus constituent in the absence of any light, and concludes that "From the standpoint of theory, this distinction between a *visual reaction* and a 'visual image' is important." Thus was mental imagery buried for half a century.

However, after forty-some years in the desert, mental imagery made a noisy come back to the main stage of nascent cognitive science, with a fiery debate between Steve Kosslyn and Zenon Pylyshyn. This debate was obscured by the simultaneous consideration of several issues, e.g.: "what is a mental image?", "what does it mean for a representation to 'depict', or be 'pictorial', as opposed to being 'descriptive' ", "are mental images functional or epiphenomenal?". "are the mental representations involved in imagery similar to those involved in perception ?", "what neural pathways are common, if any, to imagery and perception?"... Block (1981) in introduction of a collection

of essays, writes: "One side, the *pictorialist* side, takes the naive response to the experimental and introspective evidence to lead us in the right direction. The pictorialists agree that we don't literally have pictures in our brains, but they insist nevertheless that our mental images represent in roughly the way that pictures represent. The other side, the *descriptionalist* side, regards the introspective and experimental experience mentioned as misleading.. On their view, we should think of mental images as representing in the manner of some non-imagistic representation – namely, in the manner of language rather than pictures". Pictorialists, spearheaded by Stephen Kosslyn, included mostly psychologists, such as Allan Paivio, Ulrich Neisser, Ronald Finke, Roger Shephard, Lynn Cooper, and Martha Farah, but also for example philosophers like Jerry Fodor. Descriptionalists, lead by Zenon Pylyshyn, also included scientists of different horizons, such James Gibson and Daniel Dennet. Much like the interplay of muscles in the body, these two sides were antagonist and yet synergetic, because the constant nagging of the descriptionalists forced the pictorialists into better definitions and more controlled experiments.

After a decade and a half of controversy, the scale finally tipped towards the pictorialist side, mostly as a result of Kosslyn's unyielding courage at smiting craftier and craftier experiments in response to Pylyshyn's criticism, and also because of the growing evidence from neuropsychology, most of it brought forward by Martha Farah (1984; 1988). Although some rearguard skirmishes are still possible, the conclusion, we believe, belongs to Kosslyn's "Image and Brain, the resolution of the imagery debate" (1994). Kosslyn's book is mostly an analysis of vision, and argues that imagery is the use of the top-down mechanisms necessary for vision, in the absence of actual visual input.

Imagery in the brain

Nowadays, several issues remain unclear about imagery. The hottest question is probably the neuropsychological question of identifying the brain regions involved in imagery. A recent debate (July 1994) in "Trends in Neurosciences" opposed partisans and adversaries of the involvement of the occipital cortex. The evidence, at this date is still inconclusive, but one point that comes up is that, in assessing which regions of the brain are involved in mental imagery, great consideration should be made of the

individual differences between subjects. Quite possibly, different brain areas can be involved in different types of imagery.

The variability of imagery in its implementation by the brain, has been first addressed by Hebb (1968): "It is proposed (a) that eye movement has a specific function, (b) that first order cell assemblies are the basis of vivid specific imagery, and (c) that higher-order assemblies are the basis of less-specific imagery and non-representational conceptual processes." In more recent parlance, Hebb would suggest that similar top-down mechanisms may elicit the reconstruction of the mental representation in different brain regions; and that when these mechanisms are at work in occipital cortex, they create a vivid, pictorial mental image, but that when they are at work solely in down-stream areas of visual processing, they create non-pictorial representations, which for some subjects may constitute the only mental "imagery" available. We should remark how Hebb's concept is remarkably close to that advanced by James, and indeed provides a neural account which supports the idea that "a vague image is as much a mental fact as a sharp image".

Putative rôle of pictorial imagery

Now, this view of things, following closely Kosslyn and Hebb, is still vulnerable to a critique of the mental image best formulated by Gibson (1979). Gibson wonders what pictorial imagery could be good for: "An imaginary object can undergo an *imaginary* scrutiny, no doubt, but you are not going to discover a new and surprising feature of the object this way. For it is the very features of the object that your perceptual system has already picked up that constitute your ability to visualize it [...] Can you obtain new stimulation and extract new information from it?". Gibson intended to prove that a mental image cannot be scrutinized, but his critique also carries the idea that a mental image brings no new information. Surely this point is right if *all the mental image does is to represent what the top-down specifications require.*

However, Gibson's critique misses its target if imagery brings in information which is new *with respect to the non-pictorial representations* which were used to form the mental image. For example, regularities in the world are likely to be discarded by high level visual processing – all the more likely to be discarded as they are systematic

regularities – because they are of no help in identification. Universal regularities are not likely encoded in our memory traces, just like universal laws of nature are often not stated in descriptions of the world. Many tourists see Pisa, in Italy, and its famous tilted tower. Yet no one, in their description, says "Well, the place has a gravity of 9.81 meters per square second..."? That fact is so true it is irrelevant, because it is totally unspecific to the town. Yet, it is totally relevant if one gives details of Galileo's experiment on falling bodies. The uninformative fact becomes essential for the very reason it was uninformative: because it is true everywhere. Likewise, the property of continuity of objects – a tautological property, since we define physical objects, by default, as continuous solids. This property is irrelevant to abject representation. Many objects, being convex, can be specified by their apexes, forgetting the in-between matter. However, an apex-based representation is not optimal to find out what portion of space object occupy. Visualizing an object automatically makes use of the universal law of object continuity. Imagery therefore brings in information which is new with respect to the simple apex coordinates. Sure enough, that information is implicit in the apex coordinates; nevertheless it is not directly available for cognitive tasks; whereas it is directly available in a mental image. The information which is restored by imagery is not relevant for many a task, and therefore does not belong with more abstract representations.

So, the useful thing with imagery is that it recreates a fuller representation of the object under consideration. This is not only true of visual objects and scenes. It can be crucial, for example, of motor images. It is more than likely, as suggested by (Droulez & Berthoz, 1990), that one key reason why top athletes benefit from rehearsing mentally their motor images is that the timing obtained by mentally sequencing these images is close to the timing of executing the sequence of movements, much closer than, for example, the timing of sequencing symbolic representations of these motor goals. Matching to expectation might be another reason to have an expanded version of the representation of what is expected – we will come back to this in the discussion. Thus, there can be a utility to forming a mental image which makes explicit information otherwise implicit.

A mechanism for imagery: Attractor dynamics

This expansion of the symbolic representation into a mental image should require some processing. It is possible that the processing is made solely by the descending connectivity. However, a more economical design would be for each brain area involved to force the induced representation into well-formedness. Thereby, in vision, the necessary reduction of information would be made locally, not during the transit of information from one area to the next. Reciprocally, in imagery, top-down influence could restore approximate patterns, to which local processes would give the right "finish". So, this local mechanism would be useful for both imagery and perception. All it takes is that regions of representation enforce, by their proper dynamics, a certain consistency of representation.

We suggest that an attractor dynamics may be the agent of this reshaping of activity. By having as attractors the "well-formed" representations, each region can take care, by its own dynamics, of reshaping the incoming pattern of activity. For vision, this reshaping is an active process of simplification or classification. For imagery, it allows an approximate reconstitution to be improved. In particular, because bottom-up and top-down input both are submitted to the same procrustean bed, a pattern of activity not corresponding to the representation of something in the world would soon be altered to be made 'realistic', that is, identical to a pattern achieved during perception. Thus, the expansion of a mental image would be a process contributed to both by the descending connections and by the local connectivity of the brain region where the image is to be sustained.

Properties of fixed point networks

It is well-known that networks with lateral connectivity implements fixed point dynamics (Anderson, Silverstein, Ritz & Jones, 1977; Hopfield, 1982) when the connection weights are symmetrical. The final pattern, the fixed point, is a compromise between a constant pull towards the eigenvalues of the weight matrix and the limit that a non-linear transfer function (the thresholding) puts on neural activity. In the above models, this property is used in fully connected networks where the activity has relatively few available attractors, which can be individually impressed upon the network

by learning or by imposing a pre-computed weight matrix.

Another line of study, illustrated for example by Grossberg (1973) or Wilson and Cowan (1973), is to use such recurrent network with a lighter connectivity, usually restricted by some constraint of spatial extension. As the weight matrix gets sparser, it becomes degenerate, and it becomes possible to have many more attractors in the network; in fact individual eigenvalues are replaced by eigensubspaces and thus degeneracy increases the number of attractors by orders of magnitude. In that line of study, the connectivity is usually fixed, or learned from some simpler scheme, not attempting to capture individual states. Wilson and Cowan (1973) present an example of such models. Because it uses a non-symmetrical connectivity, this model presents a more complex dynamics. However, the important property in our view is that the resulting eigenstates (patches, waves,...) are families of eigenstates.

A model with similar properties is put forward by Somers, Nelson and Sur (1995). In Somers'model, the orientation selectivity of simple cells is sharpened by their lateral connections; in other terms, these cells, left to themselves, favor sharp orientation selection. Because that model, more detailed than ours, was tuned to avoid runaway excitation, considered as quasi-epileptic; hence the isolated cortical layer would not display non-zero attractors per se. However, the local dynamics strongly pulls any activity state into a state of one family: those with sharp orientation selection. In brief, the lateral connections provide the attractor dynamics which can be used to alter and form into a well-formed mental image the pattern of activity provided by the descending connections.

Conclusion on past models

Putting these elements together, we see that the most desirable aspects of attractor-based processing still need to be demonstrated in a relatively realistic model. Those desirable properties are bistability, the existence of families of attractors, the ability to use attractor dynamics to improve an activity pattern, and finally likeness to some area of visual cortex. (Somers, Nelson & Sur, 1995) shows with realistic details, that attractor dynamics can improve the computational properties of low-level cortical areas, but the model is not able to sustain activity patterns. Wilson and Cowan (1973)

show that lateral connections allow the sustain of activity patterns, but the model is a one-dimensional line, and the neural populations modeled do not have the physiology, function, or receptive field, of any visual neuron. Fixed points models such as those of Anderson, Silverstein, Ritz and Jones (1977) or of Hopfield (1982) have a very limited number of attractors because their weight matrix is not (or hardly) degenerate. Thus, none of these models present the full range of desirable properties.

III.A.2 Our model

Goals

In this work we have tried to present an attractor network combining all the properties deemed desirable from the theoretical viewpoint, and otherwise using minimal assumptions:

- Bistability, or hysteresis, is the first property for which we are looking. It should allow a cortical region to sustain an activity pattern when input stops.
- The existence of families of attractors rather than discrete fixed points, is necessary to allow a great flexibility of representation in low-level visual areas.
- A useful attractor dynamics, whereby activity patterns are improved by the convergence towards an attractor.
- Reasonable similitude between the neurons of the model and some subset of visual cortex.

Assumptions

The first core of assumptions made concern the representational space. The representation involved is a Gabor wavelet transform of an input 2D greyscale image. The wavelets used are varied systematically in position along an XY grid, in orientation, in phase (between sine and cosine), and in size, although they stay self-similar (the trigonometric and gaussian bandwidths stay in the same ratio). This is not to say that there is such a region of representation in the brain. There are neurons which perform orientation selection, and many appear to implement a Gabor-like filtering of

the visual field; moreover these neurons come in different bandwidths, but are roughly self-similar. However, two things about these neurons are unlikely: first, that their set forms a substantially independent dynamical system, and second, that they are varied systematically just like the neurons in our model are. In particular, the variations in neuronal receptive field size are related to retinal eccentricity rather than to a systematic encoding of different wavelengths. Despite these weaknesses, the bank of Gabor filters was chosen as the primary representation because it is a decent approximation to a subpopulation of the visual cortex and also is used extensively throughout the field of vision modeling.

The second core of assumptions regards the weights connecting these neurons. We have assumed that these weights would be those that arise from hebbian learning at the wavelet layer of the activity patterns arising from random activity in the greyscale representation region. If the greyscale representation takes many different random states, connections between two wavelet-encoding neurons will be strengthened, or weakened, according to the convolution of their receptive fields: thus, cells whose receptive fields are disjoint will have no connection, whereas cells with almost the same receptive field will grow strong excitatory connections, and cells with opposite responses spatially overlapping will, through the mediation of interneuron, develop inhibition. Hence, the weight connecting two Gabor units will be, to a network-wise multiplicative constant, the overlap of their signed receptive fields. As detailed in the discussion, it is likely that these weights are a very good approximation to the actual cortical connection strengths

Finally, several modeling assumptions follow established tradition: each of the wavelets-encoding units, which we will abusively call "a neuron", actually stands for a pool of neurons, and takes a continuous activation value from zero to one, taken to be an average spiking rate. The model is updated synchronously. We have allowed for units of our model to have both excitatory and inhibitory connections, which can be justified by including the local interneurons in the pool of modeled neurons that each modeling 'neuron' stands for.

Thus, the type of representation was chosen according to convention and using minimal assumptions, rather than tailored out of complexity to fit our expectancies. We hope that this honesty strengthens the argument that the properties observed shall

apply to many different models, provided that they include sufficient reafference. In particular the robustness of the model suggests that the properties observed could extend to biological tissue, with a somewhat stronger explanatory power than that of a mere metaphor.

Simulations

Given this model, we have studied in turn several properties of an isolated region: its reaction when it is submitted to brief bottom-up input, its ability to reconstitute a well-formed pattern from noisy top-down input, the dependence of these properties on the lateral connections, and how top-down and bottom-up input combine in this model.

III.B Methods

III.B.1 Architecture of the model: logical components

The essential component of our model is a layer of representation by wavelets. Each neuron in this layer represents the loading of the image onto a particular Gabor filter. Below this main layer, a pictorial representation (greyscale image) is used both for holding the input image, and as an intermediary stage of computation. Above this main layer is a layer for holding the 'seed' of top-down input.

The model is updated synchronously by cycles. In each cycle, each neuron of the main layer receives input from the greyscale image filtered by the Gabor filter associated with this neuron, plus lateral input from all other neurons in the main layer, plus top-down input from the seed layer. This total input is then reduced by a sigmoid-like transfer function, described in appendix 1.

Depending on the simulation, all or only part of the connections may be activated. For example, in the "short term memory" simulation, the main layer is standalone from the second cycle onwards. In fact, all the components of the model run together only in a minority of the simulations (the combination and control combination simulations).

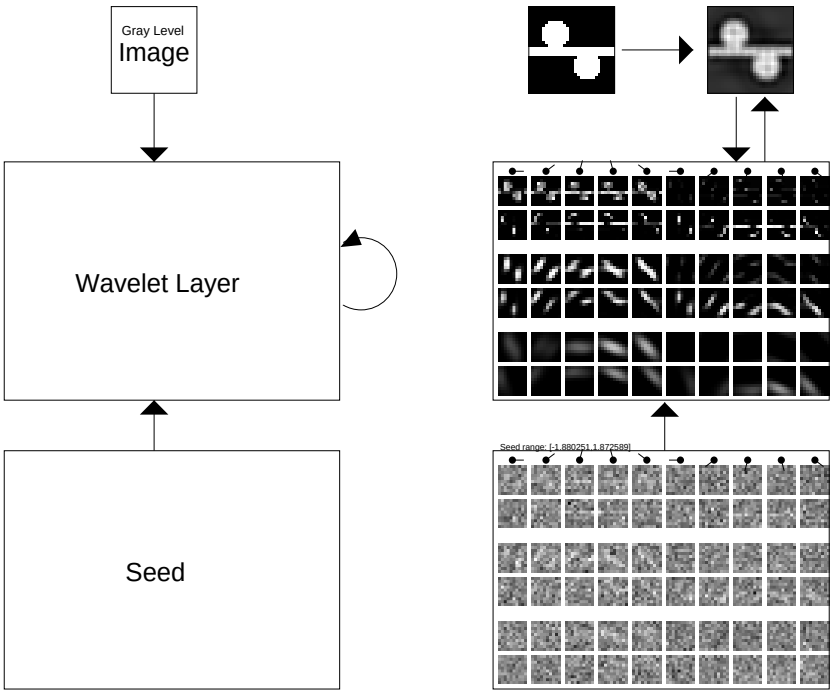

Figure III.1: **Architecture of the model:** The logical architecture (to the left) is implemented by a practical architecture (to the right) adding a buffer which facilitates computation and supervision. The input comes as a grey level image. Input and recirculating activity are added in the buffer (top rightmost image). The buffer is a grey-level representation of the same size as the input. Shown here at the end of a cycle, it presently contains only recirculating activity from the wavelet layer, giving an idea of wavelet activity. The blurring is normal, since the lowest bandwidth of the wavelet filters is 2.4 pixels. The main layer is the wavelet layer, where the sampling follows a 10x10 grid, computed at 3 increasing spatial scales with both sine and cosine phases, for ten different orientations. The "seed" for top-down input is made of the state that the wavelet layer gets from bottom-up input only, plus a gaussian noise, to insure that top-down input is not of equal quality as the bottom-up input. The noise added to the seed in this figure has a bandwidth of 0.4. Even that moderate level of noise destroys all apparent structure in the top-down input. Arrows indicate connections. Bottom-up input is obtained by turning on the Image to Buffer connections. The recurrent connections of the wavelet layer are implemented through the buffer (see text). Top-down input is obtained by turning on the connections from the seed layer to the main layer.

III.B.2 Architecture of the model: logical connections

The three logical components of the network are connected by three bundles of connections: a bottom-up bundle from the greyscale to the wavelet layers, a lateral bundle, within the wavelet layer, and a top-down bundle from the seed to the wavelet layer. For each bundle, the connection strength is modulated for all connections of the bundle by a bundle-specific parameter. Except the adjustment of the parameters of the model (see appendix 3), which include these overall connection strength and the size of the Gabor filters, all weights are fixed and were decided by deduction, not by experimentation or learning.

As said above, the greyscale image is connected to the wavelet layer by Gabor filters, shown in Figure III.2. The classical mechanism of filtering is illustrated in Figure III.4.

For the lateral connections of the main layer, we have chosen a "natural" weight: the overlap of the filters associated with the connected neurons. Thus, the weight of the connection between two neurons corresponds to the spatial correlation of the filters through which these two neurons get their bottom-up input. This weight is natural in the sense that it is the weight which would arise by hebbian learning from a equiprobable distribution of random images.

For the top-down connections, we use a one-to-one mapping between the seed and the wavelet layer. The seed layer retains a noised version of the activity originally induced by the target in the wavelet layer. Therefore, using the one-to-one mapping simply pushes the wavelet layer towards that noised version of its previous activity. In a previous model we used a more complicated data compression to associate each image with an separate attractor state of another network; that attractor was then used as a seed; in that case the top-down weights were learned.

III.B.3 Architecture of the model: actual implementation

Thus, in abstracto, the model has three layers (greyscale, wavelet, seed), and three types of connections: bottom-up input filters, recurrent connections of the main layer, and top-down connections to the main layer. In practice, a computational trick decreases considerably computational time at the expense of introducing a supplementary

layer of grey level representation, called “the buffer”.

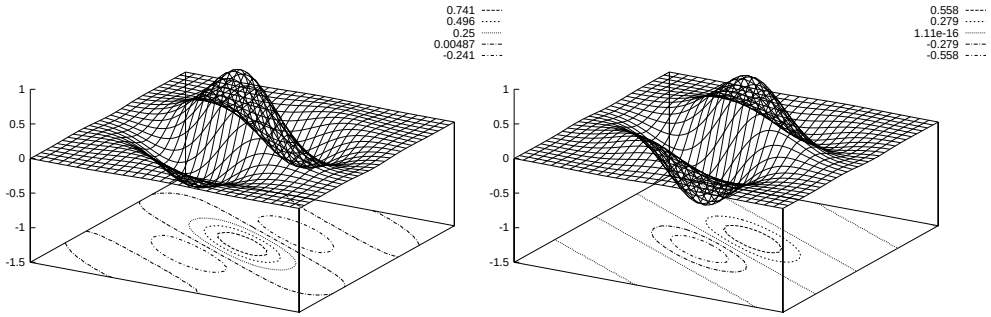

Figure III.2: **Profile of the Gabor wavelet functions used:** Gabor wavelets are made of a gaussian filter convolved with a sinusoidal filter. The wavelet layer used for the quantitative simulations is a bank of such Gabor filters, sampling a 10x10 grid for x and y location, ten different orientations, the two phases shown (sine and cosine), and three increasing spatial scales. Wavelets of different scales are similar, i.e., the ratio of their spatial bandwidth to the frequency bandwidth is the same. These filters are not only involved in the bottom-up input, but also in the recurrent connections: the weights connecting two units correspond to the overlap (the spatial correlation) of their filters.

The wavelets are varied across the following dimensions: x and y location, phase (sine or cosine), orientation, size (which regulates both spatial frequency and bandwidth). Figure III.2 shows the shape of the wavelets used. The wavelets of different sizes and orientation are similar. Since the cosine is an even function, half the orientations would be redundant, so we have used these units to encode for -cos(input). The quantitative results presented in this paper were obtained with a wavelet layer of 10x10 points, with 10 different orientations, and 3 sizes. Detailed parameters are given in appendix 3. Qualitatively identical result were obtained in a previous model where the wavelets had only one phase and are represented in Figure III.3.

Below the wavelet layer there are two pictorial representations (greyscale images). One is the image, and stays without change during each phase of simulation. The other is the buffer, used both for getting input from the input image, and as an

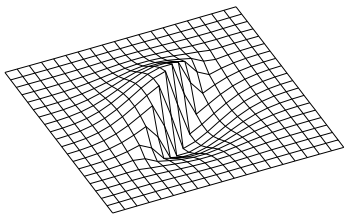

Figure III.3: **Filter used in the preliminary work:** In a pilot model, the wavelet was a difference of gaussians where the off-surround is small, with a sharp sign reversal across the orientation line. There was no "cosine" phase. Despite the greater simplicity, the qualitative results were identical.

intermediary stage of computation. Using the buffer reduces the computation of the wavelet-to-wavelet influence to a two-step computation, from wavelet layer to the buffer, and back. This is possible because the weights connecting two elements of the wavelet layer are equal to the overlap of their input filters. Thus, by expressing the activity of each wavelet layer neuron through the corresponding filter onto the greyscale buffer and then reloading that greyscale buffer into the wavelet layer, again through the filters, we feed each neuron the input that comes from all the neurons of the layer through a weight equal to the spatial correlation of the filters. Mathematically, whereas the logical equations are:

$$net_{i,t+1} = \Sigma_j activity_{j,t}.weight_{i,j} + \Sigma_k greyscale_{k,t+1}.Gabor_{i,k} + seed_{i,t+1} \qquad \text{(III.1)}$$

$$activity_{i,t+1} = TF(net_{i,t+1}) \qquad \text{(III.2)}$$

where

$$weight_{i,j} = \Sigma_k Gabor_{i,k}.Gabor_{k,j} \qquad \text{(III.3)}$$

the equations implemented are:

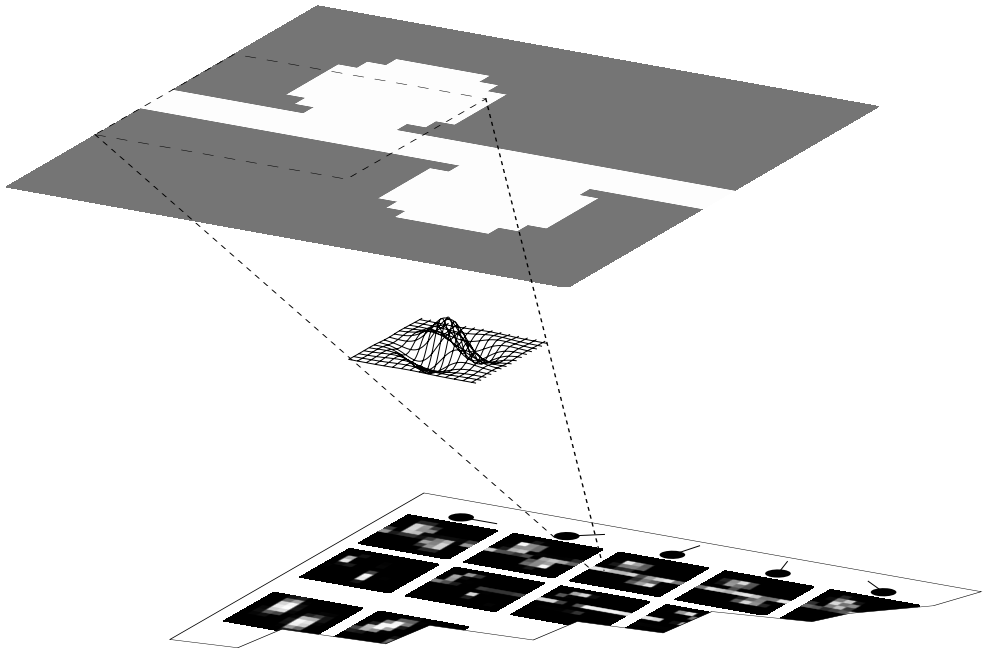

Figure III.4: **Mechanism of filtering:** To obtain the bottom input to location (X,Y), the filtering takes a window in the grey level image centered on (X,Y), multiplies pixel by pixel with the filter, and adds it all together to give the input value. For the (X,Y,orientation, small scale, cosine phase) illustrated here, the filter is almost aligned with a strong feature in the window, so that the bottom-up input to that wavelet neuron is high.

$$buffer_{k,t+1} = \Sigma_j activity_{j,t}.Gabor_{k,j} + greyscale_{k,t+1} \quad \text{(III.4)}$$

$$net_{i,t+1} = \Sigma_k buffer_{k,t+1}.Gabor_{i,k} + seed_{j,t+1} \quad \text{(III.5)}$$

$$activity_{i,t+1} = TF(net_{i,t+1}) \quad \text{(III.6)}$$

Since there are 6000 neurons in the wavelet layer and only 900 in the buffer, this method reduces computations due to bottom-up and lateral connections by a factor 3.3 for the computation results[1]. For the full-scale Mona-Lisa picture presented in Figure III.13, the factor increases to 8.5.

In addition to saving considerable computing time and weight storage space, this method provides us with a visible representation of what is going on in the wavelet layer. Especially, when the activity pattern is nearly stationary, this intermediary representation is almost the deconvolution of the wavelet transform. It would be the exact deconvolution under two conditions: first, that the pattern be stationary, and second, that there be no non-linear transform in the wavelet layer; since the patterns of best fit are still fairly in the linear range of the dynamics, and since their shape is decently stable, the visual representation obtained, for free, in the course of this computation, gives an acceptable idea of what is going on.

Above the wavelet layer, there is a layer for holding the 'seed' of top-down input. In this model, that upper layer is identical in dimensions to the wavelet layer; the seed of imagery is a noised version of the original wavelet pattern. (In a previous model, there were actually two layers operating a data compression and mapping the images to symbols arising from unsupervised learning).

Thus, in summary, the architecture of the model is that shown in Figure III.1. The input image stays without alteration in the first grey level structure, the "image". The second grey level structure, the "buffer", receives feedforward input from the image, and feedback input from the wavelet layer, through the Gabor filters. These two inputs are added pixel by pixel. Depending on the operating mode, the relative importance of these two inputs is changed: for example, for testing the static memory, the buffer is initialized with the image, but receives no feedforward henceforth. The buffer outputs

[1] to 2x900x6000+900 instead of 6000x6000+900x6000.

through the Gabor filters to the wavelet layer. The basic loop wavelet–buffer–wavelet is the core of the model, as it implements a wavelet representation with lateral connections. The image–buffer input is the feedforward input. When simulating imagery, the projections seed–wavelet are turned on, and implement the descending information reaching the wavelet layer.

III.B.4 Computation and averaging of the results

What measure of goodness?

Analyzing the behavior of this model raised the timeless question: "in the absence of a complete stream of processing, how can one evaluate the quality of the representations obtained in a model?". Because we only consider the behavior of a low-level stage of representation, assessing the quality of a sustained pattern, or of a reconstructed representation, is a difficult task. Fortunately, several constraints dictated by common sense limit the arbitrariness of the choice of a measure of goodness.

First, for want of further processing, the goodness can only be obtained by a comparison of some kind with a target. If there were further processing, the goodness could be estimated by comparing the behavior of the whole system to a target behavior, but this is not the case in this model. Also fortunately, the target presents itself naturally: whether the task is to sustain the representation of an image, or to recreate it, this image should be the one that the target is about. Because the main layer of the model is the wavelet layer, the comparison should be done in the wavelet domain. For these reasons, the target which was chosen is the representation obtained by pure feedforward input of the image. One could question why this should be the ideal target, since in a functioning system, the basis for behavior should rather be the representation obtained from perception aided by imagery. However, if we chose the latter as a target, we would unduly favor the very hypothesis that we want to demonstrate; and our point would become all the weaker as our figures would get stronger. Hence the choice of the pure feedforward representation as a target.

Second, the choice of the comparison method can be dictated by Occam's razor. The global level of activity increases through an execution run; for that reason, a measure of fit must be based on the pattern shape, not on its activity level. Also, it is desirable

that a random pattern, whose neurons independently take on random activity values, give statistically a measure of neutral goodness, being neither akin nor opposed to the target it is compared to. Likewise, uniform activity, arising from non-specific arousal, should be of neutral goodness when compared to any non-uniform pattern. From these remarks, it comes that the simplest available comparison is to compute the spatial correlation between normalized versions of pattern and target. If the activity of all neurons is zero, the goodness is conventionally zero, else it is:

$$goodness = \frac{\Sigma_i (S_i - \frac{\Sigma S_i}{N})(T_i - \frac{\Sigma T_i}{N})}{\sqrt{\Sigma_i (S_i - \frac{\Sigma S_i}{N})^2}\sqrt{\Sigma_i (T_i - \frac{\Sigma T_i}{N})^2}} \qquad \text{(III.7)}$$

Where S_i is the present state, neuron by neuron, of the wavelet layer, T_i the state of the same neuron in the reference target, and N is the total number of neurons in the wavelet layer

It can be verified that if $S_i = K.T_i$ where K is a constant, we will have $goodness = 1$, and that the measure does not vary when either the S_i's or the T_i's are multiplied by a non-zero constant (thus scaling the whole pattern). Patterns inversing the target do correspond to a goodness of -1. However, it must be pointed out that inverse patterns at the wavelet level correspond to nothing at the grey level. Inverse grey level patterns create a symmetry between the wavelet input for half the units only (sinus-phased Gabors), and then after the non-linear transform, this change of sign corresponds to nothing any longer.

Software techniques and details on simulation runs

The results were obtained on Unix platforms. The programs were written in C and ran on three different platforms: a sparc station running SunOS 4.1, a sparc station running SOLARIS, and a PC running linux. With the adaptations necessary to include equivalent libraries under each system, the code was identical on all three. Calls to random functions were made when necessary (essentially to noise the seed) by calls to the "ran1" routine described in Press, Teukolsky, Vetterling and Flannery (1992). The runtimes produced identical results on all three machines. A tar'ed version of the source and result files is publicly available by anonymous ftp at cogsci.ucsd.edu under the file name "/var/spool/ftp/pub/rebotier/imagery.tar.Z".

Two runtimes were made for the simulator. The first included on-line redisplay of the network as computation advanced, and interactive menu commands. The second was exclusively for batch use. Both runtimes admitted as parameters a file name for the network parameters, and a file name for a list of commands (optional for the interactive version). The commands, identical to the menu commands, allow amongst other things to change parameters on the fly, to load or save images or wavelet representations, to configure the network to activate or desactivate some parts of it, to store a reference target, and to store, noise and resize a seed of imagery, and to switch on or off a "spy mode" where the network reports once per cycle the goodness of fit to the reference and the mean activity level.

Each Figure in the result section was obtained from automatically processing intermediary files with tools like "sed', and "awk". These intermediary files were produced by running script files which automatically generate the right parameter and command files for the simulator. All these files are available in the public .tar.

Figure III.5: **Images used in the simulations:** Simulations used the first eight or all nine images (details in text).

Images used

Most computations were done on eight simple sketches made of lines segments and circles. In addition, a picture of Mona Lisa was used to give an idea of what is

represented in the case of a more complex image. That picture was sometimes used in conjunctions with the other eight in the averaged results, and sometimes not – individual results specify whether eight or nine images were used. The images can be seen in Figure III.5.

Imagery seed

To demonstrate the pattern-completion ability of the network, the top-down input, or *imagery seed*, had to be of lower quality than the bottom-up input. There are several ways to degrade the top-down input, which is encoded in the wavelet space. In the pilot model, we used linear compression of the wavelet state, thus reducing the amount of detail that could be specified by the top-down input. In this model, we used the more simple scheme of adding gaussian noise to the reference wavelet state. As explained above, the reference wavelet state is the state obtained from strict feedforward operation of the model. Gaussian noise of zero mean, with varying bandwidth (from 0.2 to 1.0) was added to that state to ensure that top-down input has a lower quality than input from the image itself. As seen is Figure III.1, even a moderate level of noise (0.4) can erase all obvious structure in the wavelet space. The level of disturbance is also illustrated by the goodness of fit between the noised pattern and the reference state, which is usually the first point in the figures showing the evolution of the wavelet state in *imagery* mode. The goodness of fit goes from .85 (for a 0.2 noise level) to 0.3 (for a 1.0 noise level). This drastic reduction in top-down input quality shows that the simple gaussian noising scheme is efficient.

Obtaining short term memory results

The short term memory results were obtained for each image by imprinting the image on the buffer at the beginning of the first cycle and then having the network run in a disconnected, wavelet-only mode.

Obtaining imagery results

The pattern reconstitution results are averaged over ten runs for each image, with a different random seed for each run, to ensure that the particular noise pattern

inflicted on the seed is not responsible for the effects observed. During each run, the descending input from the seed to the wavelet layer is left active throughout the run.

Obtaining threshold results

For each image, the goodness and activity levels are measured for different bottom-up input strengths, after eleven cycles during which the bottom-up input is left on permanently. This gives the bottom-up (perception) threshold presented in Figure III.8. The operation is then repeated with top-down input only, left on permanently, at different strengths, to give the results presented in Figure III.9.

Obtaining gain dependence results

The threshold testing is repeated with different strengths of lateral connections, both for the bottom-up and top-down situations. Based on the previous results and on these, a value of activity was chosen to make the discrimination between sub-threshold and supra-threshold final activity; The lowest value of input strength which resulted in supra-threshold final activity was automatically selected as the threshold associated with the current value of lateral connection strength. These thresholds as a function of lateral connection strength were averaged over the nine images and are presented in Figure III.10.

Obtaining perception/imagery synergy results

The detection of perceptual threshold was repeated for each image at various levels of additional top-down input representing the same image. The resulting activity, averaged over the nine images, and presented as a function of the two variables [bottom-up input strength, top-down input strength], is presented in Figure III.11. The same operation was repeated with nine pairs of different images: (ima,imd), (ima,ime), (ima,imf), (imb,imd), (imb,ime), (imb,imf), (imc,imd), (imc,ime), (imc.imf), where the first image was input from the bottom up, and the second from the top down. This control condition used the same input strengths as the initial condition, and its results, averaged over the nine pairs, are presented in Figure III.12.

III.C Results

III.C.1 Short term memory by reverberation

When it is disconnected from any external input –be it bottom-up or top-down– the wavelet layer reverberates activity in such a way that substantial alterations of the shape of the activity pattern take several cycles to occur.

Figure III.6 shows how perceptual input can be sustained in the wavelet layer. To give an idea of variability, we have plotted individual results for all nine images used in the simulations. In this experiment, the perceptual (bottom-up) input has been let through only during the first cycle. In the following cycles, as activity in the wavelet layer evolves without external input, we observe the correlation with the initial state (top graph) and the global activity (bottom graph). Because the gain of the feedback is greater than one, activity slowly increases up to a maximum value. Simultaneously, the shape of the activity pattern changes slowly and drifts away from its initial structure. One can also make note of how the final state can be very close or fairly far from the initial image, depending on the image – the range is from 0.17 to 0.69. Here, the "Mona Lisa" image ends up with a 0.32 correlation. One image also lingers over 20 cycles into an almost-attractor state of high quality, before escaping that basin and falling into a much worse representation. Compared to a one-dimensional attractor dynamics, we observe much more variability in the dynamics. Also, the speed of convergence does not necessarily depend on how close the final attractor is according to the correlation metrics; this is counter-intuitive since we would expect that a closer final state would be reached faster; intuition here, based on the visualization of 3-D "hill-climbing", forgets the complexities due to the non-linear transfer function. In summary, Figure III.6 illustrates that reverberation through lateral connections cause a sustain of pattern shape, and that all patterns are not equals for that sustain.

To carry on with the classic "hill-descent" analogy, one could say that fairly close to every interesting starting state, there is an attractor close enough that during the descent the activity pattern keeps for several cycles a strong resemblance to its initial shape. In terms of correlation between the initial pattern and the Nth pattern, the final

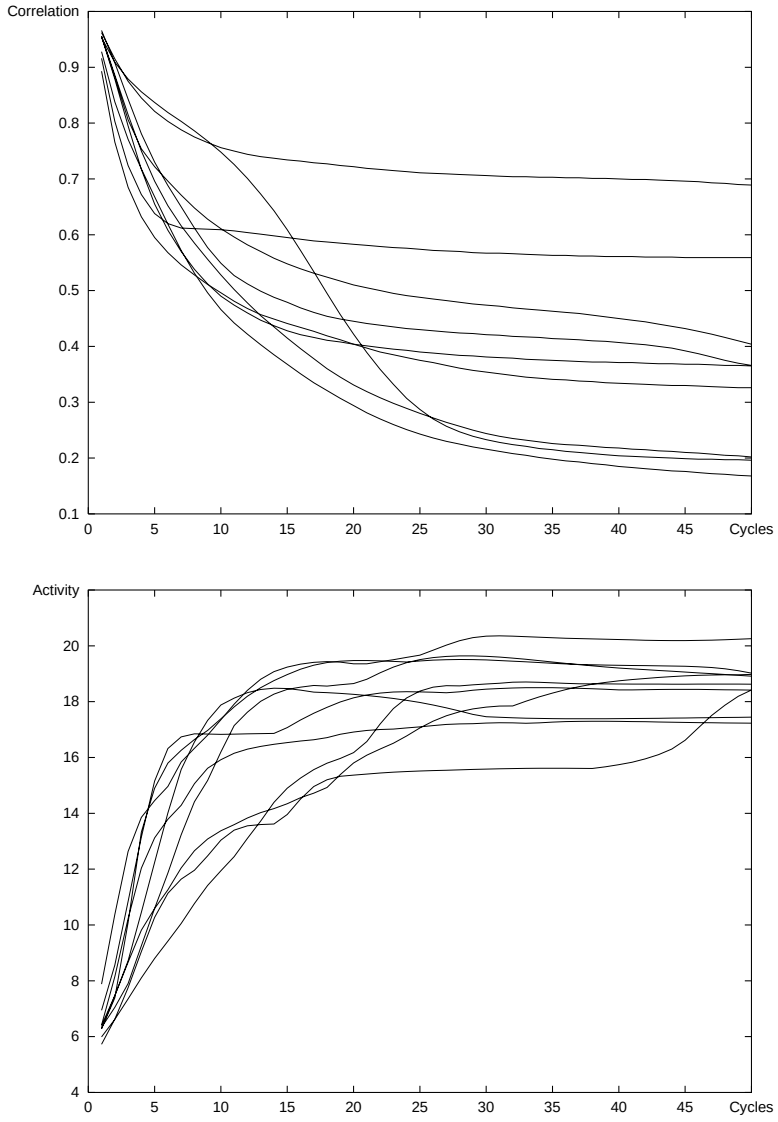

Figure III.6: **Static memory:** When the image (bottom-up) input is given only once, with sufficient strength, activity keeps increasing by recirculation (bottom) and the goodness of fit stays decent for many cycles (top). Each line represents a different image in the image set.

correlation averages 0.4 over the nine images used, and during the first nine cycles it did not go below 0.5 for any image.

III.C.2 Imagery by pattern reconstruction

Activity recirculation also develops and sustains activity patterns which have a top-down origin ("mental images"). Over the first few cycles, the reconstruction actually improves the pattern, because the connections favor patterns representing images with aligned edges and solid shapes. This simulation illustrates our theory of what goes on during imagery.

In Figure III.7, we see the evolution of the activity pattern induced by top-down input for different qualities of input. The input is made by restoring the reference state to which is added random noise. Each line is an average for the same noise level across the first eight images used previously, and across ten different random noise drawings. The noise level corresponds to the bandwidth the gaussian distribution followed by random noise added to the activity of each neuron. In contrast with the previous simulation, the input is left on throughout the simulation; this is to make up for an overall weaker input, which would be sub-threshold if it only lasted one cycle.

Looking at the evolution of correlation (top graph) in Figure III.7, one sees that the general behavior is an increase in correlation over the first few cycles, followed by a decrease which plateaus at relatively good correlation values. Except for the zero noise level (where the initial fit is near-perfect because the input is a weakened but unaltered reference pattern) , there is a strong improvement of the quality of the fit, for most curves an increment of correlation between 0.2 and 0.3. It can also be seen that the final plateau depends on the noise level. Two causes concur to this variation: first, since the input is left on, there is ongoing influence to the state of the layer; second, that with low grade input, different final attractors may be reached. Indeed, a look at the activity curves show that hi-noise trajectories show overall higher activity, including at the plateau. Because the noise is of mean zero, this must be explained by having the dynamics trapped in deeper attractors: as the random noise makes the input less specific, the activity is more likely to be captured in different attraction basins, and since in general deeper basins are also wider, activity ends up there more easily than in weaker

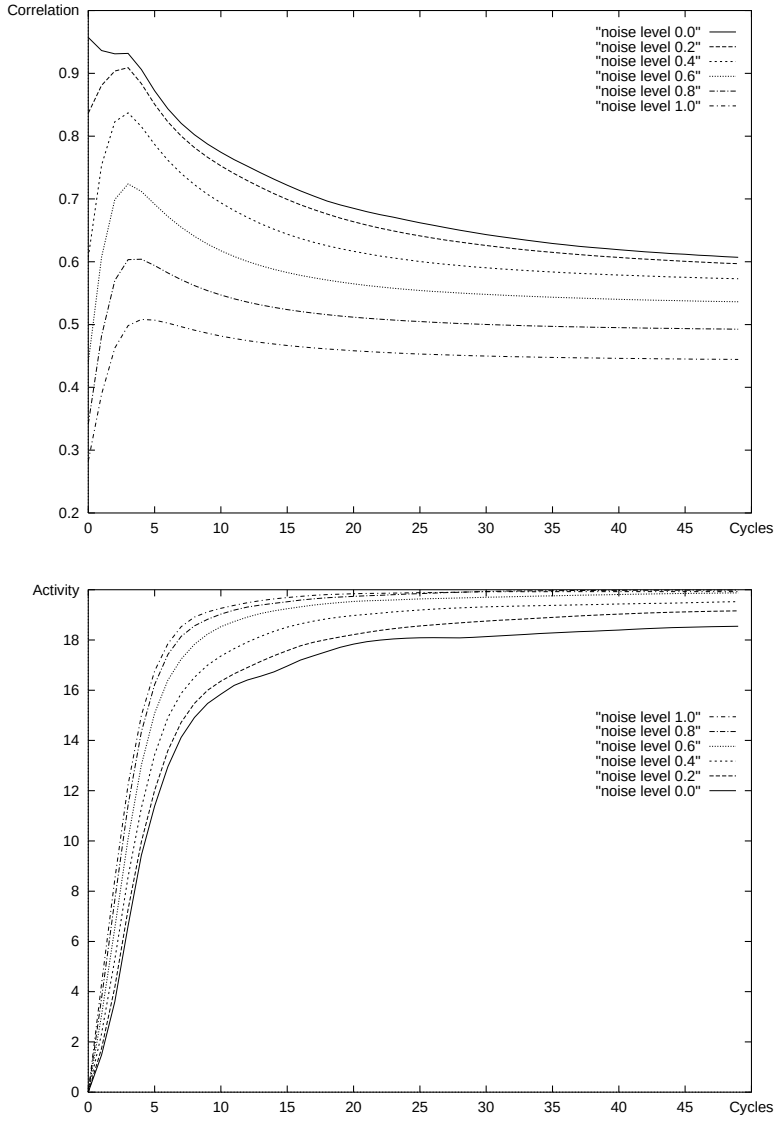

Figure III.7: **Imagery:** With continuous degraded top-down input of sufficient strength, activity increases by recirculation (bottom), and the shape of the reconstructed pattern improves over the first few cycles (top). Different lines show the simulation with different levels of seed degradation (bandwidth of gaussian noise from 0.0 to 1.0). Note how smaller noise levels allow to reach finer attractors which have a higher fit but a lower activity level. These curves are averaged across the eight line drawing images; individual curves show deviations similar to those in Figure 6.

basins.

Why does the dynamics improve rather than just preserve the pattern shape? We suggest that the top-down input no longer has to obey constraints inherent to the bottom-up induced representation. By reintroducing these constraints, the reverberation of activity recreates a better representation. Bottom-up input usually impresses a picture with many continuous edges and surfaces. Moreover, even from a random dot image, the overlap between the adjacent filters which connect the grey level to the wavelet layer will create ex nihilo a correlation structure. In contrast with this, the top-down input has, a priori, free structure. In the present model, it is made by adding random noise (activity determined randomly independently for each neuron) to the reference state; the structure is thus identical than that of perceptual input, but weaker because the noise levels it out. In more complex models, or, should the mechanism studied apply to the brain, in vivo, it may have any structure that arises from the structure of higher level representations. Thus, the top-down input is related to its target representation, but not alike. It is a seed, rather than a clone. During the relaxation phase, the seed grows into a representation which tries to align edges and provide continuity to the represented image. Since the original image is made of aligned edges and continuous surfaces, the representation obtained by the relaxation improves over the seed.

III.C.3 Non-linear transfer functions result in threshold implementation

Just as a single unit with a hi-gain sigmoid-like transfer function exhibits a bi-stable behavior when connected to itself with high enough a gain, so does a whole layer of units with lateral connections.

In Figure III.8 we see the correlation and global activity after five cycles of bottom-up stimulation followed by five cycles in isolation, as a function of the strength of the input during that first cycle. Clearly there is a threshold below which no activity is triggered, and above which activity goes onto a standard course of growth. As can be seen, this threshold varies with the image, in our stimulation from 0.8 to 1.2.

In Figure III.9 the same phenomenon is shown with top-down input. Here the input is left on throughout the simulation, which softens the activity curve. The

threshold mechanism is nevertheless obvious.

The correlation curves are more difficult to interpret. For below-threshold input, the activity decreases to zero, thus making the correlation a irrelevant measure. Fairly above threshold all input strengths are equivalent, unless they are so strong as to directly drive the dynamics in its ceiling range. Around the threshold, the correlation of the patterns which develops is not as good, presumably because the basins of the most representative attractors are fairly small, and therefore the weak input locates the dynamics from the beginning in the basin another, stronger attractor.

III.C.4 Dependance of threshold and imagery on recurrence gain

In a one-unit system with a sigmoid-like transfer function, the presence and the value of a threshold depend on the gain of the recurrent connection. Specifically, for the transfer function and parameters which we use, the threshold of a one-unit system booted with a one-shot input should appear for a feed-back gain of about 2.56 (see derivation in appendix), and for a gain above 3.0, it should obey a hyperbolic function. Is this dependence also the case for our many-units layer, where moreover the recurrent connections innervate mostly other neurons of the same layer?

Figure III.10 shows that the threshold in our model behaves almost like that of a one-unit system but with a quantitative adjustment for imagery. The top graph shows the average threshold for bottom-up input (perception) as a function of the gain. The dependence observed is very close to the theoretical dependence, also plotted. In practice, rather than appearing at gains of 2.56 the threshold does not appear for gains up to 3.0. The bottom graph shows the average threshold for top-down input, as a function of the gain. For imagery, the picture is slightly different. The theoretical threshold for a closed system with ongoing input should be $1.25/g$. The fit shown on the graph is $3.0/g$. Likely, the threshold is higher because of the bad resemblance between the top-down input (noised with a bandwidth of 0.5) and the attractor into which activity can increase.

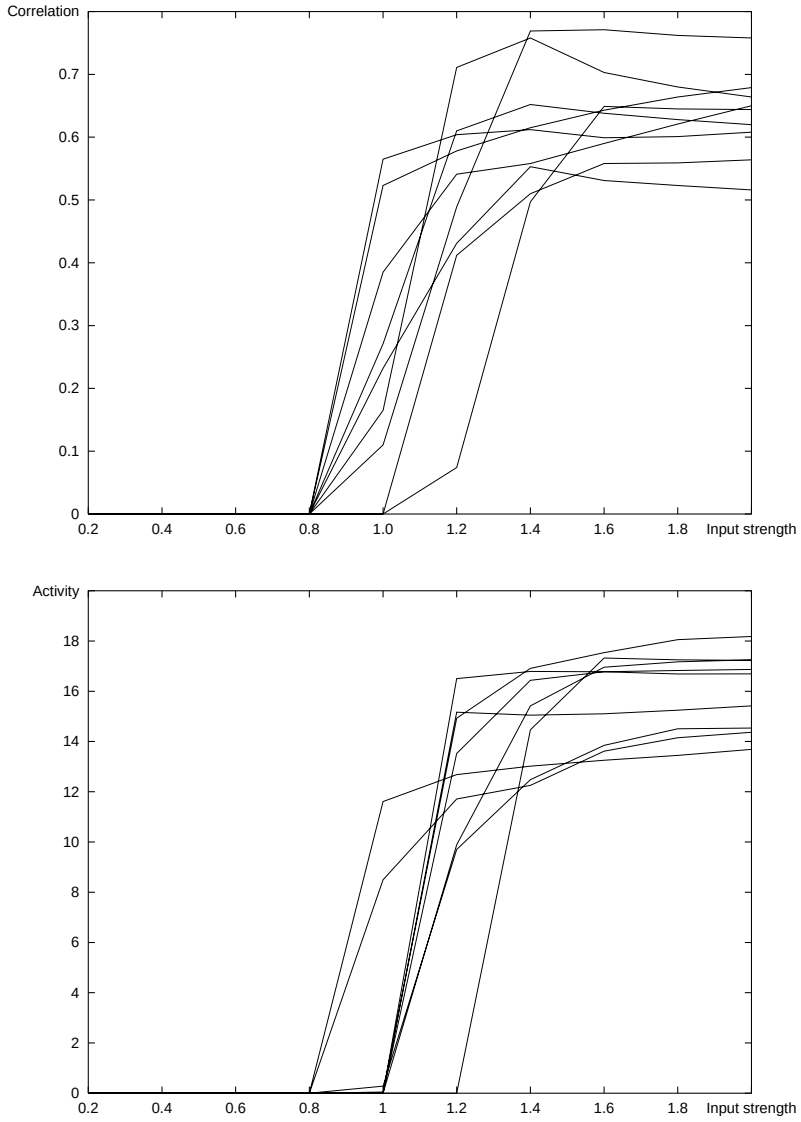

Figure III.8: **Perceptual threshold:** correlation (top) and activity (bottom) after the eleventh cycle give a snapshot of the direction the dynamic took: sustaining activity or letting it decay. Here, these are plotted as a function of the strength of the input, for each of the nine images. One sees that there exist a minimal strength, or perceptual threshold, below which activity decays. Note how the threshold depends which image is seen. Note also that near-threshold input strength yields a final state of lower quality (correlation), presumably because it puts the activity for grabs by the strongest attractor rather than by the attractor closest to the reference state.

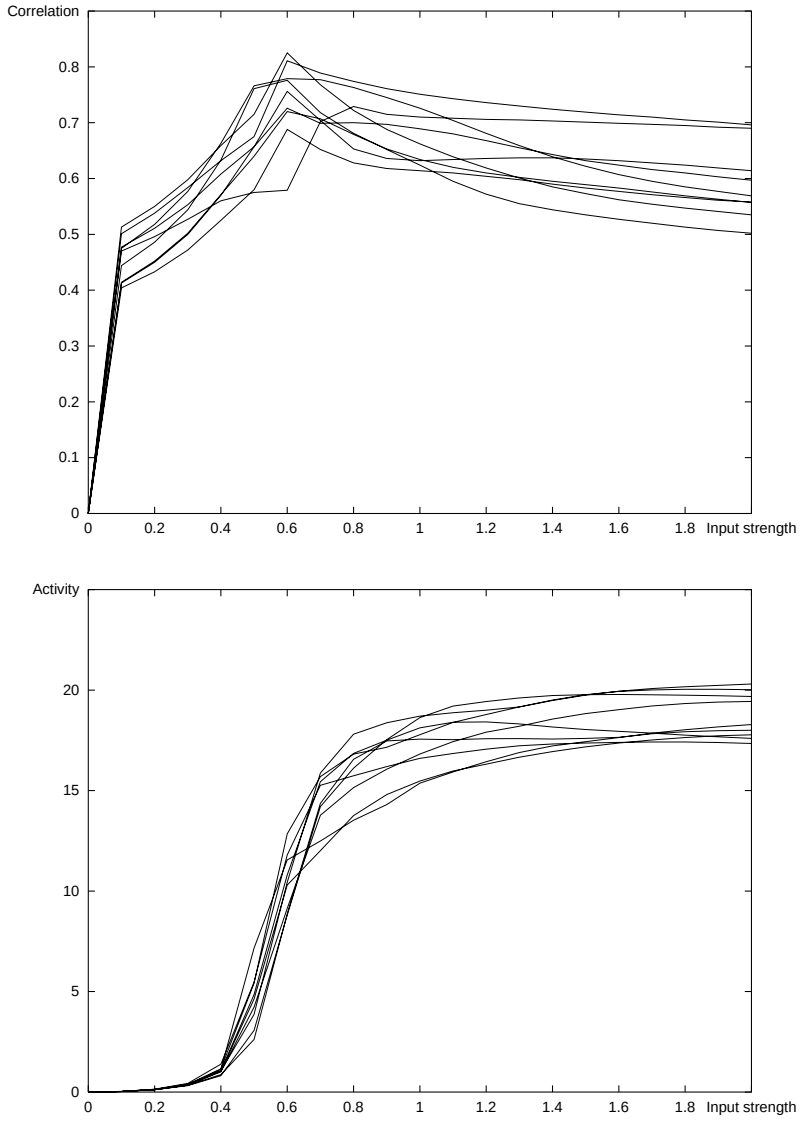

Figure III.9: **Imagery Threshold:** correlation (top) and activity (bottom) after ten cycles of continuous top-down input (imagery) as a function of the input strength, for each of the nine images. As in perception, there is an input strength below which activity does not increase. Again, the threshold depends on the image. The threshold does not appear as sharp because input is left on during on ten cycles rather than shut off after the first five cycles.

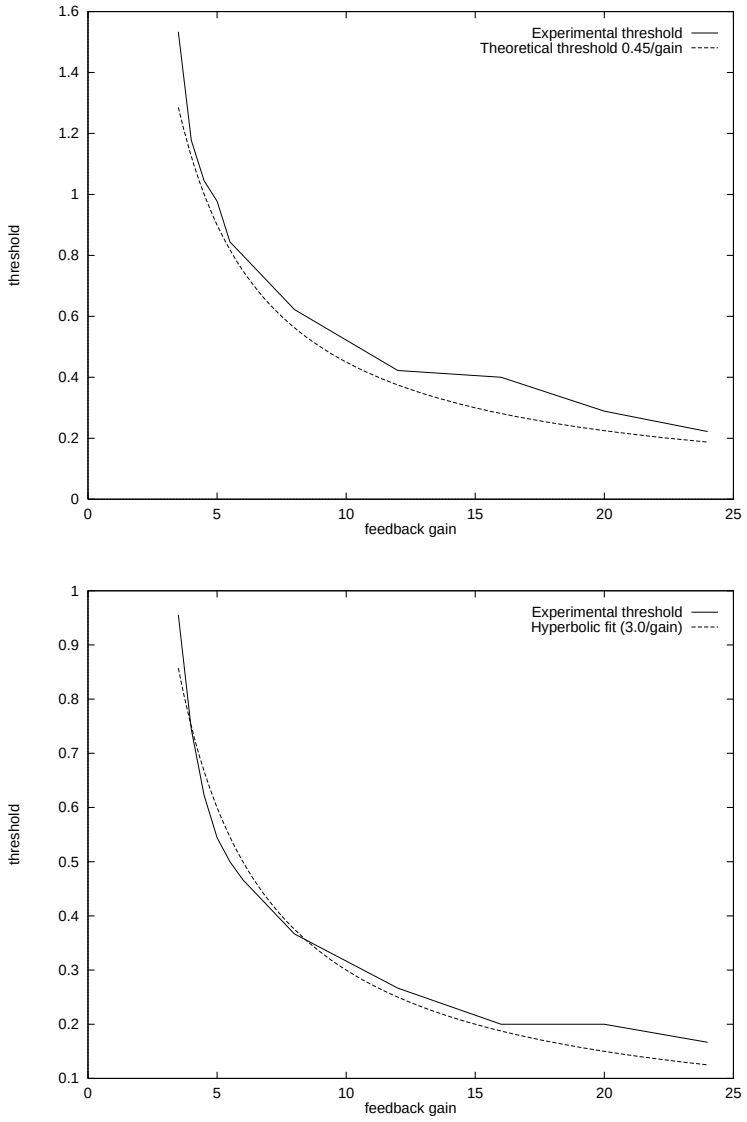

Figure III.10: **Thresholds decrease with lateral connection strength:** The average threshold at which activity starts being amplified by reverberation is given as a function of the gain of the feedback, for both perception (top) and imagery (bottom) inputs. The theoretical threshold for a one-unit closed system with the same transfer function is $4.5/gain$. That curve is drawn as such on the top graph. For the bottom graph, the shape and starting point of the hyperbolic fit are the same, but the thresholds are lower (the fit is 3.0/x), but not up to par with theory (1.3/x), reflecting that the ongoing noisy top-down input is not as efficient as bottom-up input.

III.C.5 Synergy between perception and imagery

As it has appeared more and more through this paper, top-down and bottom-up input play the same role in the dynamics of the middle layer. We are not surprised, therefore, to see that they can be combined, so that "perception" and "imagery" prime each other.

Figure III.11 illustrates how top-down and bottom-up inputs can be added to overcome the threshold. In these simulations, top-down and bottom-up inputs represent the same image. The top-down input is created with a noise level of 0.5 (see above, Figure III.7). Correlation (top Figure) and activity (bottom Figure) are plotted after five cycles of combined stimulation followed by five cycles in isolation. This Figure presents an average over the nine images. One can see that the activity depends on a perfect linear combination of both inputs.

Figure III.12 provides an essential control for the previous result. In these simulations, top-down and bottom-up inputs represent a different image. The pairs were generated by using images A,B and C paired in very possible way with images D, E and F. All other conditions are identical to those of the previous simulation. As can be seen, in Figure III.12, the top-down input gives a much weaker facilitation. Also, the image obtained does not have a good fit with either the image being imaged (as represented in the top graph), nor with the image perceived (not graphed).

In conclusion, imagery does prime perception, but this priming is substantially shape-specific.

III.C.6 Samples of imaged pictures

As an appetizer for curiosity, we have included two Figures presenting imagery runs. The images are obtained directly from the buffer; the grey level scale has been normalized to allow for the best visualization of the details and of the pattern shape. One must remember, however, that the first and second cycles have a lower global activity, so that the phenomenal impression would be to actually see these images simultaneously forming out of thin air and sharpening their features.

The first Figure (Figure III.13) presents an imagery run over a natural-looking image: a clipping from Leonardo da Vinci's painting, "La Joconde". For that run, the

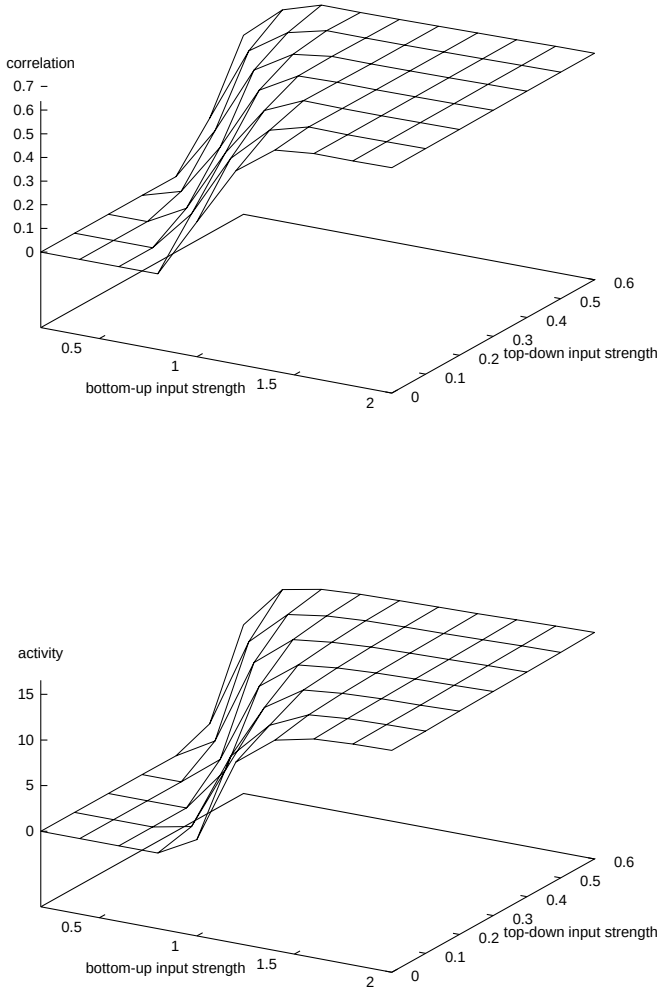

Figure III.11: **Imagery and Perception of the same image combine:** Correlation (top) and activity (bottom) after the eleventh cycle, as a function of both input strengths, in a dual input situation (perception), for the nine images used. Note how the threshold depends basically on the sum of both input strengths.

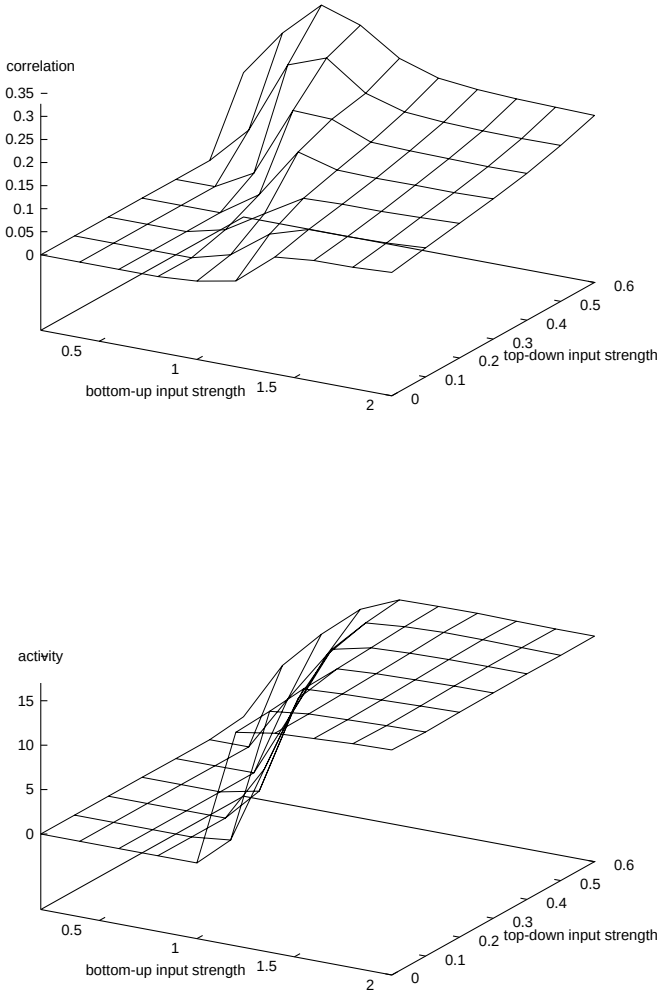

Figure III.12: **Imagery and perception of different images do not combine:** This control condition shows that imagery-induced "priming" is image-specific. Correlation (top) and activity (bottom) after the eleventh cycle, are plotted as a function of both input strengths, but contrary to the previous Figure, the images encoded by top-down and bottom-up input are different. Results are averaged over nine pairs. Compare with Figure III.11 and note that here the top-down input does not facilitate activation as much. The threshold is much less changed by top-down input than in Figure III.11.

simulator was recompiled with a bigger input (120x120) and more wavelet units (40x40, 14 orientations, 4 scales). The top line presents the input image and the reference image surrounding the graph plotting the goodness in function of the cycle number. The reference image looks like a low-pass filtered version of the original, which is not really surprising since the sharpest wavelets have a bandpass of 3.3 pixels. The reference pattern (upstream from the reference image shown in upper right corner) is noised by a gaussian noise of standard deviation 1 and of amplitude 0.4, then it is pruned by setting to zero 40randomly selected. That seed is projected continuously, but at low strength (just above threshold) into the wavelet layer, and the evolution of the wavelet layer is mostly driven by the lateral connections. The second, then third, lines of the Figure, show the state of the buffer during the first eight cycles of that imagery run. As one can see, random disparities disappear, but most features resist the diffusion until after the fourth image. It is also interesting to remark how some features, like the eyes, generate a self-sustaining sub-pattern.

The second Figure is a run on the line-and-circle images. The first column shows the projection of the reference state. The original images are similar, but presenting only sharp contours. The following columns show the first six steps of an "imagery" run on each image. The best fit to the reference state is usually the fourth or fifth image.

III.D Discussion

III.D.1 Summary of results

In this model, the central piece is a layer of semi-linear neurons with lateral connections. The bottom-up input is made from a greyscale image through Gabor filters, wherein xy-location, orientation, phase and size are varied systematically. The lateral connections are such that the weight of the connection between two neurons equals the overlap (the spatial convolution) of their input receptive fields. Top-down input is made on a one-to-one basis using noised versions of the target states. The transfer function is sigmoid-like, and the network is updated synchronously.

Left to its own device, with an activity pattern initially set from bottom-up input, the main layer keeps a decent representation of the initial target for several cycles,

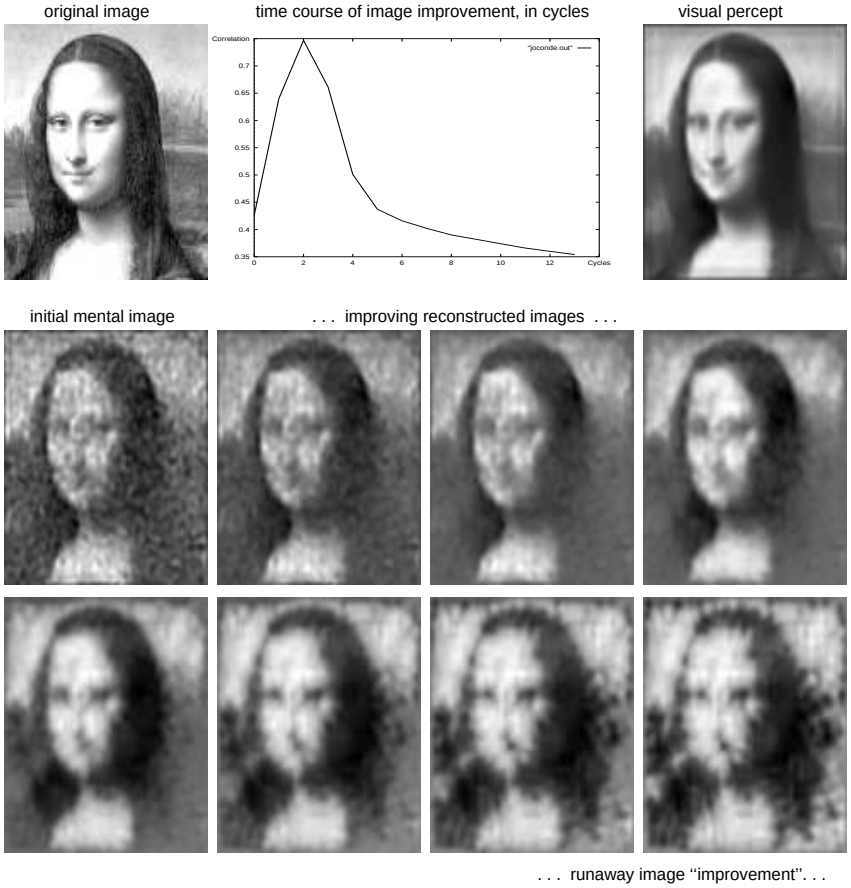

Figure III.13: **Imaging the Mona Lisa:** A clipping from Da Vinci's *La Joconde* (top left) is encoded in the wavelet layer, loosing only high frequency features. The top right picture is the corresponding mental image, the "percept". At that point the perceptual input is removed, the wavelet layer is noised , and this noised version is used as the seed for an imagery simulation. The second and last rows of pictures present the mental image as the reconstruction goes on. The initial image is fairly noisy, but imagery erases that noise without loosing the sharpness of the features, as simple diffusion would do. Ultimately, the image degrades: features of sufficient size acquire a "critical mass" which prevents their disappearing, as for example the eyes of the Mona Lisa. The top centered graph plots the goodness of fit (correlation). All images, except the top left, are projections (in the buffer), in 2-D of the 4-D wavelet representation. For a good contrast, the grey level scale has been fitted image per image; without this fitting, the effect is even more striking, as the growth of activity accompanying the improvement of shape makes Mona Lisa appear out of thin air, like a renaissance Cheshire cat.

Figure III.14: **Imaging the line images:** We represent here one imaging run for each image. In each line, the first image is a projection of the reference pattern, and the next images are projections of the wavelet layer state through the first six cycles of an imagery run.

thus implementing a transitory static memory, akin to an after-image. Bootstrapping this phenomenon requires the initial activity to be above a threshold which decreases with a higher gain of the lateral connections.

When submitted to top-down input consisting of a well-formed target state noised by gaussian noise, the main layer improves the shape of this noised input pattern, again over several cycles. This pattern completion of the shape goes along with of an increase of global activity. Thus, a representation of consistent shape and intensity arises from weak, noised input. Again, this phenomenon only happens if the input is above a threshold which is all the lower as the gain of lateral connections is high.

Finally, when sub-threshold top-down and bottom-up inputs are joined, they can together elicit a strong, well-shaped representation. However, this effect is almost non-existent if both sources of input disagree as to the target state.

III.D.2 Scope and limits of this model

In this model, the same mechanism can account for the analog of a cortical, positive afterimage, for the generation of a mental image, for the existence of a perceptive threshold, and for facilitation of imagery by vision, and $vice - versa$. The key property responsible for these phenomena is the feedback within a bank of non-linear units. As it is very likely that most cortical areas, especially visual, possess this property, the emergence within the cortex of the behaviors presented by this model depends on the quantities, rather than detailed qualities, of lateral connections.

However, it is worth reviewing the choices made by the modeler, comparing them with known details of cortical anatomy and physiology, to either point out the righteousness of the modeling choices, or anticipate the changes that would arise in a more complete model. In particular, two points need to be addressed here: encoding in a bank of Gabor Filters, and lateral connectivity.

Receptive fields of V1

Since Hubel and Wiesel (1962) have identified simple, complex and hypercomplex cells in the visual cortex, many studies have idealized the receptive field of simple cells with static Gabor filters (for a review, see (Field, 1993) or (Imbert & Buser, 1987)).

This is the first approximation on which we rely. More recent studies have emphasized other properties, especially temporal, such as receptive fields in (x,y,t) rather than just (x,y), and the fast adaptation of receptive fields size and shape to the image.

(DeAngelis, Ohzawa & Freeman, 1995), after an extensive review, conclude that receptive fields in V1 are spatio-temporal, sometimes separable ($RF = F(x,y) * G(t)$), sometimes not. When not separable, receptive fields often present a gradual shift in phase, which makes the cell as much a detector of the direction of motion as it is a detector of feature orientation. Motion direction selectivity appears principally driven by the tilt of RF in xy-t plane –a first order, linear effect– reinforced by non-linearities.

In essence, this comfort the idea that motion, along with color, texture and edges, is one of the primitive dimensions of encoding of the visual stimuli. From these primitives are extracted, amongst other things, the structure, location, color, texture, and movement of the object attended to. We suppose that, *in vivo*, all of these dimensions can be used to create a mental image. In contrast, the model presented here encodes only the edges, and its receptive fields are color-blind, texture-blind, and motion-blind. Moreover, the "higher representations" of the model are also extremely schematic. This is not really a problem, as the model could be extended to encode also these dimensions. However one must pay attention to the fact that often, V1 cells contribute to several of these dimensions. Thus, if imagery involves V1, all aspects of the image must be reconstituted. Because the same cells are responsible for movement and edges, or for edges and color, the population of cells is not made of isolated groups encoding each for one dimension, and in consequence it is impossible for a sustained activity pattern to encode only one dimension of the stimulus. Therefore, imagery, if or when it involves V1, must restitute all the primitive visual aspects of what is imaged: shape, color, texture, motion.

Receptive fields of other areas – distinguishing representational spaces

As the receptive fields used in the model are obviously patterned after those of V1, it is pointless to argue that details of the representations obtained by the model are relevant higher up. However, the principle that long-lasting transients or event stable

states can be sustained because of the lateral connectivity, can be applied at any level.[2]

Also, the argument that dimensions of the stimulus encoded together cannot be imaged separately can be carried at any level. Whatever an area encodes, it can be imaged all, or none of it is. For example, illusory contours appear to be encoded in V2 (von der Heydt, Peterhans & Baumgartner, 1984). If one visualizes the image shown in Figure III.15, with the level of detail and sharpness that requires participation of V2, then it must be impossible NOT to image along the illusory contours that one can see when looking at the actual stimulus.

Going further, one realizes that the most general version of this remark is that optical illusions will be imaged along with the image that generates them if and only if the neural wiring particularities which generate the illusion are located in an area recruited for the visualization. Thus, other kinds of illusions may not arise during visualization, in particular those illusions created by retinal or geniculate connectivity. There are several reports of imaged illusions (Berbaum & Chung, 1981; Wallace, 1984), which usually involve seeing part of the illusion and visualizing the items which alter the percept to create the illusion. Potentially, however, this constraint can become a powerful tool for identifying which representational subspaces are independent.

Conversely, this remark may be used to refute arguments that imagery is not co-located with vision because it fails to recreate *some* illusory visual effects. For example, Singer and Sheehan (1965), Over and Broerse (1972) and Rhodes and O'Leary (1985), in various psychophysics experiments, report various effects where imaged flanks fail to induce the effects of seen flanks, and all argue from this against the theories locating imagery in the visual system. The argument does not hold if these effects, though manifesting at a level reached by imagery, proceed from feedforward influences from a lower level, not implied in imagery.

Receptive field fast adaptation hint that lateral connections suffice to induce activity

Another line of research in receptive fields follows initial work by Gilbert and Wiesel (1990), and studies the change of the receptive field space-time geometry in a time

[2]See below, III.D.7 "computational power of less-than-total connectivities").//

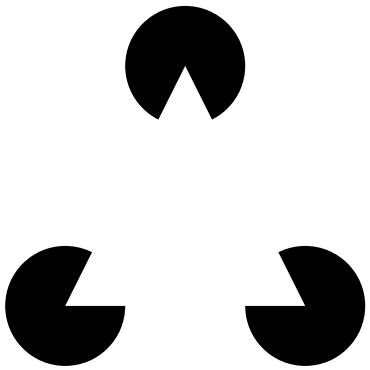

Figure III.15: **A type of illusion that attractor dynamics can explain:** This image makes us see illusory contours. One consequence of the theory illustrated by this model is that if one visualizes such an image with the vividness necessary to see sharp edges, it is impossible not to visualize also the illusory contours.

range of minutes; it appears that the RF size increases when a cell lacks stimulation, and comes back to its previous size under subsequent stimulation, see also DeAngelis, Anzai, Ohzawa and Freeman (1995) and Das and Gilbert (1995b). Das and Gilbert (1995a) suggest that this phenomenon comes from the adaptation of the firing threshold at the cellular level, so that input from the lateral connections, previously unable to activate the cell, becomes supra-liminar. This interpretation is supported by the independent results of Pei, Vidyasagar, Volgushev and Creutzfeld (1994), who uses patch clamp electrodes in the cat to measure not only spiking, but also subthreshold post-synaptic potential. Pei, Vidyasagar, Volgushev and Creutzfeld (1994) find that the main excitatory field –cells showing PSPs to the stimulus– is larger than the discharge field –cells spiking to the stimulus–; they also find that the PSP includes an early and a late component, and that cells on the side of the excitatory field receive almost only the late component. Thus, it appears that, even though lateral connectivity is weaker than feedforward connectivity, the cells themselves have a mechanism to adjust their threshold in the case where the feedforward input disappears.

These cellular changes, as fast as they are, still creep at a snail's pace compared to the dynamics which the model studies. However, what is interesting is that, after the assumed cellular adaptation, the lateral connectivity is sufficient to activate a cell, and hence the whole layer. Thus, even though the lateral connectivity is only strong enough to sharpen orientation selection in conditions of feedforward vision, in its absence this lateral connectivity becomes strong enough to drive the dynamics, just as it does in the model. Looking again at Figure III.10, we see that the gain is critical; however, as all units are arbitrary, if we diminish the firing threshold, the system behaves as if it had a higher gain, and hence a dynamics with no hope of activation becomes again bi-stable. Thus, the adaptation of the cellular threshold in the absence of input makes justice of the problem of the quantitative ratio of connectivity and threshold. Amongst the many changes triggered by consciousness alteration, this adaptation is the ideal candidate to account for the extreme vividness of dreaming imagery, and occasionally of sensory deprivation imagery.

Lateral connectivity

Reporting the connectivity of cell of known physiological function in visual cortex is a hard challenge, because it requires to match anatomy and physiology in the same study. On one hand, the anatomy of striate cortex connectivity is now well explored (see Callaway (1998)). On the other hand, because the connections are mapped *post mortem*, one does not know for certain what are the functional and receptive field properties of the cells to which these connections are made. The way through this Gordian knot lies with the systematic arrangement of the cortex in columns of orientation selectivity. Gilbert and Wiesel (1989) used slowly diffusing rhodamine-filled microspheres to map the connectivity of individual cells, then C-2-deoxyglucose to mark areas of identical orientation as the injected column. They found that horizontal striato-striate connections, as well as the feedforward V1-V2 connections, concentrate on cells of same orientation selectivity as the injected column. The patches of connectivity in V1 range mostly is from 6mm to 8 mm, corresponding to 10 degrees of visual angle, whereas the receptive fields are only of 0.5 to 2 degrees. Other studies have made use of the fact that cells of similar orientation selectivity are grouped in columns. (see, Bonhoeffer and Grinvald (1991) and

Maldonado, Gödecke, Gray and Bonhoeffer (1997), for examples of how electro-sensitive dyes can be used to determine *in vivo* the orientation selectivity of cortical columns). In a more recent study, Kisvarday, Toth, Rausch and Eysel (1997), analyzed the connectivity in cat cortex V1 and V2 using anterograde biocytin labeling. Orientation maps were obtained from intracellular recordings in cortical layers 2-5. Biocytin was injected injected in V1 in lower layer 3 and upper layer 4, in V2 in lower layer 3. Patches of labeled axons, 500-600 microns wide, appeared in v1 and v2. The average distance from patch to patch was 0.9mm in V1 and 1.2 mm in V2. Excitatory connectivity extended about 3 patches, whereas inhibitory connectivity was more local, extending 1.5 patch in average. However, even though the patchy pattern suggested a strong specialization of connectivity, the observed relation between connectivity and orientation selectivity only showed a moderate bias for iso-orientation excitation and cross-orientation inhibition. There was no significant difference for that between v1 and v2. These data are not contradictory with those of Gilbert; rather, Gilbert's must be moderated by the fact that cells of an orientation columns are not systematically selective for the column's orientation: as Maldonado and Gray (1996) state, "Tuning for orientation is more heterogeneous than previous studies have suggested". Thus, connectivity strongly constrained to a given column is not necessarily constrained to cells having the same orientation selectivity. In summary, the quantitative assessment of connectivity is still being made. Moreover, as the receptive fields appear to be defined by more parameters than just orientation selectivity (and especially, as seen above, by motion sensitivity), the cell-to-cell relationship classes presently used may well be too gross to be deterministic of connectivity. However, the current gross level of description appears sufficient to derive statistical regularities, such as the bias observed by Kisvarday, Toth, Rausch and Eysel (1997).

The connectivity adopted in this model is extremely consistent with the connectivity observed in Kisvarday, Toth, Rausch and Eysel (1997). In our model, the overlap of receptive fields tends to favor connections between cells encoding for the same orientation and aligned along that orientation. As said before, the connectivity used in the model is that which would result from hebbian learning cumulated over many random images. Several different learning mechanisms would also yield very similar patterns of connectivity, for example, mechanisms where hebbian learning occurs over a different

set of images, or mechanisms based on activity-regulated selection. Such mechanisms have been both modeled and hypothesized by neuroscientists. Kerszberg, Dehaene and Changeux (1992) have modeled a learning process based on the gradual pruning, moderated by simultaneous activity, of connections in a population of "neurons" responding to two different classes of input (as opposed to a continuum of orientation selectivities). Their Figures 2.a, 5.b and 6.a, which present the resulting connectivity, show a patchy organization strikingly alike that of V1, shown for example by Kisvarday, Toth, Rausch and Eysel (1997), in Figures 5 and 6. Callaway (1998) suggests that: "One resolution to these discrepancies is the possibility that horizontal connections can be formed and maintained during development as long as the neurons they interconnect have correlated activity at least some of the time, but not necessarily all the time". In summary, even though such a mechanism may entail slightly different quantitative distributions of connections than those we have used, it is quite possible that the connectivity in our model, conspicuous as it may be, has hit closer to the mark than we initially though.

Validity of the fixed-point systems decomposition: what can be counted as an 'isolated region' ?

As modelers and mathematicians we are naturally attracted towards the simplicity of attractor fixed point networks, especially like the brain-state-in-the-box or Hopfield models, whose dynamics can be understood in a crystal-clear geometric vision. For that reason we would like to slice the cortex in a series of BSB parts communicating on a time scale much slower than that of their internal dynamics. Under which circumstances would this approach be legitimate ?

Fixed point behavior arises in BSB models from the symmetry of the connections. Demonstrations given in Anderson, Silverstein, Ritz and Jones (1977) or Hopfield (1982), which show the existence of a Lyapunov function for two different models, apply, mutatis mutandis, to any net made of semi-linear neurons whose transfer function is monotonous. Since this is probably the case of the transfer function of most real neurons, the question becomes: "are there identifiable sets of neurons with symmetric connections and independent dynamics?"

Regarding the indenpendence of dynamics, it is a common mistake to confuse

anatomy and dynamics and thereby grant more importance to connection strength –or numbers– than connection speed. The differences in importance of connectivity that neuroanatomy finds are surely relevant for neural computation and function. However, as we try to untangle dynamics, the most important factor is time. Systems very strongly connected should still be considered independent if changes carried by these strong connections are slow compared to other dynamical events. Among the possible causes of such dynamical dissociation, two main candidates emerge. First, that the connection smears over time a sharp event in the presynaptic system, resulting in a slowly changing, quasi-static signal received in the post-synaptic system. Second, that changes in the pre-synaptic system are frozen during the active dynamics of the post-synaptic system. Are there biological cases where either situation is observed?

It is known (Rall, 1981) that passive dendritic transmission spreads the post-synaptic potential in time. However, axonal transmission keeps the action potential sharp. Since long-distance connections in the brain are axonal, we cannot assume that distance is a priori a factor smearing inter-area signal through time and making local dynamics independent.

Is there a reason why the surrounding areas could be stable during the lapse necessary for fixed-point stabilization? A point can be made that it is possible that the whole cascade of visual processing is time locked onto the end of ocular saccades. An analogous phenomenon has been observed in the olfactory system of the rabbit Freeman and Viana Di Prisco (1986), where processing is timelocked to the inspiration puff. Also, saccadic suppression suggests that there is a reset of visual processing during the saccade. In this context, it is likely that higher areas will be quiescent, or at least stable, during the feedforward sweep, which means that local activity does not have to compete with changes in the established top-down input until the time necessary for feed-forward plus feedback inter-area transmission. The situation is different with respect to bottom-up input. Ganglion cells have time-dependent receptive fields, and the information transmitted by the LGN likely reflects the time variations. At higher levels, successive areas would have to deal with the settling of all lower areas, i.e. to gradually improving input. Thus, how good an approximation temporal independence can be would depend on the ratio of sustain time to settling time. In conclusion, the hypothesis

that the independent subsystems of the brain correspond exactly to the separate areas – retinotopic maps in the occipital cortex, and functionally identified areas in parietal and temporal cortex– is still to be experimentally studied.

Summary on the realism of model details, and consequences

Of the two main sets of model assumption, that concerning the cells and that concerning their connectivity, the second is truer to the brain than the first. Striate cortex pyramidal cells often have receptive fields sensitive to motion direction, or to the time course of the stimulus. Many different kinds of cells are closely knit in the same dynamics, and we have only considered one type. The size and frequency of the receptive fields *in vivo* depend on the retinal eccentricity; not so in our model. Also, the model varied the cell parameter systematically along a limited discrete scale; in the cortex these parameters are randomly distributed along a continuum; the latter scheme allows for much finer population coding. We believe that the difference between the sharp boundaries of real vision and the fuzzy boundary of images observed in the model comes more from the quality of population coding in the cortex than from the arguable existence of high-frequency cells.

Taking into account the increase in receptive field size with retinal eccentricity requires further study, but might be an easier than it appears. The core of our difficulty in accepting the brain's non-Cartesian coordinate system is when we want to conceive of activation patterns as standalone entities separated from the particular neurons which bear them. Clearly, in our model, a shifted input image induces the same pattern, but shifted, as the original image. It cannot be so when the receptive fields have different sizes at different eccentricities. However, there is not one reason to think that pattern translation invariance could be useful in the cortex. In the brainstem, it is presently discussed that translation invariant patterns could mark the retinotopic location of saccadic target; but such patterns would consists of simple lumps of activation, encoding an amount of information orders of magnitude below what V1 encodes. The hypothetical translation circuitry is complex and requires many neurons to encode gradients along the x and y axis; similar mechanisms, to implement translation for the many aspects of a complex image as encoded in V1, would take by itself 80 % of the neural machinery.

Thus, it is extremely unlikely that there are in the cortex any mechanism for shifting detailed activity patterns. As a consequence, the use of a non-Cartesian coordinate system presents no essential difference from the use of a Cartesian referential. If all the inter-area mapping takes into account the skewed geometry of the referential, there is no reason to expect a very different behavior than that observed with our model, where the geometry has been rectified consistently at all levels. In essence, this problem reduces to the simple mathematical problem of a change of referential; what is lost in the change from Cartesian to cortical coordinates is the translation invariance property, of no interest anyway. For these reasons, even though the final proof will require further modeling, we do not expect very different results if the geometry of the model was made eccentricity dependent.

Connectivity appears, in hindsight, as the strong point of the model. Considering the simplifications which have been made for the receptive fields, the connections between units are very consistent with present measures.

The gain of the lateral connections, a parameter so critical in the simulations, appear to be modulated *in vivo*, taking values sometimes above and sometimes below the critical gain. This strongly supports the theory that very different kind of imageries can happen in the same subject depending on the state of consciousness, and that these different imageries have strong neural correlates.

The main question remaining to be addressed is whether it is legitimate to isolate a layer as an independent dynamic system. More likely than not, the visual system can be sliced into subsystem with relative temporal independence. Whether these subsystems correspond exactly to cortical areas, or whether they sometimes isolate some cortical circuits within the same area, or whether they are made of two or more areas with fast connections, determines which representations are the relevant building blocks of the variable construct which is the representational space. For example, it might have been that the vision of color is indissociable from that of shape. Our experience viewing black and white movies, or dreaming in black an white, or meditating on pure colors, strongly argues against this fact. Thus, it is likely that color and shape systems are temporally independent. Finding out which are the relevant building blocks would be, *per se*, a hefty research program, but even as we stay in ignorance of the correct way

to slice the visual system into subsystems, the basic mechanisms demonstrated by this model can be supposed to apply for those subsystems.

In conclusion about the validity of the model's detail, it must be emphasized that because the cells and connections were modeled after V1 or, at best, V2, if the model's behavior correspond *in detail* to cortical behaviors, it is most likely eidetic imagery rather than plain imagery which is modeled, or perhaps even some particularly vivid sub-species of eidetic imagery, such as was experienced, for example, by Luria's patient Vienamin (Luria, 1968) or by the eidetic subject of Purdy (1936). However, besides the particulars of the re-created representation, the same type of behavior – adapted pattern completion– should arise at several different areas in the visual stream –the criterion for its emergence being that the local lateral connectivity must be sufficient. Thus, several kind of imageries could be observed, depending on which regions have been successfully reactivated. It is our sincere opinion that the model's kind of behavior is not at work in V1 to create ordinary imagery. Rather, analog phenomena are at work in higher cortical areas, most likely V4, in ordinary vivid imagery. Only in pathological imagery, be it systematic or hallucinatory, as in vivid dreams, or when substance influence or fever create powerfully vivid hallucinations, are the phenomena herein modeled likely to account for the vividness of the imagery experienced.

III.D.3 What we believe the model will be robust to

As briefly explained in the method section, the quantitative results presented here come from a second-generation model. The first model presented qualitatively the same behaviors, yet the filters associated with each neuron, and in consequence the connection weights we originated as filter convolutions, were fairly different. Moreover, this original model only had odd filters, equivalent to the sine phase in the present model, without including any cosine-Gabor-like filter.

For this reason, we estimate that the model should be robust to changes in the receptive fields shape and size. It should also be robust to changes in the detailed shape of the connectivity, provided that the connectivity remains fairly in ratio of the overlap of the receptive fields of two cells.

Finally, the model is also likely to be robust to changes in the seed degradation

algorithm. In our pilot model we used, instead of gaussian noise, a compression algorithm: the seed layer was made of 400 neurons each connected with random weights to only 16 wavelet units. The seed was obtained by a forward pass through these random weights, and was used by a backward pass through the same weights. That seed layer was exactly to the wavelet layer what the "association area" is to the "projection area" in old-fashioned perceptrons. Thus, just like in perceptrons, it could store a classification that was still very detailed (400 neurons) but nowhere near the original level of detail (3000 neurons). Moreover, the backwards pass used the same weights, rather than a pseudo-inverse of the weight matrix, introducing a systematical skewing. The qualitative results obtained with that compression degradation algorithm were similar to those presented here. The model's robustness to seed degradation is the more important as feedback connections in the cortex are believed to be more diffuse than feedforward connections, and thereby naturally degrade the top-down input.

III.D.4 What we believe the model won't be robust to

Insulating the dynamics of a subsystem from that of its supersystem is a tricky business. This applies especially to the brain, and the *regions* to which the model applies should have a settling time (the analog of four to five cycles in this model) smaller than the time given between brutal global changes of external conditions. For vision, this time lapse might be the delimited by ocular saccades, which might play in vision a role analogous to that which inspirations play in the rabbit's olfaction (Freeman & Viana Di Prisco, 1986). In the event that the dynamical changes of the whole visual system occur on the same or a faster time scale than local changes in visual areas, it would be perspicuous to pretend to isolate regions with an independent dynamics. In this case, the behaviors observed in this local model might not emerge, or, if they emerge as the consequence of global, system-wide, recirculation of activity, they could have much more complex properties.

III.D.5 Merging older models to get pattern-specific dynamics

Almost all elements of this model have been found separately in others, but their assembling is what gives the model its full relevance for understanding pattern

reconstitution in mental imagery. The classic models by Wilson and Cowan (1972, Wilson and Cowan (1973) demonstrate a more complex dynamics than this model, but are respectively restricted to zero and then one dimension, so that one cannot speak of pattern completion. More recent, fixed point models with a simpler dynamics, such as those of Anderson, Silverstein, Ritz and Jones (1977, Hopfield (1982) have not been used with a main layer encoding an image by wavelets. Finally, the connection weights chosen for this model bring the right attractors to those fixed-point networks, and at the same time are amazingly realistic, and could have come from a quite plausible learning scheme. A similar approach was that of Somers, Nelson and Sur (1995), but the gain in Somer's model is too weak to implement pattern-completion. Thus, the merging of the classical elements which are a fixed-point network with units which are wavelet-encoders, with a new but logical connectivity, produces meaningful results which could account for pattern completion in imagery.

III.D.6 What this model suggests for the phenomena of mental imagery

Coming back to our initial concern, we will suppose that the principle of informative pattern completion, illustrated by the model can be applied at several different levels. "Informative" pattern completion, because, as shown in our results, the very shape of the pattern is improved by the completion. Let us, so, suppose that similar completion occurs in different areas. If the specification for an image's color is strong enough, with respect to the gain of lateral connection in the region which encodes abstract color, then abstract color shall be visualized. If movement is part of what we wish to see, such as when visualizing a racer, and the gain conditions are again good, then movement we shall see, in the sense that our MST, or whatever region encodes for motion, will be activated just the same by our imagery as by our vision. Likewise, in visual processing, if we see a black and white movie, soon our color area will cease to fire, and color will no longer be part of our world; we will temporarily become monochromats. This duality of state of brain regions can apply at all levels of the visual system. Several levels of encoding, mostly occipital, are still spatiotopic, and several after those are still visual, but with receptive fields so wide that it is a joke to call them retinotopic. Any, or

several, of these levels could be the loci of the kind dynamics modeled here for the lowest level only. Images, from vision or imagery, have a variable geometry; from individual to individual, between different conditions, the representational space will be composed of different dimensions.

What the model shows is that two elements govern the appearance of one dimension in the representational space. The first is the potential, governed by the gain of the lateral connections and by the strength of the input connections. The second is the actual strength of the input. It appears that most often, in imagery situations, the potential for low-level reconstitution is not there. No matter how hard we try, this famous man's face does not appear to the mind's eye as vividly drawn and colored as it appear to the body's eye. Yet, change the gain by ingesting the right amount of intoxicant, and likely hallucinations will sprout with the same vividness as reality. The potential can sometimes be there all the time, to the point where imagery is indistinguishable from reality, as for Luria's patient. But sometimes it cannot be reached, even with the help of potent hallucinogens, as for Aldous (Huxley, 1957), who, generalizing from his experience, writes: "Evidently, if you are not a habitual or congenial visualizer, you do not get internal visions under mescaline or LSD – only external transfiguration" (p.86). Sometimes the potential is there, but the input is marred, and only in conjunction with imagery can we make sense of it. Thus, also, does imagery play tricks with perception, when there was indeed nothing to see , and we form an image out of visual noise, and think we have seen, because indeed, the image was formed out of more than just imagery. Truly the model support James'position, that there are imageries, not Imagery.

We are now near the end of our endeavor. Is there yet anything to say about the representations which are take turns making up the various kind of mental imageries? Returning to one oddity of the model will suggest additional ideas to our analysis.

III.D.7 Computational power of less-than-total connectivity: the storage-computation trade-off

For those familiar with the linear analysis of totally connected neural networks, such as Anderson's (Anderson, Silverstein, Ritz & Jones, 1977) brain-state-in-the-box or Hopfield's (Hopfield, 1982) recurrent net, obtaining as many quasi-attractors as there are

plausible images in the world sounds like a paradox. Yet it is what this model does. The trick is that the real connectivity (non-negligible weights) is very local, and hence, sparse, thus creating a highly degenerated weight matrix. Because so many of the connection weights are nearly zero, the constriction of the activity pattern is much lighter than that of a hebbian-learned BSB net, and this allows whole families of states to be acceptable attractors, or at least long term transients, in a sense, "quasi-attractors".

This observation that a sparser connectivity allows more storage is interesting enough to deserve deeper probing. Let us consider the simplest recurrent net: where each neuron is connected exclusively to itself. If all weights are positive, this net can store any pattern of activity; it is the perfect short term memory; yet, it has no processing ability, because no pattern will be transformed by it. In contrast, a net with total neuron-to-neuron connectivity, in its generic shape, that is, with a non-degenerate weight matrix, can store at most as many states as it has neurons. (One must remember that degenerate matrices are a subset of measure zero in the "right" space of matrices, which means that the probability, given a weight matrix with randomly chosen weights, that it is degenerate, is zero). However, this totally connected network has classification properties: it reduces any initial state to one of the attractors. Going further, one can see that when a network includes three-way connections ("sigma-pi connections"), its dynamic becomes able to discriminate between exact opposite states –something that the Hopfield net is incapable of– thus restricting even further the number of available attractors, at the same time as it enables more sophisticated computations. Thus, it appears that a growing number of connections improves the computational ability of the network, at the expense of its representational range.

Unfortunately, we have been unable to cast this observation in mathematical form. Yet, besides the challenge of formalizing this observation, these considerations raise the question of whether the same density of connectivity is at work everywhere in the brain or not. Could it be that the attentional bottleneck is made necessary by a boundary between areas with a more locally organized connectivity, apt at representing a wide range of perceptual input, and areas with denser, more homogenous interconnectivity, more efficient at classification but less flexible in representation? Might it even be possible that some premotor or motor areas would even make use of sigma-pi connec-

tions, better suited for storing sequences of events, but limiting to very small numbers the possible attractor sequences? If this were the case, the classical view of the hierarchical brain connecting perception to action should be expanded into a view of the brain connecting essentially spatial patterns to temporal patterns. Also, the role of mental imagery in such a picture would be even more crucial: mental imagery would supply short-term memory with much higher storage capacity, and its spatial properties could complement the temporal properties of higher-level planning representations.

Such a possibility lends itself to experimental verification. At this point, there is no hint that the cortex lateral connectivity increases in complexity from the perceptual to the motor areas. The only systematic change is that of a change in the relative thickness of cortical layers. Should, however, that change of thickness implement this increase in connectivity, the theoretical possibilities opened by our remark would have to be further explored.

III.D.8 Conclusion: Classification starts in V1

As we near the end of our journey, we would like to place ourselves, modestly, in this perspective of "pure vision" (Churchland, Ramachandran & Sejnowski, 1994), which we know to be false, and deal it, from within, one more blow. In the perspective of *pure vision*, a sequence of representations analyze a stimulus until it can be classified. What appears in our view is that, even when keeping a sequence of steps, which each update their own representation, each of these steps is per se a classification. Our model does not admit any kind of activity pattern: it forces the pattern to be one among a very numerous family, but still imposes drastic constraints compared to the potential number of states. Likely, as processing goes on, attractors acquire more meaning, and eventually are symbolic. The view which we present is that these different levels of representation are meant to work together, and that the classification operated by occipital areas is no less important than that made by higher centers. Low-level classification fills the double rôle of preprocessing the stimulus for higher levels, and of reconstituting a correct representation from an approximate top-down specification. Thus, top-down feedback locates the initial state reasonably near the target representation, and on its way to an attractor, the activity pattern reaches higher likeliness to the target representation.

The rôle of the local dynamics there is to reconstruct a representation which obeys the constraints that the connectivity of that low-level area implements – in our model, the alignment of edges.

Thus, it is not only that Imagery is, as Kosslyn suggested, top-down mechanisms of vision running on empty, but rather that vision, including imagery, is a complex system made of interacting subsystems which can temporarily be reshaped to include such and such of these subsystems, so that the dimensions of space be not lost to the exclusive profit of the dimension of time.

III.E Appendices

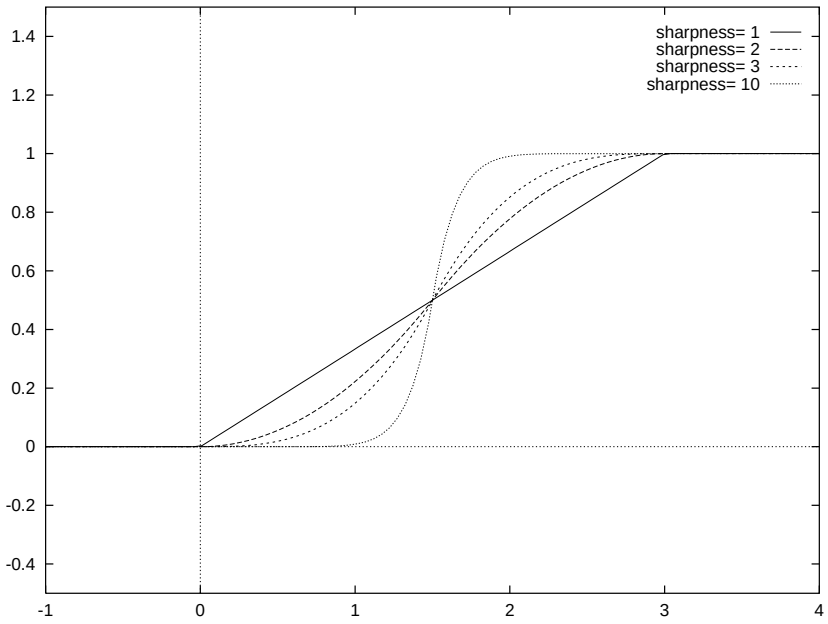

Figure III.16: **Transfer functions:** The family of transfer function which we use is illustrated here by functions corresponding to an increasing sharpness of 1, 2, 3 and 10. As can be seen, a sharpness of one corresponds to the linear case. If the sharpness were between 0 and 1 the function would swing the opposite way.

III.E.1 Appendix 1: A family of transfer functions

Most modelers are presently happy with using the sigmoid as the transfer function of continuous-valued neuron models. We are not. The reason is that with the sigmoid, if one wants to shut down the zero-input noise, the sigmoid has to be taken very sharp, thus having little linear range. In other terms, the "linear range" of the sigmoid and its sharpness depend on a single parameter.

Thus, conscious of bearing the full brunt of mathematical inelegance, we have chosen to use a family of transfer functions made by pasting polynomials. These functions use one more parameter, thereby allowing free choice of both the range and the sharpness.

The family we choose are continuous and usually differentiable functions. Given the lower bound of the range O and the threshold T, the transfer function of sharpness S is obtained by by fitting x^S from (O,0) to (T, 0.5) analytically:

$$TF(x) = 0.5 * [(x - O)/(T - O)]^S \tag{III.8}$$

and between the threshold T and the upper bound 2T-O, the symmetrical curve with respect to (T,0.5):

$$TF(x) = 1.0 - 0.5 * [(2T - O - x)/(T - O)]^S \tag{III.9}$$

Below O the function is 0, and above 2T-O, it is 1.

The two curves being symmetrical, join with the same tangent of slope S/2(T-O) at the middle point (T, 0.5). This guarantees differentiability at the center. If the sharpness is strictly greater than 1, which should be the case for any sigmoid-like function, the function is also differentiable at both O and 2T-O.

In our model, after trying several possible sharpnesses, we have settled for S=2, O=0; T=1.5.

III.E.2 Appendix 2: The one-unit closed system

A one unit system with a sigmoid-like transfer function which feeds back on itself will exhibit bistability if the gain of the feedback is high enough. It is still the case for the transfer function presented in Appendix 1, and we present here the adaptation of the classical analysis which gives the threshold as a function of the feedback. We first

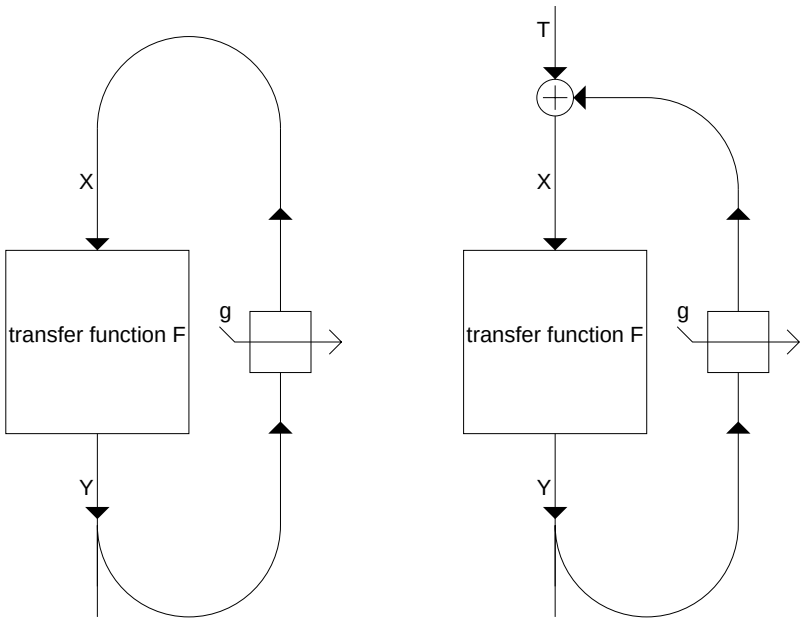

Figure III.17: **System made of a single neuron with feedback:** For non-linear transfer functions, these systems can have a threshold over which activity is sustained (or amplified) by recirculation. To the left, a system with one shot initial input, afterwards left to its own devices (our "perception" simulation). To the right, a system with constant input (our "imagery" situation).

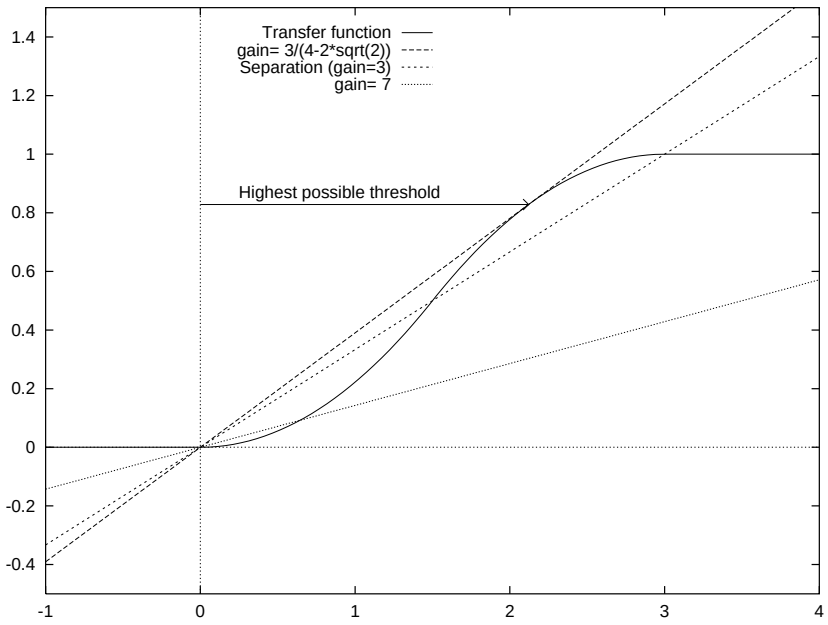

Figure III.18: **Construction of the analysis:** The lines represent the feedback and the curve is the sigmoid-like transfer function. The possible activity states are at the intersections of a line and the curve. Different lines stand for different possible feedback gains. The lower the line, the higher the gain. The line with gain=3.0 (line of slope 1/3) separates the two cases dealt with in the analysis.

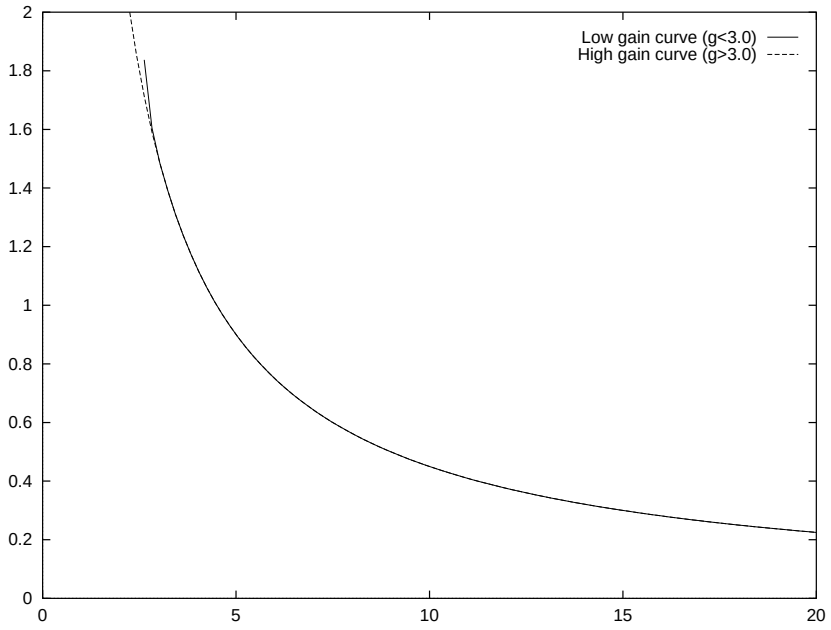

Figure III.19: **Theoretical threshold as a function of gain:** The curve of theoretical dependence of threshold on gain is made of the little bit of a line and goes on with the hyperbole. As can be seen, the hyperbole by itself is a good approximation for the whole, provided it is truncated at the same abscissa.

analyze the standalone system, with no input except the initial activity (one-shot input), then the system with ongoing input.

The condition for having a steady state is that, with a feedback of gain g and a transfer function F, the activity y can follow the equation:

$$y = TF(g.y) \tag{III.10}$$

The values of x for which this is possible are those which obey the system:

$$y = TF(x) \tag{III.11}$$

$$y = 1/g.x \tag{III.12}$$

This system is illustrated geometrically by Figure III.18. The line of equation $y = 1/g.x$ always cuts the transfer function curve $y = TF(x)$ at $x = 0$. This is the low-activity steady state. If the gain g is high enough, then there are two more solutions, one which is the threshold level activity, and is an unstable state, and the other which is the high activity steady state.

Since the transfer function is made of two pieces, the derivation follows a disjunction of cases. This is illustrated in Figure III.18. For a narrow range of gain ($g_0 < g < 3.0$), the threshold and the activity range are both on the second part of the curve. For all higher gains ($g > 3.0$), the threshold is on the first part of the curve and the activity saturates at 1.0. We will start by looking at the low gain situation, then will look at the high gain situation; afterwards we will paste the solutions. It will actually appear that the high gain solution encompasses most of the interesting range, and over the small range where it doesn't, still is a good approximation to the low-gain solution.

Low gain threshold

Expressing the transfer function and the gain, the system is:

$$y = 1 - 0.5(2 - x/1.5)^2 \tag{III.13}$$

$$y = (1/g)x \tag{III.14}$$

where, by our "low threshold" condition, 1/g must be over 1/3. giving the second degree equation of which the intersection x are the roots:

$$x = g[1 - 2(1 - 2x/3 + x^2/9)] \tag{III.15}$$

the parameter change $G = 3/(2g)$ simplifies the equation to

$$Gx = -3/2 + 2x - x^2/3 \tag{III.16}$$

that is

$$x^2/3 + (G - 2)x + 3/2 = 0 \tag{III.17}$$

the discriminant is

$$\Delta = G_4^2 - 4G - 4.1/3.3/2 = G^2 - 4G + 2 \tag{III.18}$$

with the roots $2 + \sqrt{2}$ and $2 - \sqrt{2}$.

When G equals one of these roots, Δ is zero and the threshold is confounded with the activity: it is the G of appearance or disappearance of the high regime. The corresponding tangent points have abscissas $+3\sqrt{2}/2$ and $-3\sqrt{2}/2$. The latter corresponds to the intersection with a portion of the parabola which is no longer part of the transfer function, hence the first value is the real apparition of threshold:

$$activity = 3\sqrt{2}/2 \tag{III.19}$$

$$gain = 3/(4 - 2\sqrt{2}) \tag{III.20}$$

The gain for which the high regime appears in this one-unit system is therefore about 2.56, and the function relating gain and threshold is:

$$t = 3 - 9/(4g) - 3.\sqrt{9/(16g^2) - 3(2g) + 1/2} \tag{III.21}$$

High gain threshold

Expressing the transfer function and the gain, the system is:

$$y = (2/3.x)^2/2 \tag{III.22}$$

$$y = (1/g)x \tag{III.23}$$

where, by our "high threshold" condition, 1/g must be under 1/3. This reduces to

$$x = 2g/9x^2 \tag{III.24}$$

and the relationship between threshold and gain is hyperbolic:

$$t = 9/(2g) \tag{III.25}$$

Put together: threshold=F(gain) for no-input system

Figure III.19 summarizes the dependence of the threshold on the gain. The hyperbolic part of the curve has been plotted past its application range for the sake of comparison with the part of the curve owed to the "low gain" phase. It can be seen that the hyperbolic curve all by itself (truncated to $g > g_0$) is a good approximation of the whole curve.

One-neuron closed system with ongoing input

Relying on the previous findings, we limit the analysis to the high gain case. The we can express the steady-state Y as a function of the steady input T, and there is a singularity where dY/dT is infinite, which is the threshold. Derivations are a little more complex than those made in the case without input, and show that the steady input changes the result by a factor 1/4:

$$t = 9/(8g) \tag{III.26}$$

Since the transfer function is quadratic this can be seen by realizing that if X_0, Y_0 are the steady state variables of the first system, then with $T = X_0/4$ we have the steady state variables for the second system $X_1 = X_0/2$ and correspondingly, after the quadratic transfer, $Y_1 = Y_0/4$. Thus, steady input is sufficient to kick off the system if it reaches a quarter of the activity that is necessary in the case of one-shot input.

III.E.3 Appendix 3: Detailed parameters of the model

imWidth	30	pixel width of input image
imHeight	30	pixel height of input image
gain	4.0 v	global multiplier of the recurrent connections
threshold	1.5	threshold of the transfer function
sharpness	2.00	sharpness of the transfer function
fbfff	1.0	ratio of the recurrent over feedforward input strength
powerDiv	1.0	(no influence at this value - unused degree of freedom)
scaleInc	2.6	ratio respective bandwidths of successive scales
iniBand	2.4	bandwidths of smallest scale, in input pixels
steps	varied	duration of default run
visible_image	on/off	for online display of input image
visible_retina	on/off	for online display of buffer
visible_wavelet	on/off	for online display of wavelet layer
visible_seed	on/off	for online display of imagery seed
seed_rescale	1.0	ratio of top-down input strength to wavelet layer output
seed_prune	0.4 v	ratio of seed units reset to zero
seed_noise	0.8 v	magnitude of added gaussian noise
sigratio	1.1	ratio of gaussian to sinusoid bandwidths for Gabor filters
XDIM	10 j	wavelet layer width
YDIM	10 j	wavelet layer depth
CUTS	10 j	number of different orientations; CUTS MUST be 4*k+2
SCALES	3 j	different spatial scales
PHASES	2	sine and cosine filters
GRAYSCALE	3	ratio of retinal to V1 size
XGRAY	30 j	retinal width, must be GRAYSCALE*XDIM and ¡ 127
YGRAY	30 j	retinal depth, must be GRAYSCALE*YDIM and ¡ 127

Table III.1: **Constants and Parameters used for the simulations:** "v" indicates variable values depending on the simulation, 'j" that the values were changed for the "Imaging the Mona Lisa" simulation

Chapter IV

Content-Specific Low Level Interaction between Imagery and Perception

IV.A introduction

In the first chapter of this thesis, we saw that imagery, unlike vision, does not have to recreate all the major properties of an image: color, form, shading, texture... As William James wrote, "there are imaginations not Imagination". Sometimes imagery is pictorial, sometimes not. Sometimes it is vivid, sometimes it is abstract. Respecting this fact, we proposed that imagery is like a swiss army knife. In a swiss army knife, available tools are not all pulled out at the same time; likewise, in imagery, many visual representation spaces can be unused, "folded". How? Because pattern reconstruction is best made by an attractor dynamics which makes whole regions of the visual brain bistable, so that they are either taking part to the representation of a mental image, or not. So, if the area encoding color is active, the mental image will be colored. Else, it will be like an old movie – we don't really miss the color but we can notice it is missing if we pay attention to it. Depending on the active regions, a mental image may be encoded in very different representational spaces: with or without color, details of shape, texture, movement, etc... Some of these spaces are more or less pictorial,

that is, respecting the external spatial layout. Some aren't. Thus, imagery can include low and hi-level components, and its varieties are richer than suggested by the simple propositional/pictorial dichotomy.

In the second chapter, we illustrated this attractor dynamics so crucial to our theory with a region using wavelet encoding, very analogous to what can be observed in striate cortex. One effect observed on this model is that imagery and perception can be combined, but that the combination brings fruit only when the mental image formed is that of the object seen. For example, visualizing the Mona Lisa should help you see it in the dark Louvre room where it hangs, but it won't help you seeing the El Greco paintings in the next room. This may seem obvious, but one line of influential psychologists have argued that visualizing the Mona Lisa will actually hinder seeing the Mona Lisa as much as any other scene.

Therefore, the experiment we present in this chapter addresses the importance of similarity between the mental image and the stimulus. Does visualizing the Mona Lisa have a different effect when we are looking at a Mona Lisa and when we are looking at another painting? Yes. When the mental image formed and the stimulus are similar, we predict that perception of the stimulus will be facilitated[1]. By the same reasoning, when the mental image and the stimulus are dissimilar, we predict that the perception of the stimulus will be inhibited. Evidently the key factor is how closely the stimulus and mental image resemble each other. Thus, how imagery interacts with perception depends on the content of the mental image.

Few studies with similar goals can be found in the literature, and none with controls that establish unambiguously that the effect observed is due to imagery. Classical experiments always measure some perceptual performance during the visualization of a cue. The experimental study presented in this chapter is the first which used a control condition where a visually presented cue replaces the imaged cue of the main condition. We argue that the effects observed with ongoing imagery but not with prior vision cannot be explained by information-acquisition strategies, knowledge-based strategies, or

[1] To be precise, perceptual facilitation in this context means facilitation at the levels of pictorial representation. Because the response itself depends also on higher processes that are sensitive to mental load, the facilitation has to be studied differentially.

classical priming. For example, if one expects a particular perception, one can get ready by moving one's eyes to the most salient part of the expected image, to facilitate its identification. Even though, such a strategy uses prior knowledge to facilitate perception, it is not a low-level interaction of the kind we hypothesize. Fortunately, this strategy is detectable by the control condition, because it also comes into play if one has just seen the stimulus, rather than visualized it. Hence, if an effect is observed with an imaged cue only, and not with a visual cue, it cannot be the result of such a strategy. Similar arguments apply to facilitation by classical priming or expectation. By virtue of this control condition, the results of this experiment are less ambiguous than any previous results, and show that the effect of imagery on perception depends on which image is formed.

IV.A.1 Historical Background

Ever since its birth in faraway Europe, Psychology has had as one of its primary concerns understanding the nature and the rôle of mental images. As early pioneers such as Wundt, Galton and James attacked the issue frontwise using their own introspection, as well as that of others, they soon came to realize that the bias of everyday language was taking root in the sincere impression of a consequent minority, for whom imagery was indeed the faculty of forming images similar to those summoned by the sense of vision. Thereupon, the pendulum of Science History swung to and fro and the perceptual nature of imagery was denied, then reaffirmed, then questioned, and, like the phoenix, is presently arising, in its latest metamorphosis, from joint arguments of neuropsychology and cognitive psychology.

As the works of Paivio, Finke, Shephard, Cooper, Farah and Kosslyn reconnect with the intuitions of the early masters, they also set the question in the wider, well-studied framework of top-down effects. Since auditory top-down effects facilitate the parsing of natural language and the perception of words, one could wonder if visual top-down effects could not be found to facilitate parsing of the visual scene, figure-ground segmentation, and general visual perception. Kosslyn (1994) argues that these top-down effects, used *à vide,* are responsible for imagery. Imagery then is top-down reactivation of visual representation by mechanisms mostly evolved for helping bottom-up vision. A

consequence of this thesis is that vision and imagery should interact, priming each other when the same target is seen and visualized, or impeding each other when dissimilar targets are simultaneously seen and visualized.

Several studies have addressed the interaction of imagery and vision. There are indirect studies which we will not review, that have studied parallel effects of imagery and vision, for example on after-images such as the McCollough effect (Finke & Schmidt, 1977; Finke & Schmidt, 1978; Broerse & Crassini, 1980). There are also direct studies, which can grossly be classified in three types:

- studies which focus on demonstrating a task-related, content-independent, interference of imagery with perception,
- psychophysical studies trying to reproduce with an imaged context the same alterations of the perception of a central stimulus which are observed when the same context is really seen,
- studies looking in particular at content-dependent interaction between the mental image and visual perception. (Interaction depending on the content of the mental image).

Studies focused on content-independent interference with vision

Many studies looked at a putative inhibition of vision by imagery, called the 'Perky effect', after Cheeves Perky (1910). Perky (see below) had an introspectionist approach and wanted to show that mental images could not be distinguished from faint visual images. Her subjects indeed did not report faint visual images when they were superimposed with their intentional mental images. Later, Perky's result was reinterpreted in the context of signal-detection theory. Segal, Reeves, and now Craver-Lemley, amongst others, tried to demonstrate that there is a non-specific inhibition of perception by imagery. Their approach is not conclusive about content-specific interaction, because it shifted from early studies where some of the mental images were of objects similar to the stimuli but which had method problems to later, better controlled studies, where negative effects on perception were always observed with mental images of objects very different from the stimuli. These studies are well worth reviewing, however, because,

they illustrate well the role of some undesirable actors in the game, e.g. the confusion between mental and visual images, or the global mental load.

Cheves Perky, a disciple of Titchener, opened the field with a long paper (1910) principally dedicated to the difference between remembered images and constructed images, but mostly remembered for her first experiment, which set the paradigm for the Perky effect. In this experiment, subjects were sitting in front of a screen where, unbeknownst to them, colored shapes of blurred, oscillating edges and low contrast, slowly increasing from subliminal to just supraliminal, were being projected just at the same time as the subject, following instructions, tried to visualize them. The items were a blue book, a green leaf, a deep yellow banana, a red tomato, an orange, and a light yellow lemon. The subjects (three children, then eight graduate students and nineteen undergraduates) always reported that they were forming the images, but never thought that they were seeing them. Perky's point was to show that perceptual and mental images were of the same kind; however her result was later given the interpretation that imagery interfered with vision. Several of Perky's subjects reported that their imagery appeared more vivid in the experiment; short of really addressing our problem, this could be taken as positive influence of vision on imagery – the other side of the coin.

After a long period of fallowness, the Perky effect was again studied, especially by a line of scientists starting with Sidney Segal, who have focused since the early sixties at exploring the interference of imagery with vision, under the general hypothesis that mental imagery lowered the visual perceptual capacities.

Segal and Glicksman (1967) reproduced the Perky effect on twenty-six female students, varying their bodily position. The Perky effect, being the report of imagery where the image was actually seen, was measured on a 1 to 5 scale in three body positions: supine, sitting, standing. The subjects were also sorted as tensed or relaxed. Mostly for relaxed subjects, the Perky effect was found much higher when the subjects were lying supine (1.22 versus 4.24). The authors suggest that awareness level plays a role, and conclude that "simultaneous increase in imagery, increase in their vividness, diminution of reality-testing, and raising of the perceptual threshold in this state, all contribute to the effect. It is difficult to isolate the separate contribution of these effects, as all tend to covary along with the level of awareness". Thus, in this study, the influence of imagery

and mental state on the subjects'reports results more from the subjects'interpretations than from their perception.

Segal and Gordon (1969) compared the Perky effect in the Perky situation, where subjects ignore that a stimulus is being projected, and in an informed situation where subjects knew that stimuli were sometimes projected. The authors found that imagery interferes even in the informed condition. However, the methods can be questioned on two grounds: first, the measure is the sensitivity D' according to signal theory, but it appears to be computed after averaging the hit scores across subjects, which can considerably alter the measure (see Appendix 11); second, the results could be interpreted as order effects from gradually learning signal discrimination, because the naive task always preceded the informed task, which itself preceded the discrimination control task. Altogether, that study carries little conviction by itself, but, because it corroborates earlier and later studies, strengthens the case for the Perky effect.

Segal and Fusella (1969) addressed the problem of order effects found in the previous study. In addition, they introduced a second factor: the projection of the slide could be a pulse, with brutal onset and removal, or a gradual projection. Images formed were of common items, whereas stimuli were colored geometric forms, it is not specified whether images and stimuli were similar or not, but because the authors later say that imagery brings "noise", one is lead to believe that in most cases they weren't. Regardless of the task order, the sensitivity for pure detection was always better than that of the imaging condition. Total intensity summed over time, rather than gradualness, was determinant for the sensitivity. Again, the sensitivity appears to be computed from aggregate rates, but the repetition of similar results over several studies lowers the chance that the randomness introduced by the wrong D' transform could be responsible of all results. The authors conclude that the difference in sensitivities represent the internal signal or noise contributed by imagery.

Segal and Fusella (1970) presented two essentially identical experiments, the second replicating the first with more trials and fine tuned lower noise conditions. They compared sensitivity for visual and auditory signals between a detection task and two imagery tasks: visual and auditory. Neither of the signals were the target of the respective visualizations. The visual signal was a blue arrow, never visualized, and the

auditory signal was a harmonica cord, never imaged auditorily. Imagery was found to lower D' overall, but isomodal imaging conditions gave significantly lower D' than cross-modal conditions. Therefore, diversion of attention alone should not be responsible for the Perky effect. Looking carefully at the data, especially in the second, cleanest experiment, it appears that the difference between pure detection and imagery conditions was mostly a lower hit rate indifferently of the imagery modality. So, subjects would miss more signals when they were visualizing in either modality – presumably an effect of attention. In contrast, imaging in the signal modality increased the false alarm rate. This, along with the same D' computation problem found in previous studies, argues against a D' analysis and in favor of an effect-specific analysis, where one would argue that attentional load lowers the hit rate, and isomodal confusions increase the false alarm rate. Whatever the interpretation may be, two facts stand out from this study: mental load does have a negative effect on perception, and dissimilar mental images also have a negative effect, independently of mental load.

In a book she edited to gather cognitive approaches to the study of imagery, Segal (1971) reviewed all her experiments at that date, including a couple then unpublished. She tracked the cases where perceptual cues were incorporated in the mental images, and defended the theory that imagery always incorporates elements from perceptual signal or noise and *assimilates* them to give a mental image. According to Segal, the discrimination between imagery and perception is then a matter of *reality decision*, essentially statistical, based on both vividness and contextual cues. Our attractor model does allow imagery to incorporate elements from vision, when the possible hybrid representation is a good attractor. Yet, the kind of incorporation presented by Segal would not be low-level. In this regard, her theory would be best applied to non-pictorial imagery.

Reeves and Segal (1973), focusing on tighter methods, filmed the subjects'eyes during Perky experiments. Their fourteen subjects, four in high luminosity and ten in low luminosity, had to detect a visual symbol while doing the following tasks: no imagery, imagery, or mental arithmetic. Pupil diameter variations, which might have been systematical, were measured on film. No relation between pupil diameter and D' change were observed in the imagery nor in the mental arithmetics tasks. This rejected the hypothesis that this particular low-level influence of mental imagery could dim acuity and

thereby induce Perky effect (D' diminution). In this study D' are computed individually then averaged. The observed D' diminution is much lower than that observed in previous studies of Segal: 0.25 only, as opposed to about 0.7 usually. In particular, the false alarm rates were constant, and the diminution of D' reflected a lowering of the hit rates. Because the visual field of the subjects in this experiment was not totally occluded by an opaque hood, as it had been in previous experiments, yet more occluded than other studies entirely failing to demonstrate Perky effect, the authors suggested that ganzfeldt might be necessary to obtain Perky effect. Moreover, the Perky effect was negligible in high luminosity (.037 D' units). The authors concluded that the stricter controls used in this study argued in favor of the reality of the Perky effect. Once again, the imaging condition did not ask to image the stimuli, so that this study does not inform us with regard to our main question.

Schorr, Balzano and Smith (1978) tested a propositional account of the results of Segal and Fusella (1970). In their experiment, eight good imagers underwent two tasks, one of before-after dot detection, the other of above-below dot detection, while they imagined visually or auditorily images that had either the above-below or the before-after component. If the result of Segal and Fusella (1970) had been caused by propositional relation interference, one would expect intra-relationship interference as well as intra-modality interference. This was not the case: Segal and Fusella's results were replicated, and the propositional relationship led to facilitation rather than interference. Although they involved similar spatial relations, the mental images and stimuli were dissimilar; for example, the visual stimuli were made of superimposed dots, but the mental images had to be two colored half planes. Thus, in our view, both results of this study could be expected: dissimilar low-level appearance created inhibition, similar high-level resemblance yielded facilitation.

Reeves (1980) used a backward-masking paradigm, where a stimulus composed of one to four dots arranged on a square was made subliminal by a short projection followed by the presentation of a mask quickly afterwards. A first experiment was run to validate the paradigm. A second experiment compared the sensitivity to a particular dot location when subjects were imaging the four-dot mask or not. The sensitivity decreased in the imagery condition. This result, taken as a Perky effect, could also be interpreted

as interference of imagery with covert spatial attention: the subjects were asked to fixate in the center, but knew that the crucial dot was to appear at the lower left; hence the task is typically a covert attention task. A third and last experiment tested the detection of incongruent orientation in one of four dot pairs making up the masked stimulus. The main finding of this third experiment was that in the event where the image had the same dot alignment as the target, D' increased. The authors interpret that finding as the consequence of a decreased sensitivity to the majority of dot pairs, allowing easier detection of the occasional oddball. Needless to say, our simpler interpretation is that when the mental image was strictly identical to the stimulus, sensitivity increased.

Reeves (1981), in a within subject design using thirteen students, provided the cleanest evidence of classical Perky effect, along with the interesting result that it is dependent upon chromatic contrast. The subjects imaged either nothing, or the target (a triangle), or an unrelated object, in either a forced-choice or a yes/no detection task. The target could be filled with white, like the background, or colored in red. D' were computed subject per subject. The Perky effect appeared in both imaging conditions, according to both measures, but for the color opponency only, not for luminance opponency. So, the Perky effect would not affect detection of a white triangle on grey background, but it would affect detection of the red triangle on white background. The mean effect averaged 0.85 D' units. In his conclusion, Reeves suggested that imagery might interfere selectively with the parvocellular pathway. Of all results on the Perky effect, the color-induced difference observed in that study makes it the most difficult to account for with the priming theory. On one hand, interference rather than facilitation was observed even in the "image target" condition. On the other hand, most confound one could think of should apply, *mutatis mutandis*, to both the colored image and the white image. Thus, observing the Perky effect in one condition but not the other discards the interpretation that it could be an artifact of some yet unthought confound. However, it is possible, as pointed out by Farah (1983), that the chromatic imagery would be more demanding of attentional resources, and that the observed inhibition resulted from attentional load. Thus, adding the chromatic component takes up more of the subjects'attention and their response worsens then. It is also worth noting that, although the difference was not significant, imaging of the stimulus had less inhibition

than imaging something else. So, Reeves' result could be the superposition of worsening from mental load and true low-level, content-dependent interaction.

Craver-Lemley and Reeves (1987) presented several experiments in three parts, all using the same paradigm of measuring the detectability of the relative horizontal position of two vertical line segments almost aligned (the Vernier acuity). In this paradigm, the Perky effect is a diminishing of acuity when visualizing. They observed, first, that visual imagery reduces Vernier acuity by about 15%, and that this does not depend upon what is visualized, and that visualizing away from the center gives less Perky effect. Second, they measured how the Perky effect disappears slowly after imagery is disengaged, taking up to six seconds to totally disappear. Finally, they showed that the Perky effect appears stronger at longer stimulus durations (however, no statistics were given on this last effect). This study also included on some of the experiments very good controls against confounds such as eye movement and pupil diameter change, in particular by the use of an artificial pupil. The authors concluded that the basic low-level effect of imagery is reduction of sensitivity, and that the priming occasionally observed by others must be rooted at higher levels of the visual system. However, all the images used were different from the stimulus, thus setting the stage for inhibition according to all theories. Moreover, because the Vernier effect depends on using information from the whole length of the line, it is likely to be a non-local judgement, necessitating a fair amount of high-level processing and thereby opening the ground for many unwanted effects.

Craver-Lemley and Reeves (1992), after a long review, defend the hypothesis that the Perky effect is a "reduction of effective target energy". Using the same paradigm, they present a series of experiments following up on Craver-Lemley and Reeves (1987), and focused mostly on comparing the effects of spatial attention and imagery on Vernier acuity. Various experiments use situations of split attention, unequal division of attention, distraction with parafoveal stimuli, and auditory imagery. The results show that, generally, attention and imagery do not interact, except in one condition where visualization is centered. The authors concluded in support of effective energy reduction, maybe in parvocellular pathways, especially the LGN, which could be a defense mechanism against assimilation of mental images and visual images. From the point of view of our theory, similar gross results – a decrease of task performance– could be expected because

the pattern imaged were not the target; we believe that the fine results depend critically on whether subjects –ninety nine altogether– could visualize an image at one location and attend covertly at another location. We would expect that localized imagery would prevent or strongly hinder covert attention to another location, or vice-versa, depending on the subject's priorities. Unfortunately, performance on the attentional task in distract condition –to detect flicker frequency change– was not reported. In any case, there is an open possibility that subjects were incapable to follow the instructions in many conditions, and although this does not impair the overall finding of Vernier Perky effect, it hinders the interpretation of the detailed results.

Craver-Lemley, Arterberry and Reeves (1997) used again the Vernier acuity paradigm to test the influence on vision of the depth at which mental images were formed. Perspective line drawings were used to cue the subjects as to where in depth the mental image had to be formed. It was found that the presence of Perky effect depended on whether the image was formed in front of the stimulus plane. In contrast, real lines interfered only according to their lateral distance from the Vernier stimulus. Two main conclusions were drawn by the authors. First, the difference between the interference patterns observed with real and imaged lines argues against the equivalence of real and mental images. Second, that no present theory, including the energy-reduction theory which they previously favored, can account for the variations observed in the Perky effect for foveal stimuli. In our view, her results also strengthen the hypothesis that Vernier effect relies more on high-level processing than most perceptual judgements. If so, the low-level interaction could be lost in much stronger high-level effects.

In conclusion, the Perky effect was exciting because it appeared counter-intuitive: you couldn't tell you daydreams from reality? Not possible! Yet, the chase gradually focused on proving the existence of a low-level mechanism whereby imagery would inhibit perception. In doing so, it gave up the original meaning of Perky's study, the question of reality-testing. A posteriori, the observed interferences appear like a conglomerate of different effects. Sometimes global attentional resource is taken away, sometimes reality-testing is lowered, and, sometimes, imaging different objects than are perceived can hinder the perception processes and the response processes at all levels.

Psychophysical studies using imaged rather than visual context

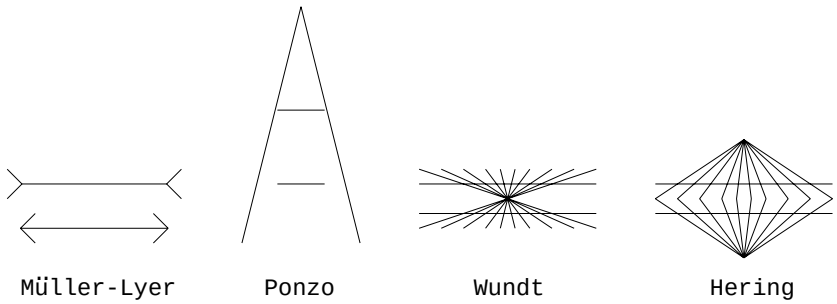

Figure IV.1: **Illusions replicated with an imaged context:** These various illusions are also obtained by good imagers when they visualize rather than see the illusion-inducing contexts: the arrows for the Müller-Lyer illusion, the wedge for the Ponzo illusion, and the hatch lines for the Wundt and Hering Figures.

The next set of studies that we will review look at effects of imagery on perception which are not necessarily simple facilitation or inhibition. In these studies, alterations of perception induced by visual *context* are replicated with more or less success using *imaged context*. Therefore, the effects observed depend on image content, although they are more subtle than the basic effects of gross facilitation/inhibition.

Singer and Sheehan (1965) warn us about demand characteristics in such paradigms. They studied the effect of imaging a slanted line onto the estimation of verticality, under different sets of instructions which manipulated the demand characteristics of the task. The same bias was introduced by *ad hoc* instructions for the imagery condition and a vision condition where the slanted line was shown. The results show a clear linear superposition of a main effect for vision but not for imagery, with a main effect for the instructions which manipulated the demand characteristics.

Over and Broerse (1972) compared the influence over the detection of a vertical line of vertical or horizontal gratings seen or visualized. Gratings seen masked the target line in a strongly orientation-dependent fashion. Imagery produced some masking compared to an orthogonal seen grating, but it was homogeneous across orientations.

The authors argued that compared to Segal and Fusella's study (1970), their stimulus was simpler, and that masking by imagery may come from a "feature-selective attention mechanism" rather than from low-level interaction.

Berbaum and Chung (1981) reproduced the famous Müller-Lyer illusion (Figure IV.1), where a line terminated by outwards-pointing arrowheads appears shorter than the same line terminated by inwards-pointing arrowheads, asking the subjects to visualize arrowheads on the end of plain line segments and report changes of apparent length. They found that nineteen out of the twenty naive participants reported changes in the expected direction.

Wallace (1984) asked high and low imagers, selected with the Vividness of Visual Imagery Questionnaire, to image different elements of the Ponzo, Wundt and Hering illusions (Figure IV.1). In the Ponzo illusion, where two line segments appear to have different length depending on their position in a two-line wedge, sixty-four subjects visualized none, one or both segments. In a second experiment, thirty-two subjects visualized the wedge while being shown the segments. In the Wundt or Hering illusions, where radiating lines crossing two parallel segments make them appear curved, forty participants had to image the radiating line while seeing the parallel segments. Wallace found that high imagers experienced these illusions with the equivalent magnitude in imagery and vision, whereas low imagers experienced the illusions only in the visual presentation.

Rhodes and O'Leary (1985) presented three experiments testing the interaction of imaging gratings onto detecting them. In the first experiment, eight subjects visualized gratings in one quadrant of the screen. The design factors were imagery/no imagery and location, the task was a forced choice between two sequential alternatives. The author used a staircase procedure, whereby stimulus duration was increased until the subjects detect it, then decreased until it is not seen any more, repeatedly, and the turning points are averaged to give an estimate of the perceptual threshold. Imagery raised the threshold to 2.48 ms of exposure, from 2.28 ms, but this increment happened indiscriminately in all quadrants. The second experiment was done showing a low-contrast grating as an anchor to imagery. When the grating had the same orientation as the target, it interfered at the target location only, but imaging that the grating had a higher contrast produced

no effect. In the third experiment, the sixteen subjects were selected good imagers, and the task was made easier by asking for centered imagery. The results show no effect of imagery any more. The authors argue from this that "internally generated images must be represented at higher levels of the visual system than the primary visual cortex". [2] Except for eidetic imagery, we also subscribe to this conclusion.

Ishai and Sagi (1995) compared the patterns of facilitation and interference by flanking masks seen and imaged. In their study on six subjects, measuring threshold change, they replicated the interference pattern of Gabor flanking masks onto a central Gabor stimulus, which is a interference at close distance and a facilitation at further distance. They also looked at the pattern created by similar masks imaged rather than seen, and found that the facilitation is observed at all distances, with the same asymptotical magnitude as that of the visual flanks. They further reported that the effect is monocular, and that a memory trace, rather that just spatial attention, was necessary to get the effect, because insufficient knowledge of the flank suppressed the effect.

In conclusion, some of these studies show identical effects of imaged context, and some do not. Typically, it appears that effects resulting from fatigue were not replicated with imagery. So, prolonged imaging of gratings or lines is not likely to have the same effects on subsequent perception that prolonged inspection of the same gratings or lines has. However, interference effects which depend on the immediate context are obtained with imaged context as well as with seen context. So, illusions which arise immediately because of proximate hatching or lines, as in Figure IV.1 will also arise if those hatching or lines are absent from the stimulus but simultaneously visualized. Because there are only a few studies, one cannot know if this pattern will generalize. However, it is clear that in about half of the studies the right mental image alters perception the same way as the right visual context. This is strongly in support of our hypothesis that the interaction of imagery and perception depends on the particular mental image.

[2] They draw further argument towards this conclusion from the fact that "electrical stimulation of PVC produces elementary sensations, such as flashes of light and colored spots, whereas stimulation of the secondary areas produces recognizable images of objects and familiar people." That argument, however, is specious, because such electrical stimulation cannot pretend to the degree of organization of top-down cortical connections.

Studies focused on priming or facilitation

The third and last group of studies attacks directly the question of content-dependent interaction between imagery and vision. As in the Perky effect studies, these studies use a mental image cue before a perceptual situation. However, they distinguish between cueing by a similar or a different image. Yet, none of these studies compares the results obtained with those from cueing with a visual image.

Peterson and Graham (1974) compared the performance of two groups of thirty subjects, one which listened to cues and the other who visualized from the cues, in a task of forced choice between sequential alternatives. The cues could be either compatible or incompatible with the target. Incompatible cueing lowered the hit rate in the imaging group only. Compatible cueing raised the hit rate from about fifty to eighty percent, in both imaging and control groups. The authors suggested that first these results falsify Segal's theory of imagery being assimilation of memory traces and perceptual input, and second that imagery appears to interfere negatively on perception when the image formed is incongruent with the perceptual signal, and that the verbal cues provide a priming effect with or without imagery. However, as pointed out by Farah (1985b), it is possible that their imaging task caused eye movements or particular attention to parts of the screen which could account for the effects. Thus, Peterson and Graham (1974) is strongly in support of our hypothesis, but lacks the decisive control condition of visual cueing.

Farah (1983) study facilitation and interference in the auditory domain. The thresholds for detecting signals at 750 Hz and 1000 Hz were studied under the conditions of simultaneous imaging of either tone, and just-before imaging of either tone. The rationale was to pull apart the effects of confusion, attention, and facilitation, and avoid confounds with eye movements, fixation or spatial attention. The threshold were measured by giving the subjects a forced choice between two intervals only one of which contained the signal, which had a varying intensity continuously adapting to the subjects performance. Their result show that there is signal-specific facilitation, and also that the attentional load under the simultaneous imaging condition also raised the detection threshold. They also discuss the absence of selective interference –the Perky effect analogue– which is surprising because many subjects reported confusions. To their ac-

count in terms of signal detection theory, we prefer the simpler explanation that the forced choice task they used gives no opportunity for a mere confusion to cause a false alarm, whereas the simpler detection task of Segal and Fusella (1970) leaves that possibility open. In conclusion, the Farah and Smith support the findings of Peterson and Graham (1974), and argue against the point that confusion causes selective interference.

Farah (1985b) followed up on her previous conclusions, and, coming back to the visual modality, focused on the content-specificity of imagery-vision interaction: "In order to show that imagery and perception share a set of representational structures, it is necessary to show that the interaction is representation-specific. That is, to rule out a capacity-based account of imagery-perception interactions, we need to show that the content of the image and percept, within a modality, determines the extent of their interaction". To explore this issue, she set up two experiments, one where the representational content is merely a shape, the other where it is a shape at a given location. In her first experiment, twelve subjects visualized either a capital 'H' or a 'T' and made a two-alternative forced-choice between a first and a second presentation where one stimulus contained the target, which was either H or T. She found a significant interaction, that visualizing H made the subject better at detecting H, and visualizing T facilitated the detection of T. In her second experiment, Farah asked the participants to either not visualize, or to visualize 'H' or 'T' in one of two preset boxes, the upper box or the lower box. The results showed that having the right letter at right place caused better performance than everything else. There was also a strong main effect for the location alone, showing that content-specific facilitation was not the sole component of imagery facilitation. Martha Farah concluded that imagery shares perceptual representation at the low level, the "array" level in Kosslyn's terms. Farah, after pointing out the confound of eye movements in Peterson's study, controls it by ordering subjects to maintain central fixation. This control cannot, however, avoid the possible influence of other imaged-based strategies of information acquisition. For example, the subjects could be paying attention more to the top of the stimulus if they were asked to visualize a "T" or more to the center when asked to visualize an "H". Thus, covert attention confounds are still present in this study, because it does not include a visual cueing control condition.

Addressing, in turn, the issue of criterion change versus sensitivity change,

Farah (1989a) presented four experiments which updated a paradigm due to Podgorny and Shepard (1978), where block letters are visualized on a 5x5 grid. The subjects had to detect dots falling in the grid. The first experiment replicated with a hit rate measure what Podgorny and Shepard (1978) found looking at reaction times: that detection was better on the cases covered by a visualized letter. In the second experiment the hit and false alarm rates were used to compute criterion and sensitivity, according to signal detection theory. Martha Farah found that the criterion was lowered, but that the sensitivity did not improve. In her third experiment, she replaced the imaged letter by a faint visual stimulus, the facilitation effects then disappeared. In the fourth experiment, the two letters were fused and the union was presented as a faint visual stimulus. The participants were told to see in that union one or the other letter; again, the criterion changed but not the sensitivity. Martha Farah concluded by arguing in favor of the visual array theory, and insists on the similarity between the effects of imagery and attention.

Although Martha Farah does not take up this point in her conclusion, several items along the way tend to suggest that imagery cannot increase sensitivity per se. We believe that this is a consequence of the setup. First, in the second experiment, ten out of twenty subjects had to be rejected because ceiling hit or false alarm rates prevented the computation of β and D'. Martha Farah argued: "Note that the elimination of data on the basis of this criterion does not bias the analysis towards any particular hypothesis concerning the effects of imagery on stimulus detection". This is true with respect to the criterion, but not for the sensitivity, since the subjects with ceiling hit rates are likely to be the subjects with infinite D' (in practice, higher D' too high to be sampled by the experimental conditions). Thus, the rejection of subjects with ceiling effects probably introduces a bias against observing a raise in D'. Second, if the influence of imagery on perception is content-specific, then the letter images could be the wrong stimulus to increase sensitivity to the dots. Kosslyn (1994) suggests that, because the visual pathways separate into encoding of object and location, top-down effects for both are to be expected. In Farah's experiment, co-location could permit the facilitation by the spatial top-down effect, but non-identity of object hinders the facilitation by the object top-down effect. Part of the possible confusion, we believe, stems from too literal an interpretation of the spatial array model of imagery: Farah writes that "If the image is

really a pattern of visual properties bound to particular locations in an internal array-format representation of the visual field, then the image should also affect the detection of probe dots falling in the region of the visual field that it covers [...] Furthermore, this should be true even though there is little similarity between probe dot stimuli and the imaged letters". This argument omits to remind us that wavelet banks, for example banks of Gabor filter, are localized and yet are an intermediate representation between the spatial domain and the frequency domain. In other terms, our everyday conception of the spatial array as a screen of NxP pixels is biasing the interpretation of what "spatial array" means. If we take into account this remark, the way that imagery can "affect the detection of probe dots" clearly becomes both a function of what is imaged, especially, the features orientation, and where it is imaged. Farah's experiment, centered on the "where", did not demonstrate that "what" effects cannot increase sensitivity. In conclusion, Martha Farah's paper demonstrated properly that imagery includes a part of spatial attention, but does not show that spatial attention is all there is to imagery. Rather, we agree with her 1985 statement about the content-specificity of facilitation, and her 1989 results should not be taken as evidence that imagery cannot improve perception.

Heil, Rössler and Hennighausen (1993) replicated and extended the study of Farah (1989a), first by adding a third letter in the design, then by varying the compactness of the images. The subpart of their results overlapping with Farah's comforted hers, but the additional results showed that the supplementary effect of imagery was all the stronger as the image was compact, arguing for a spatial attention effect. In addition, these authors suggested that the attentional effect may proceed from imagery-induced improvement in figure-ground separation, rather than from the interaction of activity patterns in a shared medium.

In conclusion about this last group of studies, we see that content-dependent interaction is shown, but not compared to the decisive control condition of visual cueing. Farah's later studies try to determine the more subtle problem of finding the balance point between facilitation for the right mental and inhibition for the wrong mental image, but in these she gives up the use of mental images similar to the stimuli and reverts to a paradigm more likely to make use of spatial attention than of genuine low-level synergy

between mental image and visual image. Besides the lack of visual control, another problem can be found with all of these studies: the set of mental images used as cues is always extremely reduced. This could entangle the effects of imagery proper with the effects of learning and fastidiousness. Our study remedies these two problems.

Conclusion of the review

The complex pattern which emerges from all those experiments is that several effects are probably at play.

- First, we can identify spatial attention effects, overt –as eye movement confounds– or covert. Even though spatial attention may be the genuine result of top-down mechanisms in the "where" visual pathways analog to those which, in the "what" pathway, reconstruct an image, we cannot legitimately call them imagery. However, until the synergy of spatial attention and object imagery is better studied, one should be wary of tasks involving simultaneously imaging and covert attention; quite possibly what goes on then is not the simultaneous engaging of two different mechanisms, but the attempt to use independently two processes where one of them (imagery) already uses the second (spatial attention). This would place heavy demands on the attentional mechanisms.

- Second, the question of whether there is a task-dependent, content-independent, low-level interference of imagery upon vision remains open. Craver-Lemley's (1992) arguments are strong, but it remains that most studies of pure interference have not provided images equal to the targets, so that the interference then observed could be accounted for by content-dependent mechanisms. The early studies and replications of the Perky Effect would argue in favor of such a content-independent interference mechanism, but, notwithstanding the individual problems with some of them, they could be accounted for by confusion between the visual and mental images; such confusion would not be a low-level mechanism, but rather a high-level, post-perceptual, classification problem. Taken together, these results force the existence of either confusion, or a low-level mechanism of the kind postulated by Catherine Craver-Lemley.

- Third, the issue of tasks demands is particularly acute in this field, because many paradigms depend on subjects reports.

- Finally, the evidence for content-dependent, low-level interaction is still contestable. The studies of Peterson and Graham (1974), Farah (1983), Farah (1985b) stand out as the best examples of content-dependent interaction of imagery and perception, but because, first, Perterson's study can be criticized on the grounds of imagery cueing subjects as to where to attend, second, Farah (1983) is an auditory rather than visual paradigm, and third, the later results of Farah (1985) appear from her later studies to result partly from spatial attention allocation, these three studies should be considered as the benchmark against which any new experiment is compared, rather than taken as conclusive.

IV.A.2 Goals of the experiment

In this context, our goal is to provide stronger evidence for the last type of effect, low-level content-specific interaction. The reason for this quest is that, as admirably stated by Martha Farah, this type of interaction puts the strongest theoretical constraints, and argues strongly in favor of a shared low level medium of representation for vision and imagery.

A second goal of this study is to feature an imagery task that stays clear of pure memory, as well as of extreme simplicity or extreme complexity. Whereas it is likely that many mechanisms of recall make use of some mechanisms of imagery, we seek to involve active constitution of mental images. Extreme simplicity, such as visualization of dots, runs the risk of confounds with visual attention, overt or covert. Visualization of symbols or letters involves altogether different risks, because when the item is overused as a symbol as it is easy for subjects to believe they are visualizing it when they only think very hard about it. In a previous experiment (unpublished), we found that subjects who had extensive exposure to a set of four symbols including a six branch asterisk would, in majority, draw an eight branch asterisk in the post-experimental debriefing (the same happened also to a less-used hexagon, which makes one regret the abandon of classical Greek as a basic study of humanities). Such anecdotal evidence suggests that the imagery cues have to be chosen carefully to stay away from stimuli which would

involve primarily some other mental mechanism. In adopting a clock which could be imaged from verbal instructions, we find a generic stimulus of reasonable complexity, which could be instantiated in many different particulars, thus avoiding effects of over-learning. Hence the choice of the type of target to be imaged: an analogic clock dial.

IV.B Methods

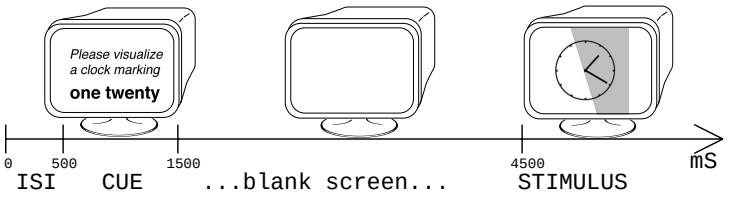

Figure IV.2: **Timeline of a trial:** A trial consists of a cue displayed for 1000ms, followed by the blank screen for 3 seconds, followed by the stimulus. Each of our 63 subjects went through 160 trials.

IV.B.1 Overview of the experiment

The experiment involved sixty-three subjects who each responded to a hundred and sixty trials. The basic trial was a sequence consisting of a cue, three seconds with a blank screen, and a stimulus (Figure IV.2). The cue was a clock picture or verbal instructions to image a clock (Figure IV.3). The blank screen stayed on for three seconds to give the subjects time to engage in the visualization. The stimulus was a clock dial covered by a half grey plane, and the subjects had to press a green button if both hands were in the grey plane, and a red button if one or both hands were out of the shade. We recorded reaction time to the button press, and analyzed mostly the difference between situations were the cue indicated the same time as the stimulus ("primed") and those where the times were different ("distracted"), all other factors being equal.

IV.B.2 Preliminary: Reasons for the choice of task and stimuli

The task was chosen to satisfy several requirements:

- First, we wanted to stay clear visualizing any over-practiced symbol, especially letters. Verbalization and symbolization has been blamed for suppressing imagery in several different situations, from the disappearance with aging of eidetic imagery to the difference in imagery between cultures, or between educated and non-educated members of our culture (Galton, 1907). The goal of this study is not to address these issues, and we felt it was wiser to require visualization of images which did not prompt quasi-automatically symbolic thinking.

- Second, we hoped to make the image to be visualized complicated enough to tap imagery rather than spatial attention, but not so complicated that visualization would be impossible for all but the most talented subjects. Although the role of spatial attention shaping cannot be precluded, it is likely that shaping the attentional window to the shape of two clock hands rather than in the shape of a dot or of one hand only demands much more resource. Indeed, if such a feat is possible, then spatial attention deserves the truer name of "spatial imagery". Had the cue to be imaged been too simple, such as a simple dot or bar, it is likely that subjects would have preferentially paid overt or covert attention to the relevant area.

  These first two constraints decided on the choice of the cue (see Figure IV.3).

- Third, the response had to be as visual as possible, excluding as much as possible verbal behavior and the possible interferences that could have occurred from it.

- Fourth, we wanted to avoid problems bound to subject report. This oriented the choice towards the measure of reaction time on a task that the subjects have to do with the dial.

- Fifth, we wanted to avoid any matching component in the task. We reasoned that if the cue gave any prior as to what the response should be, subjects might accomplish the task ahead of time, on the cue, then perform a simple match-to-cue to find out if the default response was sufficient. In the case the match failed, the subjects would then accomplish the task on the stimulus. Such a strategy might

develop even unconsciously, especially with a strong prior and when the task to be accomplished is harder or slower than matching. Even hybrid strategies, where part of the task is accomplished on the cue and gets validated by a cue-stimulus match, are likely to develop if the pay-off is positive. Priors allowing such a strategy to develop would have been catastrophic, because any content-dependent difference observed might then be solely the result of entirely different mental processes – "same" matching only v.s. "different" matching plus expected task. Thus, we strove to design a response that depended on the stimulus proper, and for which the cue would introduce no prior. Even though matching probably still occurred, due to the relationship between cue and stimulus, at least it was not directly involved in determining the response.

To prevent cue-induced prior on the response, the task depends on an additional item in the stimulus, which can be varied so that, given the clock ahead of time, the priors for the response are not changed. In our task, the additional item is a gray plane, and the task is to determine whether both hands of the clock fall in that plane. This task is also chosen with the intent of making it as perceptual as possible. We measure the reaction time on the response to that task, and analyze the facilitation (difference of reaction time between incongruent and congruent cue).

To handle the possibility that the interference observed might be a consequence of the subjects knowing where to look, we include a control condition where an analogue clock dial is seen rather than imaged, prior to seeing the stimulus. These controls are also meant to provide an identical semantic priming as the verbal cues from which the images had to be formed.

We also submitted the subjects to three pretests to assess their imaging capacity. We expected the visual cue control to show content-dependent interaction (facilitation for the congruent cues and/or slowing otherwise) because of priming, and our major expectations were that the pattern of results observed with a visual cue would be replicated with the imagery cue, at different magnitudes, decreasing as the subject's pretested measure imagery decreases.

These last two constraints oriented us towards adding the grey plane to the clock, with a random orientation, and making the task that of deciding where the hands

fell. (see Figure IV.4).

IV.B.3 Subjects

Sixty-three (63) subjects aged between 18 and 30 participated in the experiment. Twenty-six (26) were female and thirty-seven (37) were male. Only four subjects were left-handed, the remaining fifty-nine were right handed. All subjects but one were fluent English speakers. All subjects had normal or corrected-to-normal vision.

IV.B.4 Hardware, software, and physical setting

The experiment was run on a PowerMac MacIntosh, using the Superlab software. The subjects sat between 65 and 75 cm from the screen. The response was made By pressing one of two buttons made of two round pads of a diameter of 3cm glued onto the "e" and the "o" of a keyboard stripped of the other keys. The subjects ran in a closed experimental room, in the company of the experimenter. The level of illumination in the room was set prior to and following each task on the computer to accommodate the subject comfort.

The data, collected under Macintosh in Excel spreadsheets, was written on text files to be passed on Unix, where the computation of medians and means for each subject and condition was made by script files.

The ANOVA was run on PC using using the software "SPSS". For verification, simplified analyses, involving no more than one unbalanced factor were run both on PC and under Unix using the "anova" software.

IV.B.5 Stimuli

Each trial consisted of two events: the displaying of a cue, followed by that of the stimulus proper. The cue could be either a clock image, or a verbal time indication. The order of appearance of different cues was a factor of the design, explained below. The stimulus proper is the image of a clock half-covered by a transparent grey plane. Examples of stimuli can be seen in Figure IV.4.

The clocks themselves were made realistic in that the position of the hour hand between two marks of the circle depended on the minutes. The reason for this choice

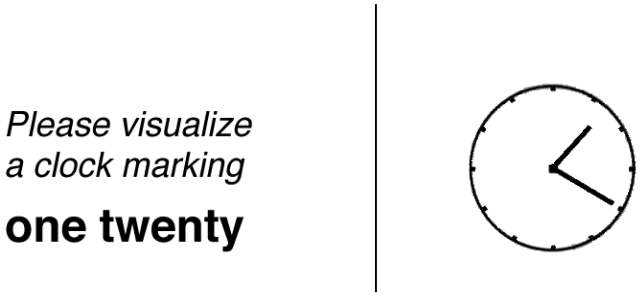

Figure IV.3: **CUE:** Cues appeared in two forms: either as verbal instructions (left) or as a visual clock image without grey plane (right). Both cues were black on white, about the same size, and appeared for 1000ms. The clock image was 52mm in diameter, and the time was displayed in a realistic manner in that the position of the hour hand depended on the minutes, as can be seen on this example there the hour hand is one third of the way between one o'clock and two o'clock. Time varied in 5 minute intervals; for example subjects never saw 5:37 or 7:19

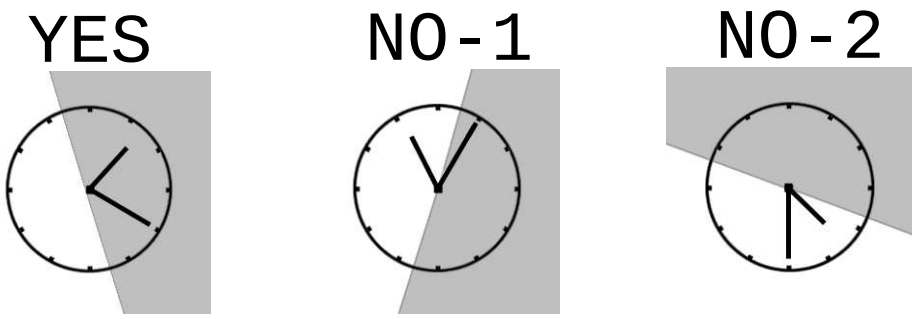

Figure IV.4: **STIMULI:** The clock image was similar to a cue clock image, but half-covered by a shaded plane coming in an unpredictable orientation. The subject is supposed to press the green button which we refer to as 'yes' if both hands of the clock are in the grey plane, and to press the red button ('no') otherwise. At the left we have a 'YES' stimulus. At the right we see two examples of a 'NO' stimulus, where at least one of the hands is at outside the grey plane. In the text, we refer to the case where one hand is outside as **NO-1** , and to the case where both hands are outside as **NO-2**

is that is had been found in pilot studies that, although most subjects initially tend to visualize a simplified version of the clock, namely, with the hour hand on the indicated hour and the minute hand on the indicated minute, some subjects would form the mental image corresponding to a real clock. In addition, some subjects could start with the right visualization and, because of certain clocks such as those showing 4:30 or 7:30, realize their mistake and change their imagery strategy afterwards. To prevent the loss of power which would occur should subjects visualize the wrong clock, it was decided to adopt the most realistic images, to sensitize the subject to this problem before the experiment, thanks to a pretest also using analogue clock visualization, and to use verbal rather than numerical descriptions of the clock times to be visualized, where the time would be specified as "after" or "till" the hour, to remind the subject about the continuum of the position of the hour hand.

The half plane covering the clocks did not extend throughout the screen, but rather was contained within a 77mm square, about half as large as the clock dial itself, which had a diameter of 52mm. The plane thus defined was a light gray shade, quite conspicuous but not at all an impediment to seeing the clock.

IV.B.6 Task, response and dependent variable

The subjects attended the stimulus with their hand ready between a green and a red button. The task was to press the green button if both hands were in the grey, otherwise to press the red button. We measured reaction time and distinguished errors. Subjects were asked to use the same hand regardless of the button they were pushing, and to place their hand between the buttons during the interstimulus interval. About half the subjects were instructed to use their right hand and the other half their left hand. The position of the buttons was the same for all subjects: on the right for the button corresponding to two hands in the grey, on the left for the button corresponding to one or both hands in the white.

To test the effect of imagery, each trial started with a cue. The cue was either a clock image without any grey plane, or verbal instructions to visualize the time (see examples in Figure IV.3). The task depended on the cue type. When seeing the clock image, subject had to "pay attention" to it. When seeing the instructions, they had to

visualize the corresponding clock.

Because of the simplicity of the task, and because our instructions did not emphasize speed, we did not expect error rates to be very high. Indeed, the results confirm our prediction, as most subjects committed no error throughout the hundred and sixty trials of the experiment. The overall error rate was only 2.3% of trials. On this basis, we feel that error rate data can be discarded with confidence, without biasing the study.

IV.B.7 Factors

See Figure IV.5 for a summary of factors. The experiment used both within subject and between subjects factors. Between subject factors can be further divided into assigned and observed factors.

There were three within subjects factors: the cue type (image, instructions), the expected response (yes or no), and the priming situation (primed or distracted). The cue type could be either a clock image – a "vision" cue–, without the shaded plane, or a verbal indication from which the subject had to visualize the clock image – an "imagery" cue. The expected response was called a 'YES' if if both hands of the clock were covered by the shaded plane, and a 'NO' if at least one hand fell outside the grey . The priming situation was that the subject was "primed" if the stimulus indicated the same time as the cue, and "distracted" if the stimulus and cue indicated different times. The full range of stimuli resulting from these factors is illustrated in Figure IV.6, where each of the eight types of trials is illustrated by an example.

Before proceeding further, a note should be made that the "yes" and "no" labels were assigned a posteriori for analysis. As can be seen by reading the instructions in Appendixes 4-10, careful precautions were taken to prevent the subjects from using such verbal labels. In fact, pilot studies showed that when subjects cast the task in terms of a verbal yes/no question, they would tend to automatically assign a "yes" to the "primed" situation, and a "no" to the "distracted" situation, and the response to this automatic matching was interfering with the task. In particular, with a "vision" cue (clock image), a strong delay for "primed, no" and "distracted, yes" was superimposed with the facilitation of "primed". We were not so worried at having such supplementary

effects, which the statistical analysis would have isolated, as at involving mental processes including matching and response interference. As it came, in our final data, the interference pattern observed in those pilot studies did not show up, which we take as a confirmation that the instructions correctly manipulated the task demands to prevent the subjects from phrasing the task explicitly as a yes/no question.

There were two between subject assigned factors: the task order and the hand used for pressing the button. The task order took three values. In two situations, the trials were blocked by similar cue, and the order was varied (visual cue first or verbal cue first). In the third situation, the trials were interleaved. Because the subjects complained about the difficulty of the interleaved task, and because their overall much higher RT in that situation confirmed that their short term memory was taxed beyond comfort by the complexity of that task, that condition was only applied to nine subjects. The hand which was used for pressing the buttons was evenly varied between right hand and left hand.

We expected only two between subject observed factors: gender and handedness, but due to the extremely small number of left-handers who signed up for the study, we forsook the handedness factor. Instead, observation of the results as the experiment progressed revealed another factor related to subject fatigue. A little over half the subjects ran the experiment during the exams week. As it was found that subjects were getting worse and worse during the first days of that week, we decided to make the fatigue level a factor of the analysis, by separating subjects run during the exam week from all others. We called this third factor "week", and the subject could belong either to the "term" or to the "exams week".

IV.B.8 Blocking and order of trials

Each subject underwent one hundred and sixty trials, which makes twenty in each basic cell of the design (identical cue type, expected response, and priming situation). For most subjects, the experiment went in two large blocks of eighty trials, separated by a five minute break, and interrupted by a break of user-chosen duration after the fiftieth trial. The nature of the cue was constant during each of these blocks. Within each block, the order of the trials was determined as follows: the trials were grouped by

```
WITHIN SUBJECT
        Cue         (visual/imaged)
        Response    (yes/no)
        Priming     (Primed/Distracted )

BETWEEN SUBJECTS
 assigned
        Hand used (right/left)
        Order     (v-i, i-v, interleaved)
 observed
        Gender    (male/female)
        Fatigue   (Term/Exams week)
        (Handedness)
```

Figure IV.5: **Factors of the experiment:** Factors not in the analysis appear in parenthesis: priming (implicit, because the facilitation is the difference measured across that factor), and handedness, because there were too few left-handed subjects.

subblocks of eight trials (two of each basic cell of response and priming) randomly chosen, then the trials were randomized within each subblock. In addition, the first block had the pattern ABCDDCBA. The reason for this process was that the software did not allow a different randomization for each subject, and since it was expected that the subjects would have a learning curve, we wanted to minimized the bias that could be introduced by having imbalanced temporal placement of the different conditions. For nine subjects, the trials were interleaved, so that these subjects ran one large block with three breaks of user-chosen duration. The trials were taken from each block of the non-interleaved design, in the same order, and placed one after the other.

IV.B.9 Pretests

Before the experiment, the subjects were given three pretests inspired by Craik and Dirkx (1992). The pretests are given in appendix.

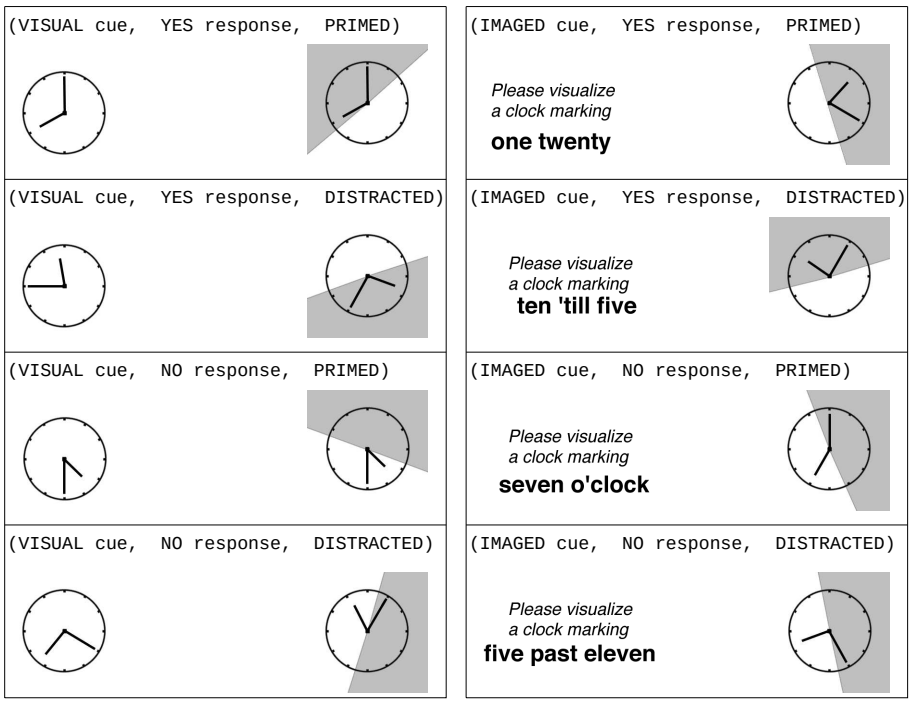

Figure IV.6: **Different cases of trials:** Taking into account all variations on the three within-subject factors, there are eight cases of trials, shown here.

- The first pretest asked the subjects to visualize clocks given a digital time and to decide whether the hands marked an angle wider or narrower than a right angle. For example, 12:10 was narrow but 1:30 was wide. The subjects had two minutes to complete as many trials as possible.

- The second pretest asked the subjects to use their imagery to keep track of their North/East/South/West orientation in a natural setting of their choice, as they mentally followed a sequence of movement directives. For example if they read the string "E L L L R" they had to visualize facing East initially, then turn left three times , then turn right once, and write down the direction they were now facing (West). After reading the instructions and practicing on a few examples, subjects had to complete as many trials as possible in three minutes.

- In the third pretest, adapted from Brooks (1968), subjects had to follow mentally the outline of a letter, reporting the direction of the turns they had to take doing so. The letter, "F", had been shown at the very beginning of the session, along with three other letters (L, T, E), and its shape was plain and sans seriffs. When they were able to do so, their time around the letter constituted the score on the pretest, otherwise they were considered as non-imagers according to that test.

IV.B.10 Interview

After the experiment and before their results were presented to the subject, the subject was interviewed about his impression of the experiment. Although the interview varied in content and form, we always asked the subject what they thought their strategy was, and especially if they were paying more attention to some aspect of the stimulus or to another. For the great majority of subjects, their response allowed them to be classified as testing the hypotheses (1) that 'the hands were in the grey' or (2) that 'at least one hand was in the white'. Because not all subjects could be conclusively classified, and because the interview was informal, the classification based on the interpretation of the answers, and the two resulting classes very unbalanced in numbers, we did not include estimated subject strategy as a factor. However, where indicated, we performed parallel analyses on the minority group (hypothesis 2) to verify that the subject strategy

was not the source of the observed effect.

IV.B.11 Timeline of experiment

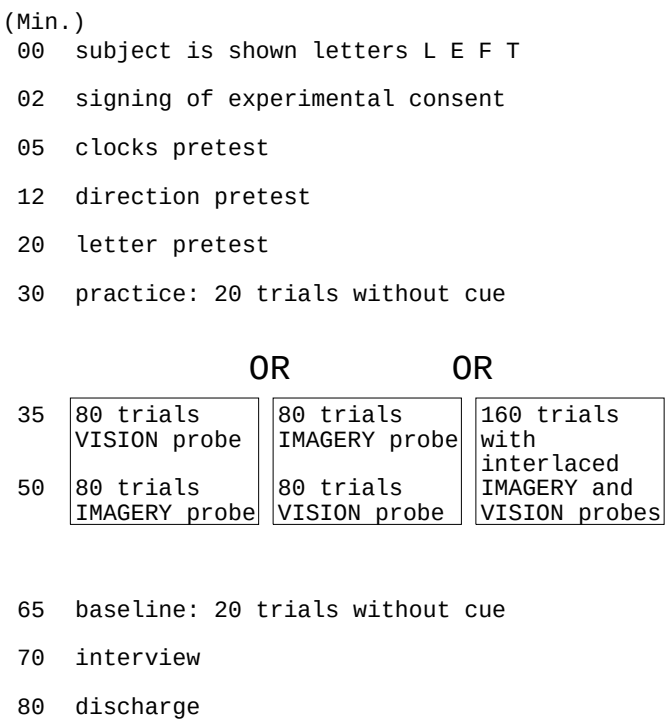

Figure IV.7: **Experimental schedule:** This schedule summarizes the various steps taken for each subject.

The schedule of the experiment is summarized in Figure IV.7. The subjects stayed for about an hour and a half, of which about thirty minutes was dedicated to the actual experimental task. Upon their entrance, the subject was first shown the letters to be used in the letter pretest, then given a brief description of the coming events and signed the disclaimer. After this, the subject took the three pretests. The subject was then familiarized with the stimulus and the response by a practice of twenty trials with no cue. Following this practice, the subject took the experiment itself, that is, either two blocks of eighty trials each, with a break within each block after the fiftieth trial and a break in between the blocks, or, when the stimuli of different cue type were interleaved, one block of a hundred and sixty trials, with three breaks at regular intervals within the

block. The duration of the breaks was user-controlled. Then, the subject was run on the same block of 20 trials without cue, to provide a lower bound of their baseline reaction time. Finally, we conducted the experimental interview, and discharged the subject.

IV.B.12 Facilitation was the analyzed variable

Because the base reaction time varied so much with the subject, and in a very predictable fashion with the task, we used a difference of RT as the analyzed variable. For each subject we obtained a global RT for each of the basic eight cells, then subtracted the "primed" RT from the "distracted" RT, to get the effect in each of the four conditions of cue and response. To get the global RT, we used for separate analyses the median, the raw mean, and the mean of RTs within two standard deviations of the raw mean ("pruned mean"). We will in the result section report the p-values in the "pruned means" analysis; the effects themselves are observed on all three scores. For most results the median and pruned mean criteria selected the same effects as significant.

IV.C Results

IV.C.1 Preliminary results: pretests, base RTs, error rates

Irrelevance of the pretests

Pretests bear no relation to facilitation After agonizing over the right pretests of imagery, it came as a surprise to us to discover that pretests scores do not predict in any way the global facilitation, or the superior facilitation of imagery over vision. The two following tables illustrate that, if anything, the bad scorers have a small, non significant advantage.

The complete ANOVAs reveal that none of the effects reported in the above table even approach significance. Looking at the interactions which come out to be significant, one finds first a response X fatigue X clock-test interaction, significant on the pruned means, $F(61,1)=5.8$, $p=.019$, but the interaction itself is different on the medians, which suggests that it is a false effect on the means.

Second, the interaction response X cue X clock-test approaches significance in both analysis, ; but the effect, if any, appears to be is that NO's give a stronger facilitation

test:	Clocks	East-West	Letter
statistics:			
medians	-5	-7	+2
pruned means	-3	-7	+3
all means	+5	+2	+9

Table IV.1: **Advantage of being a good tester for overall facilitation:** how much more facilitations good testers get is shown as a function of the test with the three possible statistics.

test:	Clocks	East-West	Letter
statistics:			
medians	+10	+4	-6
pruned means	+4	-7	-8
all means	-17	-13	-21

Table IV.2: **Advantage of good testers specifically for imagery:** the difference of facilitation between visual and imagery cue is a good index of the imagery-specific effect. This table shows the difference between good and bad testers on this index. Because the difference appears random, the tests probably do not select properly the subjects with the highest imagery effect.

for bad imagers according to the test. This effect appears to be due mostly to the imagery condition, but again, even if it were a real effect, it would argue against the hypothesis that the clock pretest is an index of imagery facilitation, since the boost affects the 'bad' imagers. Finally, another interaction involves the letter test: handused X letter-test, $F(61,1)=4.38$, $p=.041$, but again the interaction, also seen on the medians is that the facilitation for the right hand appears to be much superior to that on the left hand mostly for the 'bad' imagers according to the test. In conclusion, it is possible that these tests indicate something about the subject's ability to accomplish the task, but they are not principally indexes of imagery, as it is revealed by the facilitation observed in our task.

Pretests indicate global performance In contrast, these tests appear to index more strongly a simpler capacity of global performance. An analysis of the raw reaction times scores showed that the good test takers are faster than the bad test takers. The effects are of the order of 100 ms for the clock pretest, of 40 ms for the EW pretest, and about 20 ms for the letter pretest. Of those, only the clock pretest main effect is significant, $F(61,1)=6.252$, $p=.015$. One might have thought that the bad imagers according to the pretests are overall slower because they are especially slow with imaged cues. Yet, this is not true, because the effect of the cue type is about the same for good or bad test takers. Thus, what the pretests really seem to indicate is how fast the subject performs, regardless of the task given.

As it is likely that overall subject performance decreases with fatigue, and it appears (see below) that fatigue influences strongly the accomplishment of visualization, it is quite possible that the pretests indirectly measure something correlated with the subject's temporary visualization ability. However, there is no guarantee that this alertness-induced increase of ability corresponds at all to the qualitative differences that these pretests are attempting to sort.

The correlation between pretests replicates previous studies Finally, if indeed these pretests measure mostly subject alertness, it is logical that they would correlate to some degree, thus explaining the results of Craik and Dirkx (1992). And this is what we found. Because many subjects totally failed the letter test and were assigned an arbitrary

low score, we could not rely on the actual test scores to compute a linear correlation coefficient for each par of tests. However, once the subjects were divided into groups of bad and good scorers for each test, the overlap of the groups between tests appeared strong; reassigning *a posteriori* a score of -1 to the bad scorers and +1 to the good scorers, it is possible to compute the following coefficients, which mark the correlation of the groups rather than of the underlying measure:

test:	Clocks	East-West	Letter
test:			
East-West	.20		
Letter	-.04	.37	

Table IV.3: **Relation between pretests scores:** The groups of good test scorers overlap slightly, indicating a weak correlation between Letter test scores and East-West test scores. This replicates the results of Craik and Dirx (1992)

Because of our reassignment, the sign convention is different from that of Craik and Dirkx (1992), but our result replicate theirs, that the best correlation –for young subjects– to be obtained between these three tests is between the letter and the east-west pretests. In a last attempt to find out if what these two tests share might be imagery-related, we looked at how facilitation in vision and imagery fluctuates depending on what subgrouping of test results the subjects belong to. In the table just below, the first column indicates the east-west pretest result, the second column the letter pretest result, the third column the cue, the fourth column the number of subjects in each subgroup), and the fifth column the facilitation score. If the imagery facilitation depended on the common quantity measured by these two tests, there should be a stronger gap between vision and imagery facilitation for the (good, good) than for the (bad, bad). Clearly, this is not the case.

Thus, even though these two pretests are less correlated with raw overall reaction time than the clock pretest is, thus being less suspect of being tests of mere alertness, and even though they correlate one with the other, whatever they measure is not what causes a stronger facilitation for an imagery cue.

subjects		cue type	facilitation
good scorers	(21ss)	vision	13.6145
		imagery	14.6944
bad scorers	(22ss)	vision	11.2158
		imagery	23.0457

Table IV.4: **Facilitation scores for good and bad scorers:** The subjects with similar success in East-West and Letter test could give an idea of what this 'common factor' measures. Obviously it is not imagery, because the 'bad scorers' get more facilitation for imagery.

Conclusion on pretests Because the pretests were so unrelated to the facilitation or inhibition observed, we did not use them any further in the analysis.

A word about basic RTs

Nothing much surprising came out of the analysis of the reaction times themselves (see table IV.5) and of the error rates (see table IV.6). We basically confirmed the findings of others concerning the effects of mental load, of response type and of gender. Because the mental load for imagery is greater than for vision, the overall response is slower for imagery than for vision. Subjects generally respond faster for the YES conditions than for the corresponding NO conditions. Males show a bigger difference in reaction time between right and left hand response than females.

The effect of cue type appear related to the mental load. Overall, trials with a visual cue were responded to much faster than trials with an imagery cue (116 mSec effect). However, that difference almost disappears when trials with different cue type are interleaved rather than blocked by cue type[3]. Oddly enough, interleaving appears not so much to slow down the visually cued trials as to fasten the imagery cued trials. Seeing this, we looked at the error rates to find if the fastening was made up by more errors, but what we found was that the error rate pattern confirmed the RT finding. Errors rates for the subjects responding to blocked trials are 1.5% with visual cues and 3% with

[3]And yet the difference of facilitation between imaged and visual cues is stronger when trials were interleaved than for the overall population (47 mSec).

Condition	RT, mSec	F factor	p value
Grand Mean	811		
vision	753	F(61,1)=22.6	p<.001
imagery	869		
vision, coming in second block	772		
imagery, coming in first block	912		
vision, coming in first block	728	F(60,2)= 3.8	p=.033
imagery, coming in second block	852		
vision, interleaved	770		
imagery, interleaved	790		
yes	777	F(61,1)=41.3	p<.001
no	845		
'looking in the grey'	871	F(61,1)=4.05	p=.052
'looking in the white'	821		
yes, 'looking in grey'	784		
no, 'looking in grey'	857	F(61,1)=3.8	p=.06
yes, 'looking in white'	746		
no, 'looking in white'	794		
Females, NO, left hand	866		
Females, NO, right hand	843		
Females, YES, left hand	784		
Females, YES, right hand	786	F(61,1)= 13.2	p=.001
Males, NO, left hand	788		
Males, NO, right hand	890		
Males, YES, left hand	741		
Males, YES, right hand	801		

Table IV.5: **Effects observed on basic reaction times:** Each subgroup illustrates a different effect. Reaction times, in mSec, are averages after pruning outstanding trials. Trials with visual cues are notably faster than with imaged cues (first group), but this effect almost disappears when trials are interleaved (second group). The order effects (second group) are best explained by considering that imagery trials are harder and generate more fatigue than learning, but the vision trials, easier, cause only learning. A 'YES' response is faster than a 'NO' response (group 3). This is true even when the strategy is to look at the white (testing the 'NO' hypothesis first) (group 5). Strategy effect (group 4) and interaction with the YES/NO response type (group 5) are marginally significant. The last group (group 6) is an interaction between gender, response and hand used, and shows that males are more strongly lateralized than females. Also, note the absence of any effect directly related to the term vs. exams factor.

Condition	no response	response error	sum	%age
overall	84	143	227	2.3%
yes	13	65	78	1.6%
no	71	78	149	3.0%
vision	28	52	80	1.6%
imagery	56	91	147	2.9%
primed	25	65	90	1.8%
distracted	59	78	137	2.7%
vision primed	9	26	35	1.4%
vision, distracted	19	26	45	1.8%
imagery, primed	16	39	55	2.2%
imagery, distracted	40	52	92	3.7%
vision, yes, primed	2	11	13	1.1%
vision, yes, distracted	2	10	12	.95%
vision, no, primed	7	15	22	1.7%
vision, no, distracted	17	16	33	2.6%
imagery, yes, primed	5	20	25	2.1%
imagery, yes, distracted	4	24	28	2.2%
imagery, no, primed	11	19	30	2.4%
imagery, no, distracted	36	18	64	5.1%

Table IV.6: **Error rates:** We made no statistical analysis on the error rates but they confirm the reaction time data. Different groupings are listed in increasing detail.

imagery cues. But when the trial are interleaved, the error rates become respectively 1.9% and 2.2%. Thus, interleaving the imagery trials with the vision trials make the subjects better at those imagery trials. The only possible explanation is in terms of fatigue and mental load. Imagery is a more demanding task. When imaging comes only every other trial, it is easier, and the subject, more at ease, responds overall faster and with fewer errors. The same reason explains why order effects for the blocked cue types seem to contradict learning effects. When all vision trials, blocked together, are taken first, the subjects are less fatigued, and the responses overall faster. When they are taken second, the negative effect of fatigue from having just done the imagery block is stronger than the positive effect of learning, so that the vision trials are a bit slower (50 mSec). In contrast, because the vision trials do not fatigue the subjects as much, imagery trials coming second benefit only from learning, and are therefore faster than imagery trials coming first. Thus, it is clear that the imagery task, per se, makes the response more difficult, independently of the image content.

The 'YES' response was on average faster than the 'NO' response, by about 68 mSec. Little can be concluded from this because the green (YES) button was always to the right of the red (NO) button. Thus this effect could come from purely physical factors. Considering its size, however, it is likely not to. The fact that 'NO' came as a disjunction of cases probably contributes to this effect. [4] Taking into account the subject's reported strategy supports this interpretation.

There is a main effect of strategy (50 mSec) actually in favor of the minority of subjects who chose to turn the instructions around and look first for hands in the white. Perhaps it is only because these subjects (a minority of twelve) were more awake. But there is also a tendency to have less advantage of YES over NO for these subjects (52 mSec against 73 mSec). This marginally significant effect (p=.06) confirms the subjects'reports. The effect of the response is also paralleled by the error rates, with the YESes having fewer errors than the NOs.

The third effect is an interaction of gender, hand used and response. It shows

[4]Later, we consider the difference between NO-1 trials, where only one hand is in the white, and NO-2 trials, where both hands are in the white. There is very little difference between the basic reaction times for the two types of NOs. The strong advantage of the YES over the NO (68 mSec) dwarves the advantage of NO-2 over NO-1 (12 mSec).

that for males the responses with the right hand are much slower than those with the left hand, for all answers, but more for NOs (92 mSec) than for YESes (60 mSec), whereas the females have very little effect of hand used. [5] Thus, this result is one more speck on top of the body of evidence that males are more lateralized than females.

A word about error rates

Besides the points made above, the error rates generally confirm the reaction times data. Throughout the reaction time results obtained, the slower the response, the higher the error rate. Error rates can be used to compute a second variable measuring facilitation: Errors(distracted) - Errors(primed). Although we did not do a full factorial analysis on this index, it follows the same pattern as the reaction times pattern at least at the gross, most reliable level of within-subject factors. Table IV.6 presents the detail or error counts and rates, and several different groupings which will make this point clear.

In conclusion to this preliminary section, no excessive surprise has found us. The pretests are irrelevant to our main question. The basic reaction times hide nothing more than common sense effects: the higher difficulty of imaging compared to being passive, a greater facility to press the YES button, a higher lateralization for males than for females. Finally, as expected in a situation where instructions emphasized neither speed nor accuracy, error rates confirm the reaction time pattern, rather than compensate for it.

Thus, with these preliminary considerations behind us, we can now look at the variable which is most relevant for our question: the facilitation, given by the difference in reaction time between situations equal in all respect, except that one is primed (had the right cue) and the other is distracted (had a wrong cue).

IV.C.2 Main result: imagery facilitates the task even more than vision

Overall, we observed significant facilitation in the "primed" situation, $F(1,62)=8.176$, $p=.007$.

[5] Ignoring the response factor, the gender by response interaction alone is not very far from significance, $p=.09$).

The main result of this experiment is strongly counterintuitive: as shown in Figure IV.8, imagery produces more facilitation than vision, $F(1,62)=4.565$, $p=.038$.

It is to be observed that if the data are analyzed in two separate ANOVAs, one for the "vision" data and the other for the "imagery" data, the main effect is only significant in the latter ($p=.005$ against .317 for vision). Because our point is to compare imagery with vision in their induction of facilitation, we do not need to answer the question of whether there a real effect of priming by vision. Whether or not such priming affects the task, it is superseeded in magnitude by effects proper to imagery.

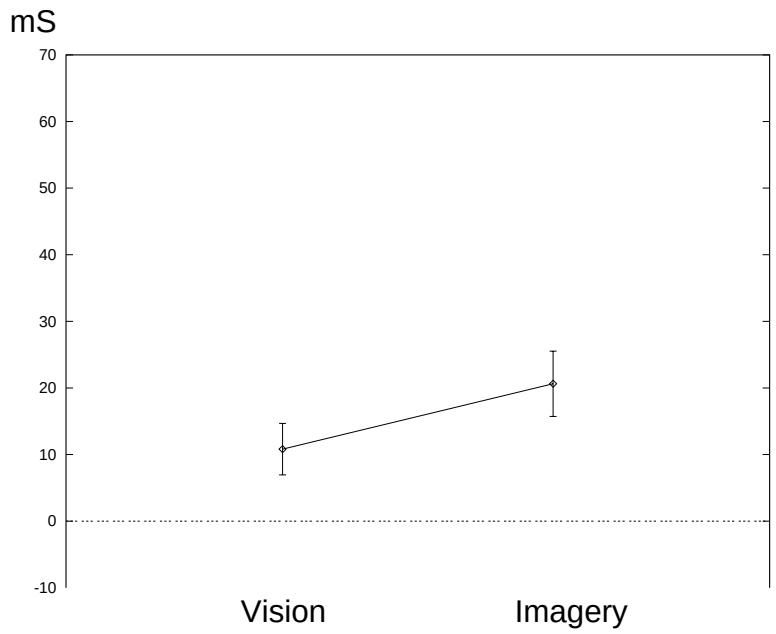

Figure IV.8: **Effect of the cue type:** contrary to common sense, imagery cues facilitate judgement even more than visual cues. Points represent the difference between primed and distracted mean reaction times; vertical lines depict standard errors of the means.

IV.C.3 Fatigued subjects have less facilitation

One important qualification on our experiment was that subjects who participated during exams week showed significantly lower facilitation than other subjects, $F(1,62)=5.385$, $p=.025$.

Although the interaction between week and cue type is not significant (p=.325), it can be observed that most this main effect comes from a change in the imagery condition (facilitation drops from 30ms to 12ms), whereas the vision facilitation actually increases (gain from 8 to 13 ms). The reason why the interaction is not significant despite the facts that is is a cross-over and that one leg alone accounts for a significant main effect is that the design is not orthogonal; in particular, order effects are not equally balanced between the two "week" conditions, firstly because we did not expect *a priori* that the week would be an important factor, and second because the "interlaced" order was abandoned early, so that most subjects in that condition are in the first, non-fatigued week. If the "interlaced" subjects are removed, the interaction remains a cross-over in the same direction, but is much weaker (10ms instead of 23ms). Because this interaction is not significant, it can be questioned, nevertheless its presence is consistent with a common sense refinement of the interpretation of the main effect of fatigue, that is, that fatigue will not affect the facilitation *per se*, but rather the ability to create, or sustain, the mental image which causes that facilitation.

IV.C.4 Gender effects

Gender did not appear to be of much influence. The main effect of gender, and the interaction of gender with cue, have ridiculous levels of significance (resp. p=.5 and .9). One three-way interaction involving gender, shown in Figure IV.11, comes out to be just significant, $F(63,1)=4.22$, p=.046. Possibly, this effect indicates a tendency for males to have a stronger hemispheric preference for forming composite images with the left hemisphere, but more evidence would be welcome to support this point.

IV.C.5 Response effects

Because we had set up the task to be as visual as possible, and because our dependent variable is a facilitation, subtracting the time taken by similar processes, we were extremely surprised to find strong response effects. We will present all effects which are significant, but suggest that they all stem from the basic level interaction between cue type, hand used and response, which is the last effect reported.

The first effect, presented in Figure IV.12, is that there is overall more facilita-

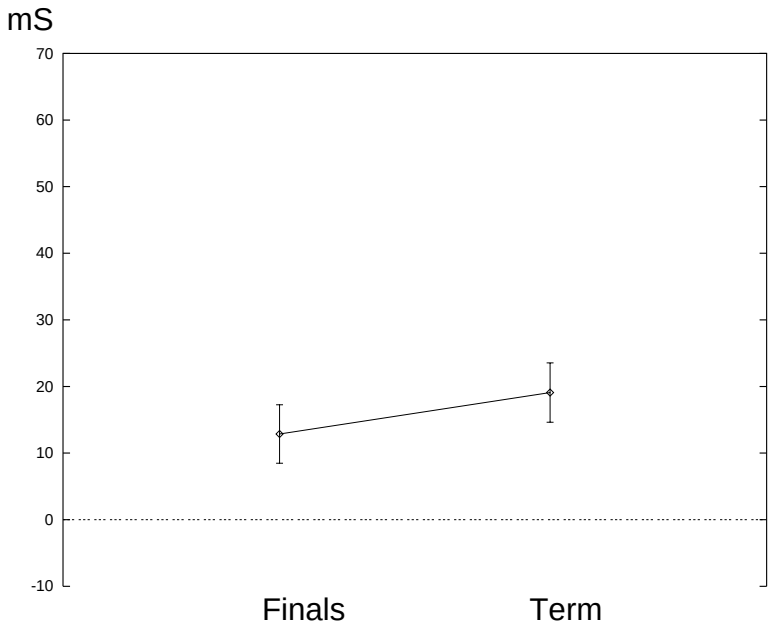

Figure IV.9: **Effect of fatigue:** subject run during the examination week show significantly less facilitation. Points represent the difference between primed and distracted mean reaction times; vertical lines depict standard errors of the means.

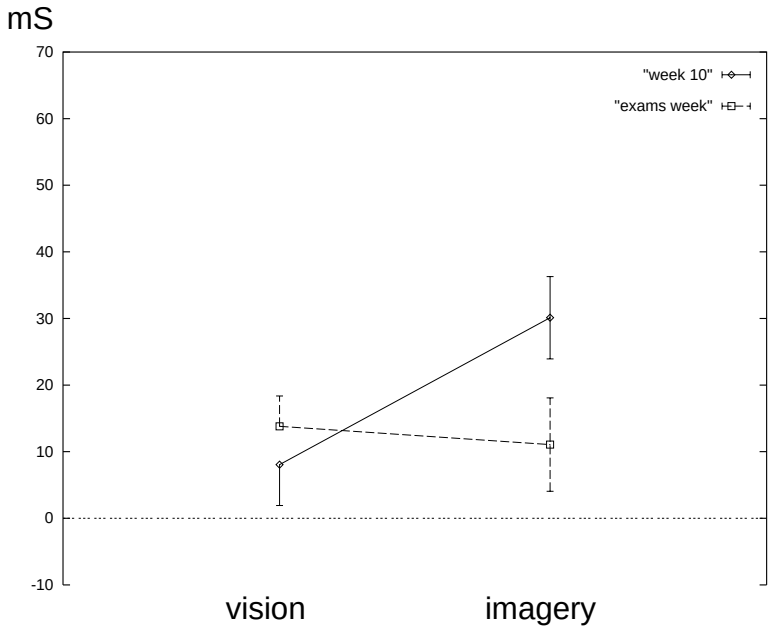

Figure IV.10: **Effect of fatigue split by cue type:** Although this interaction of the week with the cue type is not significant, it supports the common sense idea that fatigue (and lower motivation) reduces a subject's ability to visualize and hence to benefit from facilitation from imaged cues. Points represent the difference between primed and distracted mean reaction times; vertical lines depict standard errors of the means.

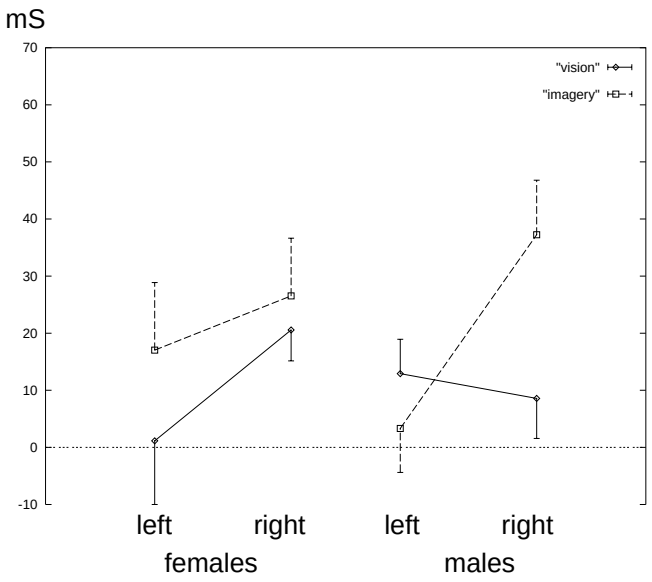

Figure IV.11: **Interaction of gender, cue type and hand used:** The only effect involving gender is this interaction between gender, cue, and hand used. Male imagery is more sensitive to the hand used for response. Points represent the difference between primed and distracted mean reaction times; vertical lines depict standard errors of the means.

tion on 'NO' responses than on 'YES' responses, F(1,63)=4.18, p=.047.

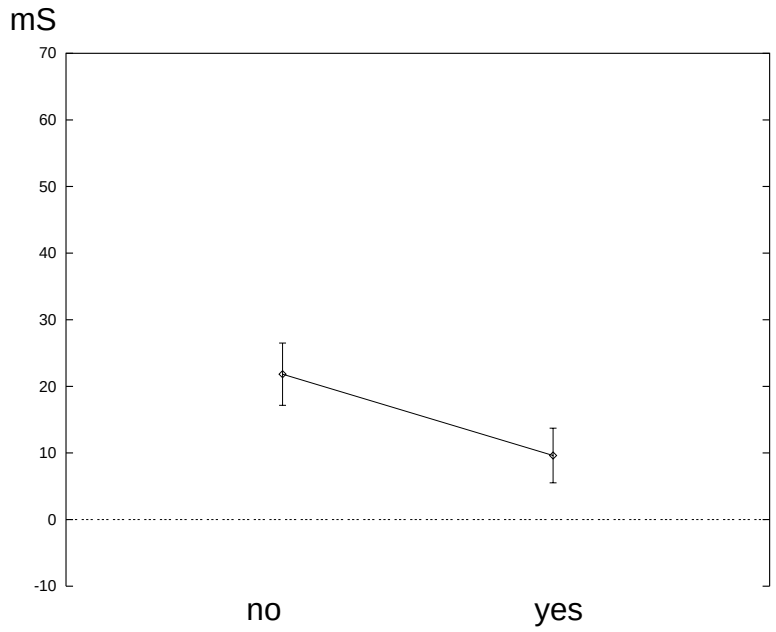

Figure IV.12: **Effect of response:** There is more facilitation on NOs than on YESes. We speculate that this is due in part to eye movement preparation in the NO-1 condition. Figure 20 will illustrate that point. Points represent the difference between primed and distracted mean reaction times; vertical lines depict standard errors of the means.

The second effect involving response is that facilitation was higher when subjects used their right hand than when they used their left hand, F(1,63)=5.514, p=.024.

The third effect involved an interaction between the hand used and the response. It is good to remind here that the response keyboard always had the same orientation, so that the 'YES' button was always at the subject's right, and the 'NO' button on the subject's left. The interaction of hand and response shows that contralateral responses, i.e., responses using the flexors rather than the extensors, gave place to stronger facilitation, F(1,63)=8.33, p=.006. This effect is shown in Figure IV.14.

The fourth and final effect casts some light over all the above. It is a three-way interaction between cue, response and hand used, illustrated in Figure IV.15. This effect shows that the contralateral facilitation disclosed above appears exclusively in the

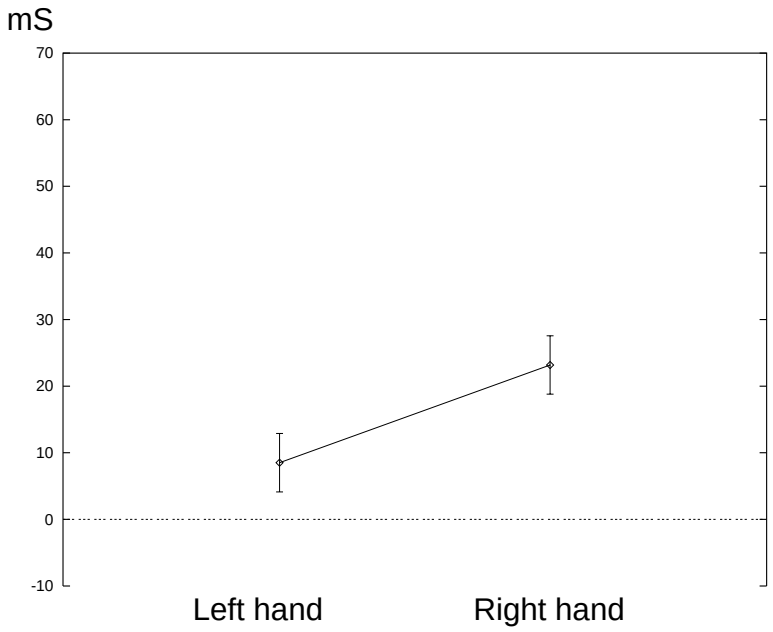

Figure IV.13: **Main effect of hand used:** There is greater facilitation when the right hand is used than when the left hand is used. Points represent the difference between primed and distracted mean reaction times; vertical lines depict standard errors of the means.

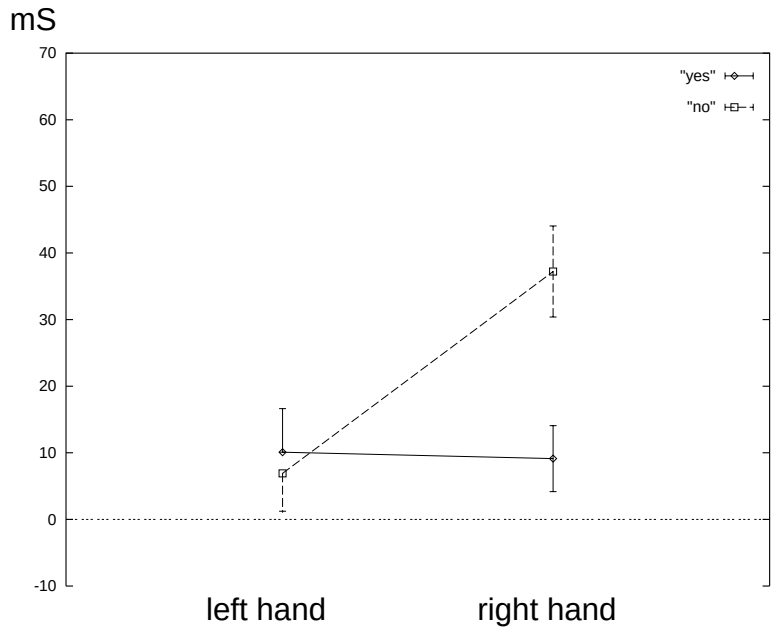

Figure IV.14: **Interaction between hand used and response:** The right hand advantage appears almost exclusively for NO responses. We speculate that NO responses are more reasoned and indirect because they result from a disjunction of cases, and therefore rely more heavily on the analytical abilities of the left hemisphere-. Points represent the difference between primed and distracted mean reaction times; vertical lines depict standard errors of the means.

imagery situation, F(1,63)=7.32, p-.01. Indeed, if separate ANOVAS are made for the "vision" and for the "imagery" data, when the visual cue is used, the main effect for the YES/NO response is significant (p=.003) but the response X hand interaction is very weak (p=.903); whereas when the verbal cue is used, the main effect of YES/NO response is weak (p=.884), but the response X hand interaction is significant (p=.001).

It is difficult to give an interpretation of a three-way interaction, but the observation that forced itself as we plotted this interaction for all subcategories of subjects (left-handers, right-handers, females, males, 'fatigued' and 'unfatigued', and most recouping of these groups), is that the single most stable feature of this result is the strong rise of facilitation in the imagery situation, for the 'NO' response, for subjects using their right hand. Our first reaction was that this could be a consequence of a ceiling effect of the response, wherein response factors influenced the unaltered reaction time, allowing a true perceptual facilitation to show up, or not, in the results. To test this hypothesis, we plotted Figure IV.16, which showed that the base reaction times did not follow the same pattern, which immediately denied that interpretation. At this point, it seemed impossible to argue that the huge difference of facilitation between left and right hand response, in the 'NO' only, could result from low-level interaction.

The strong difference of pattern between 'NO' and 'YES' might have been a result of 'NO' being a second-step decision, since most subjects were testing the hypothesis that both hands were in the grey, which corresponded to the 'YES'. Focusing on the subject strategy, we used the interview information and looked at the minority of subjects who had reported focusing on the white, unshaded part of the stimulus. Again, this subgroup showed the same interaction as the whole group, which argues again the hypothesis that the 'NO' is affected more than the 'YES' because it is a reasoned decision, which might be hindered more by the mental load of the imagery situation.

Thus, we have to look at explanations involving content-specific interference of imagery at higher levels. Several hypotheses are possible. For example, one could invoke the occurrence of extraordinary facilitation when a conjunction of situations are met, such as having imagery, analysis and response all handled by the same hemisphere. Alternatively, one could argue that the decision might involve always interaction between both hemispheres, but in a symmetric situation according to the hand used, so that if

imagery facilitates only the route in one hemisphere, the gain in time will be passed on to the response only when the right hand is used. Probably, other such hypotheses are conceivable, our data does not allow us to distinguish between them. The remaining conclusion is that in its higher-levels effect too, imagery appears to be content-dependent.

One must still insist at this point that even taking away the 'NO' cases, imagery facilitates the response for the 'YES', which vision does not. So, our ancillary result that high-level facilitation is present does not invalidate our main result that low-level content-dependent facilitation is present.

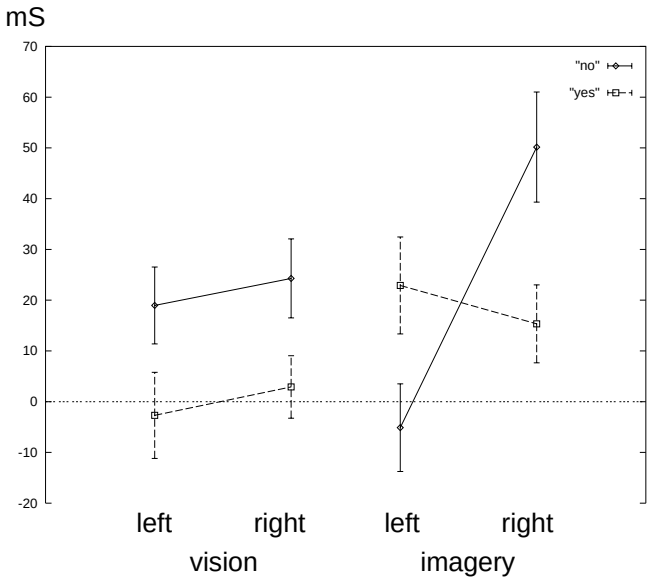

Figure IV.15: **Three-way interaction between cue type, response and hand used:** The salient fact is the strong effect of hand used for imagery NOs only. This interaction suggests that when imagery interacts with high-level representations, it is with strong hemispheric selectivity. Points represent the difference between primed and distracted mean reaction times; vertical lines depict standard errors of the means.

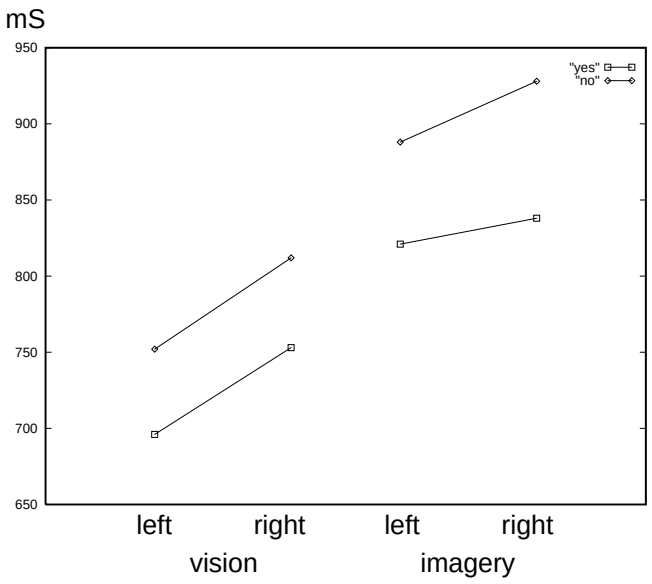

Figure IV.16: **Base reaction times by cue type, response and hand used:** The interaction between cue, response and hand used does not come from a ceiling effect: there is no interaction in the raw reaction times. Points represent the mean of reaction times, primed and distracted confounded; vertical lines depict standard errors of the means.

IV.C.6 Relation between effects and the base reaction times

Beyond the particular case of the interaction just presented, it could have been the case that the magnitude of the observed facilitation had been related to the mean reaction time of 'primed' and 'distracted' in that condition. To investigate that possibility, we looked at the coefficient of correlation between the facilitation ('distracted RT' - 'primed RT') and the mean reaction time in the corresponding condition (0.5*('distracted RT'+'primed RT')). First, we looked at the RT across all data points from all subjects, put together (four times sixty-three, i.e. two hundred and fifty two data points). The resulting scatter, shown in Figure IV.17, along with the correlation coefficient of .07, left little support for the strong hypothesis that the magnitude of facilitation depended on the base reaction time. In second thought, we realized that while such a relationship might not be identical across all subjects, it might exist within subjects, and be masked in the global correlation by between subject variance. To examine this possibility visually, we drew the plot shown in Figure IV.18. This plot repeats that of Figure IV.17, connecting the points of each subject, which have prealably be ordered by increasing abscissa. The rationale behind this plot is that a within subject correlation should appear as an overall upward trend. No such trend is visible. To test numerically the same weak hypothesis, correlation coefficients were computed individually for each subject, to submitted this group to a t-test to find out if their mean would be significantly positive. The mean of these coefficients is .047, about a seventh of their standard deviation, t= 1.045 very far from significance. Clearly, the weak hypothesis of a subject-wise correlation does not fare any better than the strong hypothesis of a nomological relationship between base reaction time and magnitude of facilitation. Thus, differences of facilitation observed in this study can be trusted to reflect true effects, even when the conditions compared have very dissimilar base reaction times.

IV.C.7 Additional analysis of the two cases of 'NO'

Two supplementary analyses were made, one for vision only and the other for imagery only, to see if there was a difference between the two cases of 'NO' responses. As shown in Figure IV.4, the two cases of NO response are when one hand is in the white and the other in the grey ('NO 1'), and when both hands are in the white ('NO

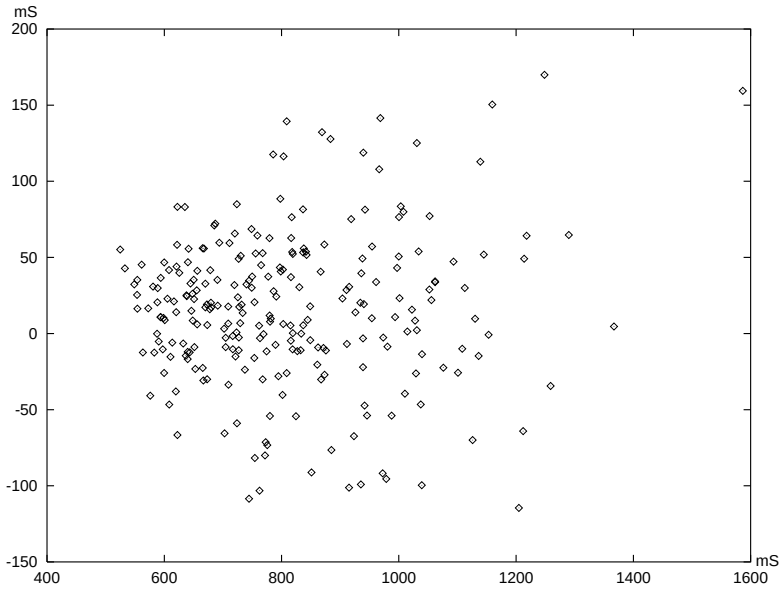

Figure IV.17: **Absence of correlation between base RT and facilitation:** This figure plots the facilitation as a function of the mean reaction time in each of the four conditions (vision, imagery) X (yes, no). Given that the coefficient of correlation is a paltry 0.07, it is unlikely that the greater facilitation observed for "imagery" than for "vision" is due to the relative slowness of "imagery" in basic reaction time.

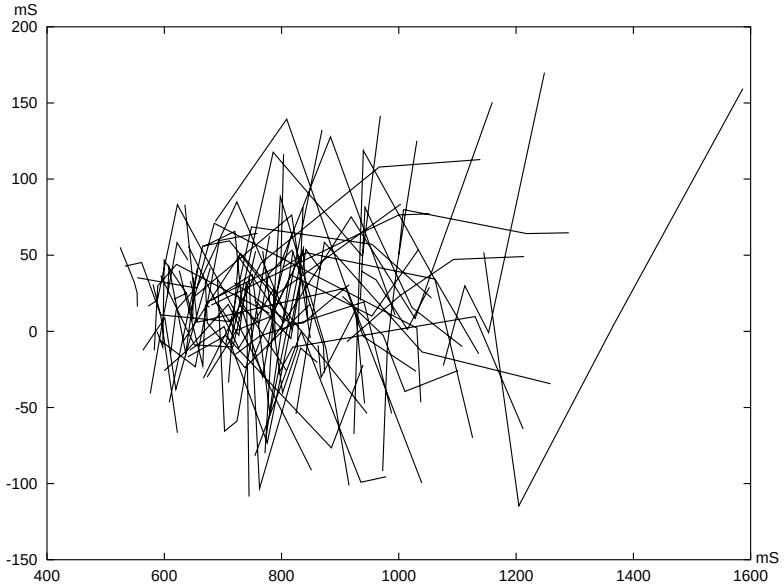

Figure IV.18: **Absence of within-subject correlation between base RT and facilitation:** Here, the previous Figure has been replotted and the points belonging to each subject connected together, because one could reason that the relation between facilitation and base reaction time should be observed on a subject per subject basis. If that relation was generally a correlation, the lines would overall present a positive slope. It is not the case. The coefficients of correlation subject by subject vary wildly, their mean is 0.047 and their standard deviation 0.35.

2'). Because some basic cells contained as few as ten trials, we based the analyses on the mean reaction time, without pruning. A clear pattern arises, as seen in Figure IV.19: with a visual cue, facilitation occurs only in the 'NO 1' case. In contrast, with an imagery cue, an overall facilitation is superimposed to the first pattern. Looking at the raw reaction times, the 'NO 1' response is in average 21 mSec slower than the 'NO 2', $F(61,2)=$, $p<.001$. This confirms the remarks that several subjects made, that they were confused more frequently in the 'NO 1' situation. However, this 21ms difference in the base reaction time is smaller than even the difference in facilitation; thus, and again remembering the results of the previous section, 'NO 1' facilitation should be a true effect.

We believe that the facilitation in the 'NO 1' case must be interpreted in terms of eye movements. Knowing in advance the location of the hands could facilitate a variety of processes, from information acquisition by attention allocation or eye movements to local comparison. However, compared to other factors, it is almost guaranteed to influence data acquisition, especially since we gave no fixation dot or instructions to the subjects. In particular, it is quite possible that deciding the response in the 'NO 1' situation involves one more saccade than in the 'YES' or in the 'NO 2' situation. If this were the case, even only in a substantial minority of cases, knowing the clock in advance could help planning these saccades. The gain in data acquisition from accelerated saccades could then explain the facilitation for 'NO 1'.

What becomes of the other effects already observed, the fatigue factor and the interaction between hand and response for imagery? The role of the 'fatigue' factor seems confirmed because for the finals week subjects, there is virtually no difference between the vision and the imagery patterns: only 'NO 1' are facilitated. Thus, all the difference between vision and imagery observed in Figure IV.19 comes from the non fatigued subjects.

The hand-response interaction for imagery presents an interesting pattern: it appears to affect as well the 'NO 1' and 'NO 2' cases. As shown in Figure IV.20, the 'NO 1' and 'NO 2' have parallel increments from the group using their left hand to the group using their right hand. The main effect in favor of the right hand is significant, $p=.006$, and so is the response/hand interaction, $p=.033$. Since both 'NO 1' and 'NO

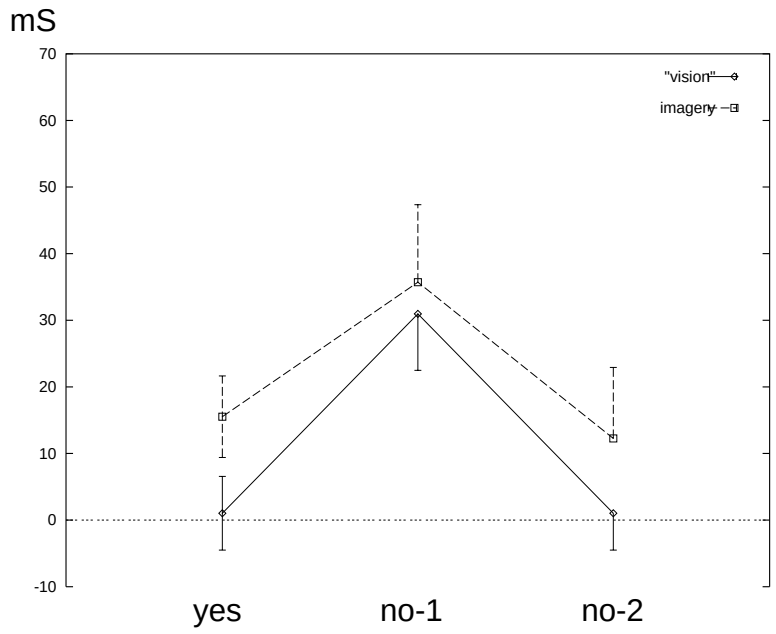

Figure IV.19: **Facilitation per detailed categories of response:** The facilitation for vision and imagery as a function of the response case, shows that for vision there is no facilitation except in the 'NO 1' case. We believe that attention pre-allocation or eye movement planning account for that facilitation. In contrast, for imagery, there is an overall facilitation, which apparently combines with the attentional facilitation, with some ceiling effect, in the 'NO 1' case. The clear difference between the pattern of facilitation for vision and that for imagery prevents an interpretation of the imagery facilitation in terms solely of eye movements or attention allocation. Points represent the difference between primed and distracted mean reaction times; vertical lines depict standard errors of the means (for readability, only half the bar is shown when overlap would occur).

2' are affected by the gain in facilitation from the left to the right hand, it is likely that the interaction does occur at fairly high levels, where the image is integrated in a representation no longer dependent of eye movements, probably, not retinotopic any more. Thus, this particular result supports our initial analysis of this interaction in terms of content-dependent high-level interaction.

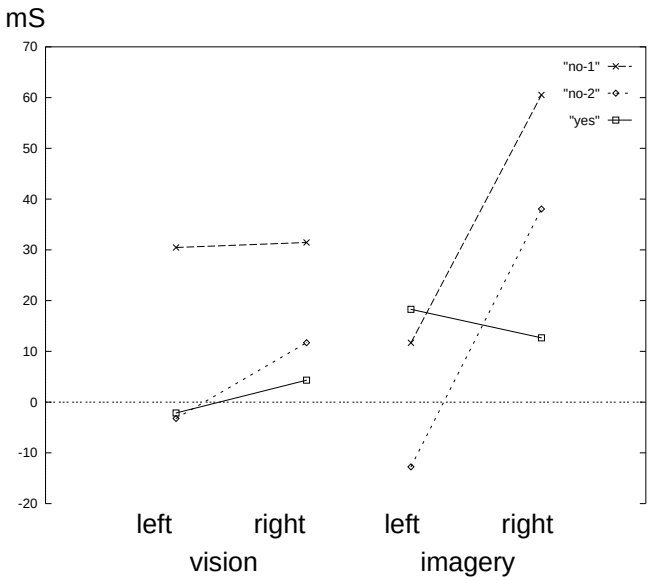

Figure IV.20: **Response X hand used X cue type interaction with detailed categories of response:** The response/hand interaction in imagery affects identically the 'NO 1' and the 'NO 2' cases. Because the effect of hand used on the NOs is additive to other, low-level effects, this supports our hypothesis that this interaction is due to high-level interference. Points represent the difference between primed and distracted mean reaction times; error bars are omitted for readability.

We also compared with the all-subjects results the results obtained from those subjects who claimed to use the alternate strategy of looking at the white, or testing if the hands were in the white. Their pattern of result is not very different from the overall pattern. This argues strongly for interpretations in terms of low-level, strategy-independent effects.

IV.D Discussion and Conclusion

IV.D.1 Summary of Results

In this study, we measured the difference of reaction time between a "primed" and a "distracted" response to a visual task, which consisted of detecting whether two hands of a clock dial fell into a grey plane (Figure IV.4). We compared the facilitation induced by priming when the priming element was a visual image of the clock dial with the facilitation obtained from a mental image of the clock dial.

The main result of this study is that imagery interacts with vision in a content-dependent way. Comparing two conditions, one where congruent or incongruent stimuli were seen to one where there were visualized, significantly more effect of congruence is observed for imagery than for visual exposure.

In addition to that main effect, we found interaction which suggest principally of a strong hemispheric asymmetry in imaging. In the imagery condition only, hand and response strongly interacted, betraying mostly a difference of 55 mSec between the facilitation obtained with the left hand (-5 mSec) and the right hand (+50 mSec). This very robust effect holds for females as well as males, and even for the four left-handed participants.

IV.D.2 Discarding the after-image hypothesis

Is it possible that the stronger facilitation with imagery is observed because in the vision condition a negative afterimage interferes with perceiving the stimulus? This would be a sensible explanation if the facilitation was evenly stronger in different subcategories. However, the difference of facilitation can be broken down and traced to two particular effects: a difference on the 'YES' and 'NO 2', but none in the 'NO 1', and a strong difference between the right hand responders and the left hand responders, for both 'NO's, but not for 'YES'. The afterimage hypothesis cannot account for either of these effects, nor can it be combined with any other simple hypothesis to account for these effects. It is therefore likely that afterimage interference did not play a substantial part in our result.

IV.D.3 Just seeing the congruent cue induces no priming

It comes as no surprise that when the visual cue is the same dial as comes up in the next stimulus, the response is overall faster. However, the detailed pattern is that the response is faster only in the "NO" condition, for those stimuli where one hand is in the grey and another in the white. The facilitation in this case is likely to arise from spatial attention allocation or from eye movement pre-planning. In the "YES" and in the other "NO" condition, seeing the stimulus ahead of time apparently does not induce priming for this task.

Although this result is initially shocking, it makes perfect sense considering the task. The task is based on the image itself rather than on the time displayed; moreover, because the grey plane can take any position, the priors are 50% for each response regardless of the clock. Thus, the response is probably based on a *visual routine* (Ullman, 1984). Knowing the time is not relevant to the task.

The interpretation that in one subcase – NO with one hand out of the grey – eye movement planning can have facilitated the task reinforces Martha Farah's hint at eye movement confounds in the Peterson and Graham (1974) experiment. In their experiment the task was detection, so that their measure of imagery facilitation was more direct, but did not allow splitting in stimulus subgroups the way ours does. In addition, Peterson and Graham (1974) had a control condition where subjects listened to cues rather than saw the image ahead of time. Presumably, when the subjects listen to the cue without visualizing, it is harder for them to anticipate the location of the relevant portions of the image and use this knowledge for saccade spatial planning. In contrast, the control condition used in this experiment allowed the building of spatial representation as good or better as those arising during imagery. Thus, not only eye movement confounds, but any kind of spurious facilitation based on this spatial knowledge should arise in the control condition as well as in the imagery condition.

IV.D.4 Imagery actively interacts with vision

Having observed that visualizing the right cue makes judgement faster than visualizing the wrong cue, we can use the control condition to argue that it is really imagery which causes the effect.

A first possible confound would have been classical priming. A priming effect could be obtained if a representation used for the task is easier to reactivate than to activate. For example, the semantic representation of time could be primed by the cue, and being more easily reactivated, it would facilitate judgement on the stimulus clock. Obviously, such a mechanism would have the same effect, if not a better effect, with the visual cue. If the clock has just been seen, all representations used for processing the stimulus have been activated. Imagery could at best do just as much. Therefore, observing stronger effect for imagery cannot be explained by classical priming.

A second possible confound is the use of various strategies for acquiring information. The confound of eye movements, pointed out by Farah (1985b), is just one case illustrating how expecting a particular stimulus can bias the subject's strategy so that this stimulus can be identified faster, to the detriment of others. Besides eye movements, which are overt attention phenomena, covert attention phenomena also come to mind, such as the pre-allocation of spatial attention in relevant areas of the expected stimlulus (for example, between the expected position of the clock hands). However, because all these strategies are based on expecting a particular clock, they should facilitate the judgement with a visual cue as much as the judgement with an imaged cue. This is not the case, hence the observed facilitation with an imaged cue cannot be explained by information-acquisition strategies.

Because the facilitation with imaged cues appears genuinely stronger than that with visual cues, and because that difference cannot be accounted for by other mechanisms, the remaining explanation is that imagery indeed affects the judgement more that prior vision. Thus, the interaction between imagery and judgement depends genuinely on whether the cue imaged is the same clock at the stimulus clock.

There remains to prove the case that the effect observed is a perceptive effect. Two points especially tend to contribute to this conclusion. First, the judgement that the task requires is extremely visual; especially in the 'YES' situations, there is no reasoning necessary. Second, when the task becomes a trifle more reasoned, in the 'NO' situation, a complex interaction (hand used X response for imagery cues) can be observed, which superimposes with the basic difference of effects. The best explanation for this is that the complex interaction is an asymmetric high-level effect, and that the basic facilitation

is a low-level effect. In conclusion, it appears that the most reasonable interpretation of the observed patterns is that visualizing the right cue makes it easier to *perceive* the stimulus than imaging the wrong cue does.

IV.D.5 What is the balance between facilitation and inhibition?

Is there true facilitation ? In the light of the Perky effect studies reviewed in the introduction, which argue rather in favor of inhibition, one might ask whether the content-dependent interaction comes from facilitation with the right mental image, inhibition with the wrong mental image, or both. Including a neutral condition with equal mental load, especially mental load equally taxing to the same subsystem, looks almost impossible. This puts into a new light the initial question: "facilitation or inhibition?" – compared to what?

In good logic, perception during imagery should be compared to the situation without imagery; when this is made, it is found always worse, presumably because the "mental load" of imaging hinders the response. In our experiment the content-dependent effect favors congruent imaged cues but the base reaction time is slower in the imaging condition, even with congruent cues. We can, however, look at the subjects having run with interleaved trials, because likely the difference of mental load between imagery and vision cues will be reduced for them. When we look at the subjects for whom the trials were interleaved, vision and imagery times come much closer, and for those subjects it appears that in the "primed" situation the reaction has the same speed for both types of cueing, and this equality holds in subpatterns due to response and hand used. Thus, provided the cueing clock was the right clock, whether those subjects had seen or visualized that clock did not affect their time. All the difference came from the "distracted" priming situation. Even then, when trials were interleaved, there was likely more mental load when the cue was imaged, probably slowing the response to those trial. So, since the times were identical, there must have been some real facilitation at the low level.

However, we believe that the balance is tipped towards inhibition in the case of bad cues, rather than towards facilitation in the case of good cues. The reason for this is that most likely top-down effects of the visual system are used by vision too, not just

by imagery! So, the benchmark for comparison is not a "neutral", feed-forward process. It is a visual system equipped, too, with top-down processes. When we engage imagery voluntarily, we recruit part of those processes. Some of the top-down processes must be totally automatic, "encapsulated". Maybe, for example, feedback from V2 to V1. Those are not controlled by imagery. Some must be totally under the control of volition, and not engaged unless we want them too. Maybe, for example, feedback from frontal cortex to temporal cortex. Their effect would be totally symmetrical: a facilitation for the right cues identical to the inhibition for the wrong cues. And some processes are likely to be semi-automatic, like breathing[6]. They will work relentlessly in ordinary vision, but can also be recruited by force of will, to visualize something. For those processes, an congruent imaged cue gives a little timing onset advantage over no cue, but it gives no advantage over having seen a visual cue. But an incongruent imaged cue gives a strong penalty, because the the hindering process is not disengaged until much after perception has occurred. Let us compare with breathing again: if we force our breath out when it should occur, it can make us a little faster to breathe, compared to letting it happen automatically. But if we decide to force our breath out when it's not the right time, it can make us a lot worse! So, if one considers *the whole system*, it is likely that there is less facilitation from the right imaged cue, than there is inhibition from the wrong imaged cue.

Why should the top-down influence be slower to change with imagery cueing? First because the imaging probably necessitates that one hold onto higher representations of what is imaged. If the higher representations are slower to change because of the ongoing imagery effort, then the feedback they provide to lower representation will change later. Second, previous experiments, in particular (Craver-Lemley & Reeves, 1992) suggest that imagery takes several seconds to be disengaged. In the present experiment, the interval between the cue and the stimulus was set to three seconds[7]. In

[6]Automatic processes are known to be involved up to fairly high levels. For example, automaticity is now believed to be responsible for the Stroop effect (Stroop, 1935; McLeod, Driver & Crisp, 1988), where upon naming the ink color of word, one is slowed if the word itself names another color.

[7]Mostly, because pilot tests using the clocks pretest indicated that the median scorer took three seconds per item on that pretest. We reasoned that almost all this time was used to read the time and visualize it; hence the interval of three seconds in our experiment was chosen to allow maximal discrimination between subjects still engaged in visualizing with subjects having achieved their mental image, and that discrimination was expected to follow that made by the pretest results.

retrospect, one could regret that this interval has not been varied, because it is possible that the stronger facilitation for imagery happens precisely when subjects are still engaged in forming the mental image. However, the results of (Craver-Lemley & Reeves, 1992) indicate that imagery takes four to six seconds to be disengaged; since our clocks pretest did not correlate either positively or negatively with the facilitation, we tend to agree with this finding. It is quite possible, and indeed likely, that both reasons collapse into one, because the primary reason why imagery would be slow to be disengaged could be that some areas, perhaps very remote from the perceptual processes, hold onto outdated representations and keep providing inconsistent feedback. (Craver-Lemley & Reeves, 1992) do not allow us to determine if the change of imaged content has the same latency as the total disengagement of imagery, but nevertheless in the light of their results, it seems likely that ongoing incongruent imaging delays the alignment of top-down feedback with the stimulus image.

A consequence of using otherwise automatic processes is that imagery should be relatively effortless. If this is the case, it is not the visualization process themselves which generate a bothersome mental load, but rather the rehearsal of the task instructions, and also, more generally, all the correlates of the mindset of having to visualize, of keeping the necessary concentration and determination. Thus, this consequence spells that mental load is not a low-level factor, and can be equalized by just grouping in the same blocks trials of different type. This prediction is borne out in our interleaved condition.

As a side remark, the above point that imagery should be effortless has probably been missed so often because introspection betrays us there. What is effortful with imagery appears to be *trying*, not *doing*. Surely, keeping one's mind in the most receptive state where images may form does require effort, but this technique is not imaging, not any more than the athlete relaxing his body to obtain his best spring is actually jumping. Yet our everyday introspection reports us the effort we decide to use to obtain imagery, but fails to notice how little images depend on that effort. The effortlessness of imagery appears much more clearly in reports of eidetic imagers. Both Luria's patient (1968) and Purdy's subject (1936) report this effortlessness.

This effortlessness is also a strong prediction of our attractor model of mental

imagery. Our model suggests that areas of representations have a bi-stable dynamics, so that on first approximation, a given level of representation is used totally or not at all. In this model the possibility of forming images depends first on the connectivity gain, then on having supra-threshold input; yet connectivity gain is not controlled directly by neuronal activity, and input comes itself from other areas obeying the same all-or-none activation law, so that it cannot be modulated in strength – only in duration, by the mediation of concentration. Other models, where the representations in each area can have gradual intensity, would *a contrario* predict that the intensity of effort is paramount in obtaining good visualization. Thus, those side results that vision and imagery "primed" conditions may have the very same patterns possibly bear on our deepest theoretical issue, that of which mechanism implements image formation.

IV.D.6 Hemispheric difference

The effect of which hand was used to press the buttons was the strongest with imagery cueing, for the "NO" responses. With imaged cues, on 'NO' stimuli, much higher facilitation was observed with subjects using their right hand than with subjects using their left hand. A more detailed analysis also shows that the effect affects equally "NO"s with one or two hands out of the grey. As discussed in the results section, several possible hypotheses involving low-level interaction are denied by observing various sub-cases. The major line of explanation remaining is that this difference originates in content-dependent interaction of imagery at higher, non-pictorial representations.

Lateralization of imagery, favoring the left hemisphere, has been studied particularly since the review of cases made by Farah (1984), who looked at the localization of brain lesions in patients having lost imaging ability. Many reviews and studies since have nuanced the idea that the left hemisphere is always predominant in all cases of imagery generation, but at least it is predominant when images are made of composite parts. Kosslyn (1994) reviewing the lateralization of imagery in the left hemisphere (pp311-319), wrote: "that the left hemisphere could perform imagery tasks that required adding high-resolution parts to an image [...] and could also perform imagery tasks that required only the global shape of an object [...] In contrast, the right hemisphere was impaired when parts had to be added..." Considering that the imaging part of the task

in our experiment is an archetype of task requiring to assemble parts (the two hands, almost independently specified, plus the dial), our results are strongly in line with Kosslyn's conclusions. Thus, the left-hemisphere mediation of image generation must account for having content-dependent interaction strongly detectable with a right-hand response but invisible with a left-hand response.

IV.D.7 Conclusion

This experiment shows that mental images interact with visual images. The interaction is between the *contents*, rather than just between the *processes*. Does visualizing the Mona Lisa hinder any perception? No. It hinders perceiving other scenes, and facilitates seeing the Mona Lisa. True, in the experimental condition, the imagery response, regardless of content, is slowed by mental load. But in real life, imagery is often nearly effortless and spontaneous; using it often provides the least mental load. Anyway, content-dependence is best explained by shared representations for vision and imagery. And since interaction appears at both low and high level, representations are probably shared at both low and high level. In conclusion, the major point shown by this experiment is that mental images interact on a case-by-case basis with visual images.

IV.E Appendices

Appendix 1 Clock pretest

Appendix 2 Direction pretest

Appendix 3 Letter pretest instructions

Appendix 4 Instructions for the practice

Appendix 5 Instructions for the Vision block if it comes first

Appendix 6 Instructions for the Imagery block if it comes second

Appendix 7 Instructions for the Interleaved block

Appendix 8 Instructions for the Vision block if it comes second

Appendix 9 Instructions for the Imagery block if it comes first

Appendix 10 Instructions for the Baseline

Appendix 11 Why D' must not be computed from averaged rates

IV.E.1 Appendix 1: Clock pretest

Instructions

For each of the digital times given next page, visualize a clock dial showing this time. Look at the hands and find out if they make an angle smaller than 90 degrees or larger. Then check NARROW if the angle is smaller than 90 degrees and WIDE if it is larger than 90 degrees. It does not matter which hand is ahead, for example 11:00 and 1:00 are both narrow.

Do not forget that minutes also push the hour hand further, for example in "8:55" the hour hand is almost on the "9".

Examples

12:00 is narrow. 12:15 is still narrow. 12:20 is wide. Practice: 03:30 05:50 10:20 10:10 08:25

You have two minutes to complete as many trials as you can. Trials are presented on two columns... nothing is tricky, but do not forget the second column.

Do not guess – find out what the angle is. Tell me you are ready, then turn the page over and start.

Test

NARROW	LARGE	NARROW	LARGE
12:15		06:25	
05:30		03:10	
02:45		01:55	
12:55		12:05	
11:05		05:50	
10:40		10:05	
09:30		07:30	
04:20		10:00	
08:50		08:00	
11:25		07:15	
06:50		04:50	
12:25		04:00	
03:35		01:20	
02:15		03:30	
07:20		04:15	

06:50

10:10

11:45

05:45

IV.E.2 Appendix 2: Direction pretest

Write your name here:

Instructions

In each of the lines of this test, imagine yourself in one of your favorite locations (e.g., on the beach). You must be facing in the direction given by the first letter of the line: E is east, S south, W west and N north. For example, you would be facing West on the beach, towards the ocean, or facing East towards the mountains. Select a landscape/cityscape which you know well and can visualize easily, then place yourself facing the indicated direction. From there, take the turns indicated by the following letters, one after the other: R is right, L is left, F is forward (no turn), and B is backward (a U turn). Finally, report the direction you are facing.

examples W L R L L L R

You start facing West, then turn left, so you are facing South, then right, so you are facing West, then left, so you are facing South, then left again, which takes you facing East, then left again, which takes you to the North, then right, so you finish facing East. Hence, you write down "E".

E F F F F F B

You start facing East, then nothing changes for five rounds, but at the last round you turn around, so that you end up facing West. Hence, you write down "W".

N R R R B F F right answer: E

S B L L F R R right answer: N

S R F R B R L right answer: S

You have three minutes to complete as many trials as you can.

Do not guess; do not "compute". Visualize. Tell me you are ready, then turn the page over and start.

Test

Start Point	Turns	End Point
W	R R B F L R	
S	R L F L L R	
E	L B F L B F	
N	B R B L L L	
S	L B R F R R	
N	R R L B L F	
E	L F B R R L	
E	R B F L L R	
W	L L L F B R	
E	R R L B L R	
N	L L B B L R	
S	B R B L R R	
E	L R B F R B	
N	R L R F L L	

S L R L B L L

N L L F R L B

E L B R L L F

E L R B L F L

W L F L B R L

E L L F R L R

IV.E.3 Appendix 3: Instructions for the Letter pre-test

"OK, we are going to do the last pretest using the letters which I showed to you at the beginning. You'll have to do this by visualizing one of the letters, but I'll show you on an actual letter. OK. [showing on the 'L' with a pencil] You start at the bottom right corner, then you follow the outline of the letter, clockwise, and tell me what turns you have to take, until it's over. Let's do it on the L. [simultaneously doing it on the 'L' with a pencil] You start here, bottom right corner, then you tell me 'right', 'right', 'right', 'left', 'right', 'over'. Don't count the starting corner, just say 'over', OK? So, you'll have to do this ion the mind's eye, I'm not showing you the letter. In a moment, I will tell you which letter, and start the clock. I time you from the moment I say the letter to the moment you say 'over', and I also write down the turns you tell me. Are you ready?... [waiting for subject approval] We are going to do it with the letter...'F'. "

IV.E.4 Appendix 4: Instructions for the practice

" The next set of tasks, which is the experiment proper, will take about thirty minutes, and I'll ask you if you want a break half way. The experiment has four ["three"

in the case of interleaved trials] parts: a practice session, followed by two test conditions ["a main test condition"] and a baseline, which I'll tell you about after the practice. In each case, the stimulus you will see is a clock that is half shaded and half unshaded, like this [showing on the screen a sample stimulus, like one of the clocks in Figure IV.4]. You will respond using this keyboard [placing the keyboard in front of the subject]. If both hands are in the shaded part, like this [showing a sample 'yes' stimulus], press the dark green button. If one or both hands are in the white, unshaded part, like this [showing a stimulus with hands astride the grey plane], or like this [showing a stimulus with both hands unshaded], press the red button. So, I summarize: two hands in the shade: dark green button, two hands in the the white, or one on each side: red button. Do you have questions? Let us now get in place. OK, I want you to use only your left ["right", depending on subject factor assignment] hand. The rest position is here, between the buttons, and you can move right and left like this [demonstrating]. You can rest your hand on the pad or not – depending on what you feel most comfortable with.

Now, you must be 70 cm from the screen. [getting a ruler to place the subject 70cm from the screen]. That's what this ruler is for. Make yourself comfortable in the chair, take the same position as you will during the experiment, and we'll see if we must roll the chair up or back. OK, we are all set. Are you ready?"

IV.E.5 Appendix 5: Instructions for the VISION block if it comes first

"That was good. Now we are ready for the first test. The stimulus, the clock with the shade, is the same, and the response is the same. So, you'll still have to press the green button if both hands are in the shade, and the red button otherwise. However, there is a novel element: before the stimulus, you are shown another clock, with no shade. So, in each trial, you attend a first clock but do not respond to it. It does not have the shade anyway. Then, the screen stays blank for three seconds, then the shaded clock comes up and you respond to it. Half the time, the first clock will be the same as the second, shaded clock, and half the time it will be different, so do not try to prepare – just pay attention to the first clock, then respond to the second clock. There are eighty trials in this part, and it should take six to eight minutes. Is it clear? Do you have any questions? Let's get in place. [placing keyboard, checking distance]. Are you ready?

Go."

IV.E.6 Appendix 6: Instructions for the IMAGERY block if it comes second

"That was good. The second part of the test is very similar to what you just did. The only change is that instead of seeing a clock before the stimulus, you have to visualize one. The stimulus, the clock with the shade, is still the same, and you respond to it the same way. So you'll still have to press the green button if both hands are in the shade, and the red button otherwise. But before that stimulus comes up, you are given a time, such as 'half past seven' or 'ten till nine', and you then have three seconds to visualize the corresponding clock. So, in each trial, you first get this visualizing instruction, then the screen stays blank for three seconds and during that time you visualize a clock showing the time given. Then the regular stimulus comes up and you respond just as you've been doing until now. So, your response is based on the stimulus only; the only thing you do with the written time is to visualize a clock showing it. I'm interested in what that does to your perception. Half the time, the visualized clock will be the same as the second, shaded clock, and half the time it will be different, so do not prepare – just visualize the first time, then respond on the shaded clock, regardless of the first one. Except the first part, everything in the design is similar to what you just ran through. So, there are again eighty trials with a break after the first fifty, and it should again take six to eight minutes. Any questions? OK, let's go."

IV.E.7 Appendix 7: Instructions for the INTERLEAVED block

"That was good. Now we are ready for the main test condition. The stimulus, the clock with the shade, is the same, and the response is the same. So, you'll still have to press the green button if both hands are in the shade, and the red button otherwise. However, there is a novel element: before the stimulus, you are given a probe. The probe can be either another clock, without shade, or a written time, such as 'half past seven' or 'ten till nine'. After you see the probe, the screen stays blank for three seconds. If the probe is a clock, you pay attention to it. If the probe is a time, you have to visualize the corresponding clock. After three seconds, the regular stimulus comes up and you

respond to it just as you did during the practice. So, I summarize: each trial, you will have a probe: either a clock without shading, or a written time, then you have three seconds to visualize the clock corresponding to this time, or do nothing, then a shaded clock comes up and you respond to it. Half the time, the probe and the stimulus are the same, and half the time they are different, so do not try to prepare – just pay attention to, or visualize, the probe clock, then respond to the shaded clock. There are a hundred and sixty trials in this part, with two breaks, and it should take about fifteen minutes. Do you have questions? OK, let's go."

IV.E.8 Appendix 8: Instructions for the IMAGERY block if it comes first

"That was good. Now we are ready for the first test. The stimulus, the clock with the shade, is the same, and the response is the same. So, you'll still have to press the green button if both hands are in the shade, and the red button otherwise. However, there is a novel element: before the stimulus, you are given a time, such as 'half past seven' or 'ten till nine', and you then have three seconds to visualize the corresponding clock. So, in each trial, you first get this visualizing instruction, then the screen stays blank for three seconds and during that time you visualize a clock showing the time given. Then the regular stimulus comes up and you respond just like the practice. So, your response is based on the stimulus only; the only thing you do with the written time is to visualize a clock showing it. I'm interested in what that does to your perception. Half the time, the visualized clock will be the same as the second, shaded clock, and half the time it will be different, so do not prepare – just visualize the first time, then respond on the shaded clock, regardless of the first one. There are eighty trials in this part, and it should take six to eight minutes. Is it clear? Do you have any questions? Let's get in place. [placing keyboard, checking distance]. Are you ready? Go."

IV.E.9 Appendix 9: Instructions for the VISION block if it comes second

"OK, That was good. The second part of the test is very similar to what you just did. The only change is that instead of having to visualize a clock before the

stimulus, you get to see one. The stimulus, the clock with the shade, is still the same, and you respond to it the same way. So you'll still have to press the green button if both hands are in the shade, and the red button otherwise. But before that stimulus, you are shown another clock, with no shade. So, in each trial, you attend a first clock but do not respond to it. It does not have the shade anyway. Then, the screen stays blank for three seconds, then the shaded clock comes up and you respond to it. Half the time, the first clock will be the same as the second, shaded clock, and half the time it will be different, so do not try to prepare – just pay attention to the first clock, then respond to the second clock. Except the first part, everything in the design is similar to what you just ran through. So, there are again eighty trials with a break after the first fifty, and it should again take six to eight minutes. Any questions? OK, let's go."

IV.E.10 Appendix 10: Instructions for the BASELINE

"OK, we have one last really quick one. It's just like the practice: only 20 trials and just the shaded clock, no first part. This is going to be really fast. I just want to see how much faster you are at the end than you were at the beginning... OK, are you ready?"

IV.E.11 Appendix 11: Why D' must not be computed from averaged rates

Mathematicians would tell that the computation of D' does not commute with averaging operations, because it includes a non-linear transformation. For the rest of us, an example will illustrate the point more clearly.

Let us suppose we have two subjects with keen judgement, A and B. By keen judgement, we mean that their D' for the task required is rather high, say, 2.58. However, they have very different criteria: A is very conservative and B is adventurous. So, the rates end up like this:

```
A; hit: 50 %   false alarms: 1 %
B: hit: 99 %  fa: 50 %
```

Our well-intending grad student computed the population D' from averaged rates:

```
hit:  74.5 %   z= .67
fa :  25.5 %   z= -.67
```

And would find a resulting D' of 1.34, much lower than the common D' of our two subjects.

Moreover, should we retest A and B under a new experimental condition, which raises their actual D' considerably, the D' computed from averaged rates will hardly change. Even worse: if we test A and B under a new condition which changes only their criterion, for example, a condition which makes both of them more conservative, the D' computed from averaged rates will increase, enven though the real D' is unchanged...

With a little imagination, one can construct examples of every situation where effects observed on D' computed from averaged rates are pure artifacts of the rate averaging. The solution to this problem is, of course, top compute individual D' for each subject and situation, and to do the statistics on those.

Chapter V

General Conclusion

"...Such tricks hath strong imagination
That if it would but apprehend some joy,
It comprehends some bringer of that joy.
Or in the night, imagining some fear,
How easy is a bush supposed a bear?
But the story of the night told over,
And all their minds transfigured so together,
More witnesseth than fancy's images,
And grows to something of great constancy;
But howsoever, strange and admirable."

William Shakespeare, *A Midsummer Night's Dream*, V.i.

In three different parts, using different techniques, we have explored the same corpus of ideas. We see the visual system almost as a federation rather than a unitary whole. Each subsystem in the federation is submitted to two pulls: its input, which can be feedforward –in vision– or feedback input –in imagery, and its own dynamics, which pulls it towards attractors proper to that subsystem. The resulting activity pattern is a compromise of these two pulls. Subsystems can be active or inactive because they are bistable, and "opt out" of the current configuration. In particular, during mental imagery, occipital brain regions can be active, causing a vivid imagery, or inactive, making mental images more abstract.

Experience, more than theory, tells us that there is greater variety in imagery

than in vision. Tallying different states of vision, one finds that color and black-and-white vision appear to be fairly different states, and also that scenes segmented and understood, also appear as a different state from raw visual sensation. Past these two cases, it is difficult to find non-pathological variations in the configuration of the visual system. In contrast, imagery appears to present in addition a scale of vividness with at least three different levels: non pictorial or abstract imagery, ordinary pictorial imagery, and eidetic imagery. These levels correspond probably to the activation of increasingly occipital brain areas. The greater variation in imagery likely results from balancing evolutionary pressures. Feed-back connections and lateral connections, which together are responsible for imagery, are beneficial for the visual system until they become dominant to the point of smothering visual input[1]. It could also be the case that there are evolutionary pressures for having in the whole human population individuals with both dominant types of imagery, abstract and pictorial, because their problem solving-skills are complementary. Altogether, because vision is more immediately useful than imagery, one expect evolution to produce it on a more dependable basis. Hence, it makes sense that the flexibility of configuration of the visual system is better expressed in imagery than in vision.

If the visual system is indeed a functional hierarchy of representations, imagery must present a computational advantage. Abstraction in the visual system is as much a *loss*[2] of information as it is a reorganization. Imagery allows to compensate for that loss by recreating the most likely original representation. Each area active in imagery recreates the best representation consistent with the top-down requirements; that representation is *richer* than those top-down requirements, because it incorporates knowledge about the world encoded in the connectivity of that area. The modern view of *knowledge* differentiates between explicit and implicit knowledge. The knowledge brought into an image by pattern reconstitution is even more structural than implicit knowledge. In the case of eidetic imagery, probably involving low-level occipital areas, it could be species-specific knowledge, residing in the capture of statistical regularities of

[1] About this, the case of Vienamin, Luria's patient (1968) once more comes to mind. Vienamin suffered from social inadequacies because of his hypertrophied imagery: in particular he would easily believe he had fulfilled his duties after merely visualizing the corresponding scenes.

[2] In our particular view, a substantial part of this loss is operated by the attractor dynamics at each level. Classification operates a contraction of the representational space.

the visual scenes, by the visual system anatomy, over eons of evolution. Thus, imagery appears as a primary function of the brain, the mirror-image of abstraction.

To study imagery as a consequence of attractor dynamics, we took three different apporaches. These three parts of the thesis correspond respectively to a theoretical analysis, a model, and an experiment.

In the first part, we explain how attractor dynamics in the brain leads to imagery. A review of the visual system presents it as a functional hierarchy of areas encoding different representations. We then follow Steve Kosslyn (1994) in his presentation of imagery as a percept arising from top-down input in the absence of bottom-up input. To introduce our vision of *how* mental images are formed according to this principle, we first present several examples of attractor networks, and show how they tend to implement bistability and pattern-completion. These two basic properties mean that imagery uses brain areas in an all-or-none fashion –bistability– and is able to restore lost information –pattern completion. The consequence, a visual system where imagery can have various configurations, is then matched with the variety of imagery types which have been observed. We conclude that the more vivid types of imagery correspond to configurations of the visual system where more of the occipital areas are wholly active.

In the second part, we present a computer model of a single layer illustrating the properties of attractor dynamics. Each "neuron" in this layer is connected to others by weights which favor the representation of aligned edges and continuous solids. This layer can get bottom-up input from a grey-level image, through a different Gabor filter for each "neuron", and/or top-down input, from a noised previous version of itself. The neurons have a sigmoid-like transfer function –monotonous, continuous and bounded. The model does present the properties which were expected on theoretical grounds. It can function in two different dynamical ranges: sub-threshold, where activity decays, or supra-threshold, where it grows to a maximum. Over several cycles of simulation, there is a decent sustain of the shape of activity patterns. With noised top-down input, the first few cycles see an improvement of the pattern shape, akin to what we expect of mental image formation. The gain of lateral connections is crucial for the dynamics. Finally, the model shows that top-down and bottom-up input can be combined to overcome the threshold. Altogether, these properties follow the intuitive expectations one can have

from looking either at large linear models or at small non-linear models. The model thus validates the statements made about attractor networks in the theoretical section.

In the third part, we test a particular prediction of both the theory and the model, on how imagery interacts with perception. We expect the interaction to depend on the similarity of the mental image with the visual image. Subjects are cued by a visualization, and must afterwards make a visual judgement on a clock image: they decide if both hands of the clock are in a shaded plane which covers half of the image. We find that when visualizing the right clock, subjects are faster than when they visualize the wrong clock. A control condition uses a visual rather than imaged cue[3]. Surprisingly, the effect is smaller with the visual cue, and actually is restricted to a particular sub-case where one might expect confounds from information acquisition strategies. This experiment thus demonstrates that imagery does influence perception at a fairly low level, and that the influence depends on the content of the image. Thus, there is an interaction between *mental images* and *visual images*, rather than an interaction between "systems". Besides this main finding, lateralization results confirm the growing hypothesis that composite imagery is strongly lateralized. These two results: low-level content-dependent interaction, that we expected, and high-level lateralization of image composition, unrelevant to our point but nicely fitting previous findings, make the experiment a success.

Now that we have presented out view of imagery, it is good to reflect on the questions which arise from its consideration.

Determining which areas of the visual cortex are activated by imagery in different types of imagery combines two issues. First, the increasingly easier issue of brain imaging. The spatial and temporal resolution of FMRI is already sufficient for this. Second, one must solve the antique problem of reliably sorting subjects with archetypal imageries[4]. Many studies in the literature make use of self-reports inspired by Betts (1909). Although the score to these questionnaires usually correlates with expected im-

[3] Lack of control for acquisition strategies, such as eye movement planning or attention allocation, was the Achille's heel of previous studies.

[4] We consider here the problem of identifying imageries of different quality. Other factors such as speed or controllability of the evoked image are, in our theory, subordinate to the basic ability to activate the areas responsible for a given level of vividness.

agery effect, this statistical effect is far from reliable on a case-by-case basis, and the interpretation of the language used in the questionnaires is extremely dependent upon personal experience[5]. The problem of classifying the subjects could be sorted in two ways. One would be to design test where objective landmarks could be used (for example, comparison with the vividness of daylight visual images, night time visual images, dreams, and mental "images" of unimaginable objects such as abstract nouns. The other way would be to rely entirely on the FMRI data and see whether the activation of occipital cortex has a bipolar distribution, showing subjects who always do and some who always don't. Both of these methods have their own problems. New tests would still be liable to some claims of subjectivity. Relying on FMRI data would require data of much better quality to establish bipolarity than is necessary to simply prove more activation in the "high imagers" group. Altogether, however, it seems that FMRI testing of the existence of distinct imagery types is now within reach.

Another set of questions concerns the nature of the actual dynamical regime. It is easier to theorize and modelize with fixed points, but Nature usually couldn't care less about what theorists prefer[6]. In our model, activity remains at a high level unless reset – the reset requires mechanisms not considered. It seems that in sensorial cortices, activity occurs rather in bursts. The rôle of oscillations is still very controversial; in our fixed-point theory, mean firing rate is the variable, which leaves the phase of oscillations, and the coefficient of variation free to be either non-significant, or to fill some other functional rôle such as binding and segmentation (Gray & Singer, 1989). Beyond mere oscillatory behavior, for example if cognition is implemented by ever-varying strange attractors, as suggested by Freeman's later work (Freeman & Skarda, 1990), it is difficult to give any theoretical detail about pattern completion. Inasmuch as the theory relies more on the existence of two regimes than on their temporal implementation, this question is not urgent, but details of timing, of controllability, and other properties of imagery could

[5]The typical item in any such questionnaire is a description of a mental image to be formed, followed by a scale of self-evaluation from "very vivid" to "abstract". No objective landmark allows the subject to determine what is vivid and what is not. In the words of Aaron Cicourel (personal communication), the subjects then define the very language with which they are supposed to be tested.

[6]However, it is also the case that mathematics often allows to consider a reduced image of the phase space, where things otherwise more complex can be reduced to a point. Whereas such transformations usually loose explanatory power, they remain legitimate to analyze classification.

depend upon the detailed dynamics responsible for the bistability.

In conclusion, one should say that by exposing a general mechanism of imagery, one only scratches the surface of things. The variety not only of imagery types, but of other properties of imagery, and of peculiarities depending on the actual implementation of the visual system. *Nature's hacks*, so to say, are the meat of the matter. Finally, because imagery is a top-down effect, its study opens onto phenomena chaining elementary cognitive events, from quick reasoning to long term evolution of personality. Thus, imagery is truly the middle land of the Cognitive Science, straddling perception on one side, and thinking on the other.

Chapter VI

Résumé en français – Summary in French

Cette thèse traite de l'imagination. Elle examine un mécanisme possible de formation des images mentales et s'inscrit en ligne droite dans le prolongement des idées de William James, Donald Hebb et Steve Kosslyn.

Par delà les problèmes qui furent au coeur du récent "débat sur l'imagerie" entre Kosslyn et Pylyshyn, un des plus vieux mystères de l'imagerie est la variété observée des types d'imagerie. Pour chacun d'entre nous, la qualité, la brillance, la précision, la clarté et le réalisme des images mentales varie selon notre état de conscience. La lucidité, la somnolence, l'ivresse, sont chacune le théatre d'images mentales de qualités différentes. Pour certains d'entre nous les "images" mentales sont abstraites, insaisissables, ou fugaces. Pour d'autres, elles sont précises et nettes. Cette variété témoigne de phénomènes apparentés, mais séparés. Comme disait William James (1890), capitalisant sur une étude célèbre de Francis Galton, "Il y a des imaginations, pas une Imagination".

Donald Hebb (1968) proposa un mécanisme de formation des images mentales qui rend compte de la variété des types d'imagerie. Il suggérait que les images mentales sont formées par réactivation d'assemblées de neurones dans les aires visuelles du cortex. Le premier, Hebb avança l'hypothèse que différentes conditions pouvaient amener la réactivation d'assemblées plus ou moins proches de l'arrivée de l'influx sensoriel. Ainsi, la formation d'assemblées de cellules actives dans le cortex strié devrait former une im-

age plus vive que si la réactivation s'opère dans le lobe temporal. La théorie de Hebb manquait à la fois des données biologiques sur les voies visuelles du cortex dont nous disposons maintenant, et d'un modèle rigoureux des "assemblées de cellules". Néammoins, l'intuition initiale était juste sur les deux tableaux, puisque la science a depuis confirmé à la fois une organisation fonctionnelle hiérarchique des voies visuelles, et les propriétés émergentes des réseaux de neurones.

Steve Kosslyn, plusieurs années après la clôture –en sa faveur– du "débat sur l'imagerie", revînt à la charge pour présenter une conception unifiée de la vision de haut niveau et des phénomènes d'imagerie. L'idée phare de Kosslyn (1994) , c'est que le système visuel est truffé de voies descendantes, qui par rétroaction aident les processus visuels de bas niveau (segmentation, liaison, etc.) En l'absence de stimulation visuelle, la rétroaction peut activer les mêmes aires que l'influx sensoriel. L'imagerie est donc le fonctionnement normal du système visuel à vide. Cette théorie, dépeignant un système visuel qui teste le monde plutôt que de le lire, suit une longue tradition récemment ravivée (Churchland, Ramachandran & Sejnowski, 1994) en dépit de l'influence encore très forte de schémas inspirés par l'informatique (Marr, 1982). Kosslyn analyse le système visuel en recensant les influences descendantes, qu'il compare aux différents cas de formation des images mentales. Son système est détaillé mais il n'insiste pas sur le mécanisme de reconstitution des images mentales.

Il restait donc à formaliser la proposition de Hebb et à la relier aux connaissances actuelles sur les réseaux de neurones et sur le système visuel. C'est là le but de la présente thèse.

Cette thèse se compose de trois parties. La première, théorique, détaille notre théorie de l'imagerie. La seconde présente un modèle qui illustre et appuie la théorie. La troisième montre une expérience explorant l'interaction entre l'image mentale et l'image visuelle.

VI.A Première partie: Théorie

VI.A.1 Vision et imagerie

Le système visuel apparaît comme une hiérarchie *fonctionnelle* de représentations. La magistrale synthèse de Felleman et van Essen (1991) présentait déja le système visuel du macaque comme une hiérarchie organique, où l'information visuelle passe séquentiellement d'une aire cérébrale à la suivante, le long de deux voies principales: la voie pariéto-dorsale ("où?") et la voie temporo-ventrale ("quoi?"). Les études sur l'humain par résonnance nucléaire fonctionnelle ont confirmé la séparation fonctionnelle des aires cérébrales (Tootell, Dale, Sereno & Malach, 1996). Des revues et analyses plus récentes (Churchland, Ramachandran & Sejnowski, 1994; Bullier & Nowak, 1995) ont remis en question l'interprétation stricte de cette hiérarchie. Il semble à présent indiscutable qu'il y a une hiérarchie fonctionnelle, mais la circulation d'information n'est ni strictement ascendante ni strictement séquentielle. En somme, les données et analyses récentes indiquent une plus grande indépendance à chaque étape du circuit.

Dans ce contexte, où les rétroactions semblent importantes, s'inscrit la théorie de Steve Kosslyn. Kosslyn propose que les processus visuels de haut niveau contrôlent quelques actions élémentaires qui permettent la perception même dans des conditions difficiles. Par exemple, ces processus contrôlent la position de l'attention spatiale. Ils peuvent également passer en revue les souvenirs visuels des objets suspectés d'être présents dans l'image. Ils peuvent spécifier la position et l'orientation présumée de ces objets. Les hypothèses ainsi implicitement faites sont testées par la projection des images correspondantes dans les aires visuelles de plus bas niveau. Si les hypothèses sont correctes, la perception s'améliore et l'identification est alors confirmée. Selon Kosslyn, donc, la capacité de projection est partie intégrante du système visuel. L'imagerie, quant à elle, est une conséquence heureuse, mais accidentelle, des rétroactions de ce système.

VI.A.2 Attracteurs, complétion et bistabilité

Lorsqu'on munit de connexions latérales divers réseaux de neurones, comme le sont les neurones corticaux –au lieu de se limiter au connexions verticales qui sont le pain quotidien du modélisateur–, on voit apparaître deux propriétés essentielles: la bistabilité,

et la capacité de compléter une structure d'activitée mal formée. Ces propriétés sont montrées dans divers modèles revus (Anderson, Silverstein, Ritz & Jones, 1977; Hopfield, 1982; Wilson & Cowan, 1972; Wilson & Cowan, 1973). Pour tous ces modèles, l'activité peut appartenir soit à un régime élevé, où différentes structures d'activité représentent différentes choses, soit à un régime faible, où l'activité est presque nulle. De plus, les attracteurs créés par la dynamique –généralement des points fixes– capturent l'état de ces modèles, pouvant ainsi changer la forme même de la structure d'activité. Cela équivaut à compléter ou modifier une structure "mal formée". Bistabilité et complétion apparaissent comme des propriétés génériques de ces réseaux.

Pareillement, si l'on regarde des études sur l'animal de la dynamique corticale, il semble que les mêmes propriétés soient visibles. Freeman en particulier, avec différents collaborateurs (1986, 1987), montre que les odeurs dans le bulbe olfactif du lapin sont encodées par l'enveloppe spatiale de bouffées d'oscillations[1]. Les bouffées elles-mêmes sont déclanchées par l'inspiration du lapin, indépendamment de l'odeur présente ou pas. Dans les mesures de Freeman (1986), il y a deux types de bouffées qui diffèrent peu par l'intensité, mais beaucoup par leur cohérence spectrale. Les bouffées "significatives" sont plus homogènes, et sont les seules pour lesquelles l'enveloppe spatiale coïncide avec l'odeur présentée. En somme, la bistabilité n'est pas nécessairement une affaire de niveau d'activation, mais peut se manifester dans d'autres dimensions du signal.

En supposant que les diverses aires du cortex visuel manifestent aussi ces deux propriétes, bistabilité et complétion, que peut-on en conclure pour le système visuel?

D'abord, qu'il existe pour chaque aire un seuil séparant deux régimes d'activité distincts: un régime supra-liminal, où l'activité est significatrice –elle représente quelque chose– et efficace –par sa cohérence spectrale ou son intensité, elle influence le reste du système, et un régime sub-liminal, où l'activité est au mieux une amorce.

Le système visuel présente donc une géométrie variable, selon la configuration d'aires ou de sous-systèmes qui sont activées. Par exemple, si l'on regarde un vieux film de Charlie Chaplin, ou qu'on se promène en campagne par une nuit sans lune, la couleur

[1] Freeman enregistre le champ électrique par des tableaux d'électrodes subdurales, une technique dont la précision temporelle est excellente, et dont la précision spatiale est intermédiaire entre les enregistrements intra-cellulaires et les mesures d'EEG extra-craniennes.

disparaît de notre univers mental.

VI.A.3 Les imageries mentales

Cette variabilité de configuration s'exprime pleinement dans le cas de l'imagerie. Probablement parce que les processus visuels ont évolué, malgré tout, pour donner la priorité à l'information ascendante, les connexions descendantes n'ont pas la même efficacité pour activer les aires de bas niveau. La participation du cortex occipital aux images mentales est donc variable, fluctuant selon l'état de conscience, autour d'une susceptibilité individuelle. Ainsi, on trouve à une extrémité du spectre Viennamin, le patient de Luria (1968), muni d'une imagerie si vivide qu'il ne pouvait la différentier de la réalité, et à l'autre extrémité Aldous Huxley, qui même sous l'influence de la Mescaline échappe aux visions que la drogue suscite d'ordinaire (Huxley, 1957). Entre ces deux extrêmes, il semble qu'il y ait une échelle avec au moins trois échelons: l'imagerie eidétique, l'imagerie vivide ordinaire, et l'"imagerie" abstraite. Non pas une Imagerie, donc, mais au moins trois "imageries".

La propriété de complétion, quant à elle, permet de restituer des détails inconnus des représentations de plus haut niveau. En effet, les lois quasi-universelles de notre environnement, telles que la continuité des bords et des solides, donnent lieu à des images redondantes, truffées d'informations non pertinentes pour l'identification des objets et pour la plupart des processus de haut niveau. Il serait donc normal que ces informations soient éliminées par les processus ascendants. Par contre, lorsqu'un modèle précis du monde devient nécessaire, ces détails doivent être restitués. Ils le sont par complétion. En somme, plus une régularité de notre environnement est fiable, plus l'évolution en a confié la connaissance à un bas niveau de traitement. De même qu'on parle de connaissance explicite et implicite, il faudrait pour qualifier les connaissances encodées dans une connectivité prédéterminée parler de connaissance spécifique, voire générique. La complétion de structures d'activité mal formées permet l'expression, aux bas niveaux, de cette connaissance générique, tout comme la complétion dans le cortex limbique ou para-limbique permet l'expression de la connaissance explicite. La complétion, deuxième propriété de la dynamique par attracteurs, achève donc la théorie de l'imagerie mentale en lui donnant à la fois un mécanisme et une raison d'être.

Quant à la détermination précise de l'aire la plus occipitale impliquée dans l'imagerie vivide ordinaire, les données actuelles semblent indiquer V4. Cette hypothèse est celle qui satisfait le mieux la variabilité des analyses par tomographie positronique et les arguments qui ont suivi[2]. Par ailleurs, la localisation spatiale des champs récepteurs de V4 est beaucoup plus précise que celle du cortex temporal, ce qui peut expliquer la différence entre imagerie "pictoriale" et "non-pictoriale". Dans cette optique, l'imagerie "abstraite" exclut toute activation occipitale, et l'imagerie eidétique peut correspondre à la réactivation de V2, voire du cortex strié.

Pour revenir sur le sujet du débat classique entre Kosslyn et Pylyshyn, il semble bien que l'imagerie soit impliquée dans le raisonnement. Dans la mesure où raisonner, c'est enchaîner des opérations mentales élémentaires, il n'y a pas de raison d'exclure l'imagerie du catalogue de ces opérations élémentaires. Et puisque l'imagerie est une reconstitution de représentations de nature perceptuelle, il suit que le raisonnement peut inclure des étapes non-symboliques. Comme modèle de l'humain, l'animal se substitue à l'ordinateur.

En conclusion de cette partie théorique, notre analyse insiste sur la correspondance entre la variété des types d'imagerie et l'organisation fonctionnelle des représentations corticales. Il y a donc plusieurs imageries, mais elles ne sont pas infiniment nombreuses.

VI.B Seconde partie: Modèle

La théorie présentée ci-dessus généralise les propriétés des réseaux à attracteurs sans apporter ni preuve formelle ni exemple complet de réseau encodant une représentation d'image et présentant les propriétés vantées (bistabilité et complétion). Le but de cette deuxième partie est de remédier à ce manque, en montrant un modèle qui illustre la théorie. Dans cette partie, on présente donc une couche de neurones implémentant une représentation de l'image, munie de connexions latérales qui en font un réseau à attracteurs. Comme attendu, on observe alors qu'il y a un seuil d'activation, et que la

[2]voir d'un côté Roland, Eriksson, Stone-Elander and Widen (1987), Decety, Kawashima, Gulyas and Roland (1992), ?), et Kawashita, Roland and O'Sullivan (1995); et de l'autre Goldenberg, Podreka, Steiner and Willmes (1987), Goldenberg, Podreka and Steiner (1990), Kosslyn, Alpert, Thompson, Maljkowic, Weise, Chabris, Hamilton, Rauch and Buonanno (1993), et Kosslyn and Ochsner (1994).

couche peut améliorer une représentation initialement mauvaise.

VI.B.1 Le modèle

La couche principale de notre modèle est une représentation par ondelettes[3]. Ces filtres prennent différentes positions (X,Y), différentes orientation, différentes phases (sinus/cosinus), et différentes échelles. En contraste avec les modèles linéaires, les neurones y transforment leurs entrées via une sigmoide, de sorte que la couche possède la non-linéarité qui est à l'origine de la bistabilité. Donc, chaque neurone de cette couche reçoit des informations en provenance d'une image en niveaux de gris, filtrée par Gabor, des informations latérales en provenance des autres neurones de la couche, et des informations en provenance de "tout le reste" du cortex, en l'occurrence résumé par une couche unique. Selon les simulations, certaines ou toutes de ces sources sont actives, seules les connexions latérales ne sont jamais coupées. Ces entrées sont ajoutées et appliquées dans l'intervalle [0,1] par une sigmoïde. L'architecture est illustrée par la Figure III.1, et le filtrage par la Figure III.4.

Les connexions latérales, élément essentiel de notre modèle, correspondent au recouvrement des champs recepteurs. Ainsi, deux neurones proches, dont les filtres ont la même orientation, alignés le long de cette orientation, seront fortement connectés par un poids positif. Inversement, deux neurones proches d'orientation opposée, alignés le long de cette orientation, seront mutuellement inhibiteurs. Des neurones éloignés ont des connexions quasi-nulles. Toutes les connexions ensemble sont normées par un paramètre de force globale, le "gain".

Les images utilisées pour les simulations sont montrées en Figure III.5. Ce sont principalement des images saturées (toutes noires et blanches), à l'exception d'un extrait de "la Joconde" utilisé surtout pour sa valeur démonstrative. Pour les simulations d'imagerie, le "reste du cortex" utilise une représentation identique à la représentation de référence de l'image-cible, mais bruitée. On peut voir en Figure III.1 que les niveaux de bruit utilisés altèrent considérablement la structure d'activité. Les résultats incluant une composante d'imagerie sont tous les moyennes de nombreuses simulations avec des

[3]Précisément, par filtres de Gabor, voir Figure III.2

bruitages différents.

Pour mesurer la qualité de la représentation, on utilise deux mesures. La première est une mesure de forme : c'est la corrélation avec un état de référence, fournie par la valeur de la couche principale en mode purement perceptif. La seconde est une mesure d'activité: c'est l'activité globale, calculée euclidiennement[4]. Il faut noter que la mesure de forme est indépendante du niveau d'activité de la référence et de la représentation mesurée[5].

La mise à jour du réseau s'effectue par cycles, tous les neurones de la couche principale étant mis à jour simultanément, en supposant que leurs entrées dépendent uniquement de la valeur de la couche au cycle précédant.

VI.B.2 Résultats

Les Figures III.6 à III.14 présentent les résultats suivants :

- **Mémoire statique :** initialisée avec l'influx perceptuel, la couche principale conserve une représentation de forme correcte pendant une dizaine de cycles.
- **Imagerie :** Recevant exclusivement un influx rétroactif (constitué de l'état de référence, bruité), la couche principale élimine une partie du bruit, améliorant l'image pendant les trois ou quatre premiers cycles.
- **Existence d'un seuil perceptuel :** L'entraînement de la mémoire statique décrite ci-dessus ne se fait que si l'entrée initiale est assez forte.
- **Existence d'un seuil pour l'imagerie :** L'amélioration et le maintient de "l'image mentale" ne se produit que si la rétroaction est suffisamment forte.
- **Dépendance des seuils :** La valeur de ces seuils est fonction (grossièrement hyperbolique) du gain des connexions latérales.

[4]Racine de la somme des carrés de l'activité.

[5]Si l'activité est strictement nulle, la corrélation est indéterminée et on choisit zéro par convention – ce n'est pas le cas dans les résultats présentés.

- **Combinaison positive de sources identiques :** Lorsque l'image visuelle et l'image cible d'imagerie sont les mêmes, il suffit que la somme des entrées dépasse le seuil pour que la couche principale s'active.
- **Non-combinaison de sources différentes :** Si l'image visuelle et l'image cible d'imagerie sont différentes, chacune doit dépasser le seuil séparément pour que la couche principale s'active.
- **Exemple de reconstitution d'image en niveau de gris :** Avec un réseau plus dense, on teste l'imagerie sur un extrait de "La Joconde".
- **Exemple de reconstitutions avec les images en noir et blanc :** Pour chaque image du jeu noir-et-blanc, on montre une simulation d'imagerie.

VI.B.3 Discussion et conclusion

Limites du modèle

Les neurones utilisés dans le modèle sont caricaturaux. Les études récentes montrent que les champs récepteurs ont une grande variété. Leur structure est spatio-temporelle plutôt que spatiale seulement. On peut proposer que chacun de nos neurones modèle en fait une micro-colonne, y compris les interneurones, de sorte que les connexions négatives sont justifiées et que les champs récepteurs sensibles en outre au mouvement dans différentes directions sont moyennés pour devenir indépendants du temps. Même dans ce cas, certaines propriétés d'une carte corticale ne sont pas restituées, en particulier l'élargissement des champs récepteurs quand on s'écarte de la fovéa. Néammoins, il semble à priori que ces propriétés négligées dans le modèle ne devraient pas avoir d'influence sur la dynamique, et donc sur les résultats qualitatifs.

Les connexions latérales sont un élément essentiel du modèle. Il apparaît, heureusement, que l'hypothèse de connexions entre microcolonnes correspondant grosso modo au recouvrement des champs récepteurs est réaliste (Gilbert & Wiesel, 1989; Bonhoeffer & Grinvald, 1991; Maldonado & Gray, 1996; Maldonado, Gödecke, Gray & Bonhoeffer, 1997; Kisvarday, Toth, Rausch & Eysel, 1997; Callaway, 1998). De plus, il semble que le cortex strié, au moins, possède un mécanisme d'ajustement de sa sensibilité (Gilbert & Wiesel, 1990; DeAngelis, Ohzawa & Freeman, 1995; Das & Gilbert,

1995b; Pei, Vidyasagar, Volgushev & Creutzfeld, 1994), de sorte que lorsque l'influx sensoriel disparaît, le rapport du gain des connexions à la sensibilité croît. Ainsi, le gain relatif de connectivité augmente précisément quand l'imagerie est bienvenue.

En résumé, ce modèle illustre la théorie présentée, et permet en outre quelques constatations. D'abord, deux paramètres déterminent conjointement l'activation ou non de l'aire considérée (et donc, le type d'image mentale formée) : premièrement, la capacité de former une image, qui dépend principalement du gain des connexions –si ce dernier est trop faible, aucune image ne peut être formée dans cette aire– deuxièmement, l'intensité de la stimulation. A priori, rien ne laisse supposer que nous ayons le moindre contrôle sur ce second paramètre[6], mais la *combinaison* de signaux de différentes sources, soit sensorielle plus interne, soit toutes internes, peut donner différentes intensités de stimulation. Ensuite, les attracteurs observés sont très nombreux. On doit parler de familles d'attracteurs plutôt que d'attracteurs individuels. Il semble probable qu'il y ait une gradation dans la numérosité des attracteurs, les aires corticales les plus occipitales en possédant le plus grand nombre, de façon à pouvoir représenter toutes les images possibles dans le monde, alors que les aires ultérieures en ont de moins en moins, permettant une classification de plus en plus catégorique. Finalement, ce modèle confirme une prédiction faite depuis longtemps par les avocats de l'imagerie visuelle : il doit y avoir entre vision et imagerie une interaction qui dépend de la similarité entre image mentale et image visuelle.

VI.C Troisième partie: Expérience

La troisième partie présente une expérience qui teste cette dernière prédiction: lorsque l'on combine perception et visualisation, leur interaction dépend du contenu des images. Si les images mentales et visuelles sont similaires, on s'attend à une interaction bénéfique. Sinon, on s'attend à une interférence.

Les études existantes de la relation entre imagerie et perception peuvent être classées en trois grandes catégories :

[6]Et même, puisqu'on suppose que les aires plus élevées dont provient le signal ont le même type de dynamique, la stimulation est ou n'est pas – il n'y a pas d'autre modulation possible.

- Des expériences qui s'efforcent de démontrer l'existence d'une interaction entre les *processus* de perception et d'imagerie, indépendemment du contenu des images. Ces études (Segal & Glicksman, 1967; Segal & Gordon, 1969; Segal & Fusella, 1969; Segal & Fusella, 1970; Segal, 1971; Reeves & Segal, 1973; Schorr, Balzano & Smith, 1978; Reeves, 1980; Reeves, 1981; Craver-Lemley & Reeves, 1987; Craver-Lemley & Reeves, 1992; Craver-Lemley, Arterberry & Reeves, 1997) ont commencé à partir d'une réplication du travail de Perky (1910), qui montrait que les sujets confondaient une image visuelle dégradée avec une image mentale demandée. Malheureusement, les toutes premières études ne contrôlaient pas une explication alternative de la moindre performance observée en situation d'imagerie : l'effet de l'effort dans les conditions de visualisation. Les études ultérieures sont munies de conditions de contrôle, mais simultanément, on voit disparaître la comparaison entre le cas où l'image mentale est similaire à l'image visuelle et le cas où elles sont différentes.

- Des expériences de psychophysique qui visent à répliquer avec un contexte imaginé des effets obtenus d'ordinaire avec le même contexte, mais vu. De ces études (Over & Broerse, 1972; Berbaum & Chung, 1981; Wallace, 1984; Rhodes & O'Leary, 1985; Ishai & Sagi, 1995) il semble ressortir que les effets obtenus d'ordinaire par la fatigue induite par contemplation longue ne sont pas répliqués dans le cas de l'imagination longue du même contexte. Par contre, les effets induits immédiatement par la présence d'un contexte dans l'image sont obtenus aussi lorsque ce contexte est imaginé.

- Des expériences qui se concentrent sur l'interférence entre images mentales et visuelles, et sont donc directement pertinentes pour notre question. Ces expériences (Peterson & Graham, 1974; Farah, 1983; Farah, 1985b; Heil, Rössler & Hennighausen, 1993) souffrent néammoins toutes à divers degrés de la même faiblesse : les résultats observés pourraient être expliqués par l'adaptation, dans le cas où les images mentale et visuelles sont similaires, de stratégies d'acquisition de l'information. Par exemple, l'effort de visualisation peut provoquer un positionnement des yeux sur les zones les plus pertinentes pour identifier cette image

particulièrement, à l'exclusion des autres. Si le sujet fixe un point central, un déplacement de l'attention couverte est encore possible. Si le stimulus peut susciter des saccades après sa projection, dans le cadre d'un processus d'indentification, ces saccades peuvent être accélérées par la connaissance préalable des cibles, même si leur déclanchement est retardé par une obligation de fixation initiale. Ainsi, plusieurs stratégies d'acquisition d'information peuvent être influencées par une visualisation préalable d'une façon qui accélère l'identification de l'image visuelle seulement lorsqu'elle est similaire à l'image mentale préalablement formée. A cause de cette confusion possible, aucune des expériences existantes n'apporte de preuve solide que l'interaction entre vision et imagerie est une affaire de contenus plutôt que de processus.

Dans ce contexte, notre expérience visait à apporter une démonstration plus convaincante, en incluant une condition de contrôle où le sujet pouvait voir ce qu'il avait à imaginer dans la condition principale.

VI.C.1 Abrégé des méthodes

Chacun de nos soixante-trois sujets répondait à cent-soixante essais répartis entre deux types. Dans la condition principale, un essai comporte une phase de visualisation (lecture d'instructions puis temps mort) et une phase de réponse à un stimulus qui s'affiche sur écran. Dans la condition de contrôle, la lecture des instructions est remplacée par la vision de l'image qui correspond aux instructions. Le stimulus proprement dit est une image de cadran d'horloge avec deux aiguilles, couvert par un demi-plan ombré. La réponse se fait en pressant un certain bouton si les deux aiguilles sont dans l'ombre, et un autre bouton si une ou deux aiguilles sont non couvertes. Le demi-plan vient à angle aléatoire. La visualisation ou la vision préalable ne concerne que l'horloge et ses aiguilles. La Figure IV.2 illustre un essai de la condition principale. Notez que la tâche ne permet pas de favoriser à priori un bouton ou l'autre, puisque le demi-plan est aléatoire. Par ailleurs, l'utilisation des temps permet une visualisation à la fois modérément compliquée et offrant de nombreuses situations ce qui permet d'éviter la répétition des stimuli. Enfin, les instructions évitent soigneusement de présenter la tâche comme une question à sens unique (oui/non), pour éviter autant que possible ce champ sémantique, déjà par-

tiellement accédé par la similarité entre les horloges. On mesure le temps de réaction de la réponse.

Pour chaque sujet, trois facteurs se croisent : la condition, principale (imagerie) ou de contrôle (vision) ; la réponse attendue (aiguilles dans l'ombre ou pas) ; et la congruence de l'image préalable avec l'horloge du stimulus (même horloge/ horloge différente). La Figure IV.6 montre les cas de figure découlant de cette architecture. Les temps de réactions des cellules élémentaires sont moyennés après exclusion des temps écartés de plus de deux déviations standard initiales[7]. La plupart des analyses sont faites sur la *facilitation*, c'est à dire la différence, tous autres facteurs égaux, entre les temps de réaction lorsque les horloges sont similaires et différentes. L'utilisation de cette différence permet d'éliminer la variabilité en provenance du sujet et de l'effort particulier à chaque condition.

D'un sujet à l'autre, on a varié la main utilisée pour la réponse et l'ordre des conditions (vision ou imagerie d'abord, ou, pour une minorité de sujets, alternance des essais principaux et de contrôle). On a également noté si les sujets étaient masculins ou féminins, droitiers ou gauchers, et enfin, on a remarqué la différence entre les sujets participant pendant le trimestre ordinaire, et ceux qui ont participé durant la semaine d'examens (ce dernier facteur correspondant à un niveau de fatigue moyen différent).

En outre, les sujets ont été soumis à trois prétest standard d'imagerie mentale, avant l'expérience proprement dite, et à une entrevue d'interrogation puis d'explication après l'expérience.

VI.C.2 Abrégé des résultats et discussion

En préliminaire, un certain nombre de résultats nuls ou sans surprise permettent de simplifier les problèmes considérés. Premièrement, les prétests ne sont corrélé avec aucun résultat, sinon la vitesse de réaction globale de chaque sujet. Il semble donc que ces tests mesurent l'état d'attention et l'efficacité du sujet plutôt qu'une quelconque capacité de visualisation. Deuxièmement, les temps de réaction bruts montrent que globalement la condition de visualisation est beaucoup plus lente, mais cet effet (113ms)

[7] Cette procédure améliore la fiabilité des résultats mais n'en change ni le sens ni la magnitude.

disparaît presque lorsque les essais des deux types sont alternés (20ms alors). Si l'effet global semble donner raison à ceux qui pensent que les processus de vision et d'imagerie interfèrent, sa disparition, au contraire, confirme que l'origine de cet effet est sans doute *l'effort* fait pour visualiser. Il n'est pas évident, et il est même probablement faux, que la qualité et la duration de visualisation dépende de l'intensité de l'effort. L'effort, qui est donc sans doute le facteur observé dans ce ralentissement, apparaît comme un phénomène parasite qui trahit le manque de technique de visualisation dont nous souffrons probablement tous. Finalement, l'observation des taux d'erreurs confirme que les résultats les plus lents coïncident avec les taux d'erreurs les plus hauts. Il y a donc accord entre ces deux sources d'information[8], et pour faciliter les choses, on a basé l'analyse sur les temps de réaction[9]. En somme, ces considérations préliminaires n'ont pas apporté de surprises.

Le principal résultat observé est que la facilitation est *plus forte* dans la condition principale que dans la consition de contrôle. Visualiser la bonne horloge est donc plus efficace que de l'avoir vue d'avance. Une analyse postérieure séparant non pas deux mais trois cas de stimulus (zéro, une et deux aiguilles dans le gris), jette une lumière crue sur ce résultat. Voir d'avance la bonne horloge semble n'avoir d'effect facilitatif *que si* il y a une aiguille dans le gris et l'autre non couverte. Très probablement, ce qu'on observe là est un effet de stratégie d'acquisition d'information. Dans ce cas, l'image est un peu plus complexe, et requiert parfois une saccade, qui doit être plus rapide lorsqu'on s'attend à la bonne horloge. Si l'on visualise la bonne horloge, la facilitation est également la plus forte dans ce cas, mais elle est encore présente dans les deux autres cas. Globalement, ce résultat suggère que la facilitation se produit à bas niveau, puisque la sémantique est la même entre la condition principale et la condition de contrôle. On obtient donc le résultat principal attendu pour cette expérience : une interaction qui dépend du contenu de l'image mentale.

[8]De façon générale, si les instructions insistent sur la vitesse de réaction, on verra davantage d'information s'inscrire sur les taux d'erreurs, et si les instructions insistent sur la fiabilité de la réponse, le temps de réaction est la mesure la plus significative. Dans notre expérience nous avons évité de contraindre ce facteur, car si on le fait on arrive parfois à des situations ou une des sources contredit l'autre, parce que certaines conditions sont plus rapides au point que les sujets y font aussi plus d'erreurs.

[9]Une analyse double céderait sans doute des taux de fiabilité plus élevés, mais cela n'a heureusement pas été nécessaire.

Le deuxième résultat important était inattendu. C'est une interaction entre trois facteurs extrêmement robuste. Pour l'imagerie seulement, et pour une réponse "non", c'est à dire lorsque une ou deux aiguilles ne sont pas couvertes, la facilitation est nulle (-5ms) lorsque la réponse est faite de la main gauche, et considérable (50ms) lorsque la réponse est faite de la main droite. Cet effet est identique pour les deux types de "non" (une ou deux aiguilles hors du gris). Nous l'interprétons comme un effet à plus haut niveau. On sait depuis les revues de Farah (1984) et Kosslyn (1994) que l'imagerie requiert davantage l'hémisphère gauche, analytique, que l'hémisphère droit. En particulier, la visualisation d'images composites, dont notre horloge est un archétype puisqu'on y place deux aiguilles indépendamment spécifiées, est la fonction exclusive de l'hémisphère gauche. La réponse "non", plus complexe, parce que disjonctive, fait sans doute également davantage appel à l'hémisphère gauche que la réponse "oui". Nous ignorons comment ces deux processus peuvent interférer exactement. Néammoins, cet effet confirme les données qui suggèrent une forte latéralisation de l'imagerie.

A part ces résultats principaux, on remarque quelques résultats sans surprise. Les sujets fatigués manifestent une moindre facilitation, en particulier dans la condition principale (imagerie). Les mâles sont plus latéralisés que les femelles. Les sujets ayant répondu de la main droite ont plus de facilitation que ceux ayant utilisé la main gauche.

La facilitation n'est absolument pas corrélée avec le temps de réaction moyen. Les résultats observés sont donc authentiques, les variations de facilitation ne résultent pas d'effets de limite.

En conclusion, cette expérience comble un manque dans le domaine : il ya maintenant une démonstration d'interférence dépendant du contenu munie d'une condition de contrôle convaincante. De plus, cette interférence semble être de bas niveau, perceptuelle.

VI.D Conclusion

En conclusion générale, notre théorie propose un système visuel dont les sous-systèmes sont plus indépendants qu'on veut d'ordinaire l'admettre. La principale conséquence de cette indépendance est que l'imagerie est variée. Plusieurs types d'imagerie correspon-

dent à des niveaux de plus en en plus concrets de détail. Le rôle de l'imagerie est donc le rôle mirroir de l'abstraction d'information faite lors de la vision –et au-delà. Dans l'abstraction, on perd les détails au profit de l'assimilation à une catégorie plus générale. Dans l'imagerie, on restitue ces détails, au dépens de la généralité. Par conséquent, imagerie et abstraction sont les deux opérations les plus élémentaires du raisonnement, car elles permettent un va-et-vient entre plusieurs niveaux de représentation. D'une certaine façon, parce qu'elle permet le plus simple enchaînement d'opérations mentales, l'imagerie est le pivot de la cognition, le point d'articulation entre perception et raisonnement.

Bibliography

[Anderson, Silverstein, Ritz and Jones (1977] Anderson, J., Silverstein, J., Ritz, S., & Jones, R. (1977). Distinctive features, categorical perception, and probability learning: some applications of a neural model. *Psychological Review*, *84*, 413–451.

[Arditi, Holtzman and Kosslyn (1988] Arditi, A., Holtzman, J., & Kosslyn, S. (1988). Mental imagery and sensory experience in congenital blindness. *Neuropsychologia*, *26*, 1–12.

[Behrman, Winocur and Moscovitch (1992] Behrman, M., Winocur, G., & Moscovitch, M. (1992). Dissociation between mental imagery and object recognition in a brain-damaged patient. *Nature*, *359*, 636–637.

[Berbaum and Chung (1981] Berbaum, K. & Chung, C. (1981). Müller-Lyer illusion induced by imagination. *Journal of Mental Imagery*, *5*, 125–128.

[Betts (1909] Betts, G. (1909). *The distributions and functions of mental imagery.* Columbia University: Columbia University.

[Biederman (1987] Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*, 115–147.

[Bisiach and Luzzatti (1979] Bisiach, E. & Luzzatti, C. (1979). Unilateral neglect representational schema, and consciousness. *Brain*, *102*, 609–618.

[Block (1981] Block, N. (1981). *Imagery.* Cambridge, Mass.: MIT Press.

[Bonhoeffer and Grinvald (1991] Bonhoeffer, T. & Grinvald, A. (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, *353*, 429–431.

[Borisyuk and Kirillov (1992] Borisyuk, R. & Kirillov, A. (1992). Bifurcation Analysis of a Neural Network Model. *Biological Cybernetics*, *66*, 319–325.

[Broerse and Crassini (1980] Broerse, J. & Crassini, B. (1980). The influence of imagery ability on color aftereffects produced by physically present and imaged induction stimuli. *Perception and Psychophysics*, *28*, 560–568.

[Brooks (1968] Brooks, L. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, *22*, 349–368.

[Bullier and Nowak (1995] Bullier, J. & Nowak, L. (1995). Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Current Opinion in Neurobiology*, *5*, 497–503.

[Callaway (1998] Callaway, E. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience*, in press.

[Churchland, Ramachandran and Sejnowski (1994] Churchland, P., Ramachandran, V., & Sejnowski, T. (1994). A Critique of Pure Vision. In C. Koch & J. Davis (Eds.), *Large-scale Theories of the Cortex*. Cambridge, MA: MIT Press.

[Cooper (1975] Cooper, L. (1975). Mental rotation of random two-dimensional shapes, in Cognitive Psychology. *Cognitive Psychology*, *7*, 20–43.

[Corbetta, Miezin, Dobmeyer, Shulman and Petersen (1991] Corbetta, M., Miezin, F., Dobmeyer, S., Shulman, G., & Petersen, S. (1991). Selective and divided attention during visual discriminations of shape, color and speed: functional anatomy by positron emission tomography. *Journal of Neuroscience*, *11*, 2383–2402.

[Craik and Dirkx (1992] Craik, F. & Dirkx, E. (1992). Age-related differences in three tests of visual imagery. *Psychology and aging*, *4*, 661–665.

[Craik (1940] Craik, K. (1940). Origin of visual after-images. *Nature*, *145*, 512.

[Craver-Lemley, Arterberry and Reeves (1997] Craver-Lemley, C., Arterberry, M., & Reeves, A. (1997). Effects of imagery on Vernier acuity under conditions of induced depth. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 3–13.

[Craver-Lemley and Reeves (1987] Craver-Lemley, C. & Reeves, A. (1987). Visual imagery reduces vernier acuity. *Perception*, *16*, 599–614.

[Craver-Lemley and Reeves (1992] Craver-Lemley, C. & Reeves, A. (1992). How visual imagery interferes with vision. *Psychological Review*, *99*, 633–649.

[Crick and Koch (1998] Crick, F. & Koch, C. (1998). Consciousness and neuroscience. *Cerebral cortex*, *8*, 97–107.

[Das and Gilbert (1995a] Das, A. & Gilbert, C. (1995a). Long-range horizontal connections and their role in cortical reorganisation revealed by optical recording of cat visual cortex. *Nature*, *375*, 780–784.

[Das and Gilbert (1995b] Das, A. & Gilbert, C. (1995b). Receptive field expansion in adult visual cortex is linked to dynamic changes in strength of cortical connections. *Journal of Neurophysiology*, *74*, 779–792.

[Davidson and Schwartz (1977] Davidson, R. & Schwartz, G. (1977). Brain mechanisms subserving self-generated imagery: electrophysiological specificity and patterning. *Psychophysiology*, *14*, 598–601.

[DeAngelis, Anzai, Ohzawa and Freeman (1995] DeAngelis, G., Anzai, A., Ohzawa, I., & Freeman, R. (1995). Receptive field structure in the visual cortex: Does selective stimulation induce plasticity? *Proceedings of the National Academy of Sciences, USA*, *92*, 9682–0686.

[DeAngelis, Ohzawa and Freeman (1995] DeAngelis, G., Ohzawa, I., & Freeman, R. (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neuro-Science*, *18*, 451–458.

[Decety, Kawashima, Gulyas and Roland (1992] Decety, J., Kawashima, R., Gulyas, B., & Roland, P. (1992). Preparing for reaching: a PET study of the participating structures in the human brain. *Neuroreport*, *3*, 761–764.

[Desimone and Duncan (1995] Desimone, R. & Duncan, J. (1995). Visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

[Desimone and Ungerleiter (1990] Desimone, R. & Ungerleiter, L. (1990). Neural mechanisms of visual processing in monkeys. In I. Boller & J. Gratman (Eds.), *Handbook of Neuropsychology*, volume 2. NY, NY: Elsevier.

[Doob (1965] Doob, L. (1965). Exploring eidetic imagery among the Kamba of central Kenya. *Journal of Social Psychology*, *67*, 3–22.

[Droulez and Berthoz (1990] Droulez, J. & Berthoz, A. (1990). The concept of dynamic memory in sensorimotor control. In D. Humphrey & H. Freund (Eds.), *Freedom to move: dissolving boundaries in motor control*, pages 137–161. England: Wiley.

[Farah (1983] Farah, M. (1983). Perceptual interference and facilitation with auditory imagery. *Perception and Psychophysics*, *33*, 475–478.

[Farah (1984] Farah, M. (1984). The neurological basis of mental imagery: a componential analysis. *Cognition*, *18*, 245–272.

[Farah (1985a] Farah, M. (1985a). A left hemisphere basis for visual mental imagery? *Neuropsychologica*, *23*, 115–118.

[Farah (1985b] Farah, M. (1985b). Psychophysical evidence for a shared representational medium for mental images and percepts. *Journal of Experimental Psychology: General*, *114*, 91–103.

[Farah (1988] Farah, M. (1988). Is Visual Imagery Really Visual ? Overlooked Evidence from Neuropsychology. *Psychological Review*, *95*, 307–317.

[Farah (1989a] Farah, M. (1989a). Mechanisms of imagery-Perception Interaction. *Journal of Experimental Psychology: Human Attention and Performance*, *15*, 203–211.

[Farah (1989b] Farah, M. (1989b). The Neural Basis of Mental Imagery. *Trends in Neurosciences*, *12*, 395–399.

[Farah (1989c] Farah, M. (1989c). The neuropsychology of Mental Imagery. In J. W. Brown (Ed.), *Neuropsychology of visual perception*, chapter 9, pages 183–201. Lawrence Erlbaum Associates.

[Farah, Peronnet, Gonon and Girard (1988] Farah, M., Peronnet, F., Gonon, M., & Girard, M. (1988). Electrophysiological Evidence for a shared representational medium for visual images and visual percepts. *Journal of Experimental Psychology:: General*, *117*, 248–257.

[Farah, Soso and Dasheiff (1992] Farah, M., Soso, M., & Dasheiff, R. (1992). Visual angle of the mind's eye before and after unilateral occipital lobectomy. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 241–246.

[Felleman and Van Essen (1991] Felleman, D. & Van Essen, D. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1–47.

[Fender and Julesz (1967] Fender, D. & Julesz, B. (1967). *J.Opt.Soc.Am*, *57*, 819.

[Field (1993] Field, D. (1993). Scale-invariance and self-similar 'Wavelet' transforms: an analysis of natural scenes and mammalian visual system. In M. Farge, J. Hunt, & J. Vassilicos (Eds.), *Wavelets, fractals, and Fourier transforms*, volume 43, pages 151–193. The Institute of Mathematics and Its Applications conference series: Oxford ; New York.

[Finke (1985] Finke, R. (1985). Theories relating mental imagery to perception. *Psychological Bulletin*, *98*, 236–259.

[Finke and Schmidt (1977] Finke, R. & Schmidt, M. (1977). Orientation-specific color aftereffects following imagination. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 599–606.

[Finke and Schmidt (1978] Finke, R. & Schmidt, M. (1978). The quantitative measure of pattern perception in images using orientation-specific color aftereffects. *Perception and Psychophysics*, *23*, 5515–520.

[Freeman (1975] Freeman, W. (1975). *Mass action in the nervous system.* NY, NY: Academic Press.

[Freeman (1978] Freeman, W. (1978). Spatial properties of an EEG event in the olfactory bulb and cortex. *Electroencephalography and Clinical Neurophysiology*, *44*, 586–605.

[Freeman (1983] Freeman, W. (1983). The physiological basis of mental images, Biological Psychiatry. *Biological Psychiatry*, *18*, 1107–1125.

[Freeman (1994a] Freeman, W. (1994a). Characterizations of state transitions in spatially distributed, chaotic, nonlinear, dynamical systems in cerebral cortex. *Integrative Physiological and Behavioral Science*, *29*, 294–306.

[Freeman (1994b] Freeman, W. (1994b). Neural Networks and Chaos. *Journal of Theoretical Biology*, *171*, 13–18.

[Freeman and Baird (1987] Freeman, W. & Baird, B. (1987). Relation of olfactory EEG to behavior: spatial analysis. *Behavioral Neuroscience*, *101*, 393–408.

[Freeman and Grajski (1987] Freeman, W. & Grajski, K. (1987). Relation of olfactory EEG to behavior: functional analysis. *Behavioral Neuroscience*, *101*, 766–777.

[Freeman and Skarda (1990] Freeman, W. & Skarda, C. (1990). Chaotic dynamics versus representationalism. *Behavioral and Brain Sciences*, *13*, 167–168.

[Freeman and van Dijk (1987] Freeman, W. & van Dijk, B. (1987). Spatial patterns of visual cortical fast EEGduring conditioned reflex in a rhesus monkey. *Brain Research*, *442*, 267–276.

[Freeman and Viana Di Prisco (1986] Freeman, W. & Viana Di Prisco, G. (1986). Relation of olfactory EEG to behavior: time series analysis. *Behavioral Neuroscience*, *100*, 753–763.

[Galton (1907] Galton, F. (1907). *Inquiries into human faculty and its development.* London, UK: Dent and sons Ltd.

[Gibson (1979] Gibson, J. (1979). *Ecological approach to visual perception.* Boston, MA: Houghton Mifflin.

[Gilbert and Wiesel (1989] Gilbert, C. & Wiesel, T. (1989). Columnar specificity of intrinsic horizontal and cortico-cortical connections in Cat Visual Cortex. *Journal of Neuroscience*, *9*, 2432–2442.

[Gilbert and Wiesel (1990] Gilbert, C. & Wiesel, T. (1990). The influence of contextual stimuli on the orientation selectivity of cells in the primary visual cortex of the cat. *Vision Research*, *30*, 1689–1701.

[Giray and al. (1985] Giray, E. & al. (1985). A life span approach to the study of eidetic imagery. *Journal of Mental Imagery*, *9*, 21–32.

[Goldenberg, Podreka and Steiner (1990] Goldenberg, G., Podreka, I., & Steiner, M. (1990). The cerebral localization of visual imagery: Evidence from emission, computerized tomography of cerebral blood flow. In R. J. Hampson PJ, Marks DF (Ed.), *Imagery: Current developments. International library of psychology*, pages 307–332. Routledge, London, England.

[Goldenberg, Podreka, Steiner and Willmes (1987] Goldenberg, G., Podreka, I., Steiner, M., & Willmes, K. (1987). Patterns of regional cerebral blood flow related to memorizing of high and low-imagery words- an emission computer tomography study. *Neuropsychologia*, *25*, 473–485.

[Golla and Antonovitch (1929] Golla, F. & Antonovitch, S. (1929). The relaxation of muscular tonus and the patellar reflex to mental work. *Journal of Mental Science*, *75*, 234–241.

[Golla, Hutton and Walter (1943] Golla, F., Hutton, E., & Walter, W. (1943). The objective study of mental imagery I. physiological concomitants. *Journal of Mental Science*, *89*, 216–233.

[Gomez Gonzales, Clark, Fan, Luck and Hillyard (1994] Gomez Gonzales, C., Clark, V., Fan, S., Luck, S., & Hillyard, S. (1994). Sources of Attention-Sensitive Visual Event-Related Potentials. *Brain Topography*, *7*, 41–51.

[Gray, Freeman and Skinner (1986] Gray, C., Freeman, W., & Skinner, J. (1986). Chemical dependancies of learning in the rabbit olfactory bulb: acquisition of the transient spatial pattern change depends on norepinephrine. *Behavioral Neuroscience*, *100*, 585–596.

[Gray and Singer (1989] Gray, C. & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientations columns of cat visual cortex. *Proceedings of the National Academy of Science USA*, *86*, 1698–1702.

[Grossberg (1973] Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, *LII*, 213–257.

[Haber (1988] Haber, R. (1988). Eidetic imagery. In O. L & F. D (Eds.), *The exceptional brain: Neuropsychology of talen and special abilities*. Guildford Press, NY, NY.

[Haenny and Schiller (1988] Haenny, P. & Schiller, P. (1988). State-dependant activity in monkey visual cortex. I.single cell activity in V1 and V4 on visual tasks. *Experimental Brain Research*, *69*, 225–244.

[Hebb (1949] Hebb, D. (1949). *The organisation of behavior*. NY, NY: Wiley.

[Hebb (1968] Hebb, D. (1968). Concerning imagery. *Psychological Review*, *75*, 466–477.

[Heil, Rössler and Hennighausen (1993] Heil, M., Rössler, F., & Hennighausen, E. (1993). Imagery-Perception interaction depends on the shape of the image: a reply to Farah (1989). *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 1313–1320.

[Hilgetag, O'Neil and MP (1996] Hilgetag, C., O'Neil, M., & MP, Y. (1996). Indeterminate organization of the visual system. *Science*, *271*, 776–777.

[Hopfield (1982] Hopfield, J. (1982). Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *proceedings of the NAS*, *79*, 2554–2558.

[Hopfield (1984] Hopfield, J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *proceedings of the NAS*, *81*, 3088–3092.

[Horowitz (1978] Horowitz, M. (1978). *Image formation and cognition*. New York: Appleton-Century-Crofts.

[Hubel and Wiesel (1962] Hubel, D. & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*, 106–154.

[Hubel and Wiesel (1977] Hubel, D. & Wiesel, T. (1977). Ferrier lecture: Functional architecture of the macaque monkey visual cortex. *Proc. R. Soc. Lond.*, *198*, 1–59.

[Huxley (1957] Huxley, A. (1957). In M. Horowitz & C. Palmer (Eds.), *Moksha - Writings on psychedelics and the visionnary experience*. New York: Stonehill.

[Imbert and Buser (1987] Imbert, M. & Buser, P. (1987). *Neurphysiologie fonctionnelle IV: vision.* Paris, France: Hermann.

[Ishai and Sagi (1995] Ishai, A. & Sagi, D. (1995). Common mechanisms of visual imagery and perception. *Science*, *268*, 1772–1774.

[James (1890] James, W. (1890). *The principles of Psychology.* USA: Henry Holt and Co.

[Jankowiak, Kinsbourne, Shalev and Bachman (1992] Jankowiak, J., Kinsbourne, M., Shalev, R., & Bachman, D. (1992). Preserved visual imagery and categorization in a case of associative visual agnosia. *Journal of Cognitive Neuroscience*, *4*, 119–131.

[Johnson Laird (1983] Johnson Laird, P. (1983). *Mental models.* Cambridge, MA: Harvard University press.

[Kawashita, Roland and O'Sullivan (1994] Kawashita, R., Roland, P., & O'Sullivan, B. (1994). Fields in human motor areas involved in preparation for reaching, actual reaching and visuomotor learning: a positron emission tomography study. *Journal of Neuroscience*, *14*, 3462–3474.

[Kawashita, Roland and O'Sullivan (1995] Kawashita, R., Roland, P., & O'Sullivan, B. (1995). Functional anatomy of reaching and visuomotor learning: a positron emission tomography study. *Cerebral Cortex*, *5*, 111–122.

[Kerszberg, Dehaene and Changeux (1992] Kerszberg, M., Dehaene, S., & Changeux, J. (1992). Stabilisation of complex input-output functions in neural clusters formed by synapse selection. *Neural Networks*, *5*, 403–413.

[Kisvarday, Toth, Rausch and Eysel (1997] Kisvarday, Z., Toth, E., Rausch, M., & Eysel, U. (1997). Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. *Cerebral Cortex*, *7*, 605–618.

[Koch and Braun (1996] Koch, C. & Braun, J. (1996). Towards the neuronal correlate of visual awareness. *Current Opinion in Neurobiology*, *6*, 158–164.

[Kosslyn (1973] Kosslyn, S. (1973). Scanning mental images: some structural implications. *Perception and Psychophysics*, *14*, 90–94.

[Kosslyn (1994] Kosslyn, S. (1994). *Image and Brain, the resolution of the imagery debate.* Cambridge, Ma: MIT Press.

[Kosslyn, Alpert, Thompson, Maljkowic, Weise, Chabris, Hamilton, Rauch and Buonanno (1993] Kosslyn, S., Alpert, N., Thompson, W., Maljkowic, V., Weise, S., Chabris, C., Hamilton, S., Rauch, S., & Buonanno, F. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience*, *5*, 263–287.

[Kosslyn, Holtzman, Farah and Gazzaniga (1985] Kosslyn, S., Holtzman, J., Farah, M., & Gazzaniga, M. (1985). A computational analysis of mental image generation: evidence from functional dissociation in split-brain patients. *Journal of Experimental Psychology: General*, *114*, 311–341.

[Kosslyn and Ochsner (1994] Kosslyn, S. & Ochsner, K. (1994). In search of occipital activation during visual mental imagery. *Trends in Neurosciences*, *17*, 290–292.

[Kunzendorf (1984] Kunzendorf, R. (1984). Centrifugal effects of eidetic imaging on glash electroretinograms and autonomic responses. *Journal of mental imagery*, *8*, 67–75.

[Le Bihan, Turner, Zeffiro, Cuenod, Jezzard and Bonnerot (1993] Le Bihan, D., Turner, R., Zeffiro, T., Cuenod, C., Jezzard, P., & Bonnerot, V. (1993). Activation of human primary visual cortex during visual recall: a magnetic resonance imaging study. *Proceedings of the National Academy of Science*, *90*, 11802–11805.

[Levine, Warach and Farah (1985] Levine, D., Warach, J., & Farah, M. (1985). Two visual systems in mental imagery: dissociation of 'what' and 'where' in imagery disorders due to bilateral posterior cerebral lesions. *Neurology*, *35*, 1010–1018.

[Lowe (1985] Lowe, D. (1985). *Perceptual organisation and visual recognition*. Boston: Kluwer.

[Luria (1968] Luria, A. (1968). *The mind of a mnemonnist, translated from russian by Lynn Solotarof*. New York: Basic Books.

[Maier, Dgnelie, Spekreijse and van Dijk (1987] Maier, J., Dgnelie, G., Spekreijse, H., & van Dijk, B. (1987). Principal component analysis for source localisation of visual evoked potential in man. *Vision Research*, *12*, 155–177.

[Maldonado, Gödecke, Gray and Bonhoeffer (1997] Maldonado, P., Gödecke, I., Gray, C., & Bonhoeffer, T. (1997). Orientation selectivity in Pinwheel Centers in Cat Striate Cortex. *Science*, *276*, 1551–1555.

[Maldonado and Gray (1996] Maldonado, P. & Gray, C. (1996). Heterogeneity in local distributions of orientation-selective neurons in the cat primary visual cortex. *Visual Neuroscience*, *13*, 509–516.

[Marr (1982] Marr, D. (1982). *Vision*. NY, NY: Freeman.

[McClelland (1979] McClelland, J. (1979). On the time-relations of mental processes: an examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.

[McLeod, Driver and Crisp (1988] McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature*, *332*, 154–155.

[Miller and Peacock (1982] Miller, S. & Peacock, R. (1982). Evidence for the uniqueness of eidetic imagery. *Perceptual and motor skills*, *55*, 1219–1233.

[Miyashita (1993] Miyashita, Y. (1993). Inferior temporal cortex: where visual memory meets perception. *Annual Review of Neuroscience*, *16*, 245–263.

[Moran and Desimone (1985] Moran, J. & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*, 782–784.

[Moscovitch, Behrman and Winocur (1994] Moscovitch, M., Behrman, M., & Winocur, G. (1994). Do PETS have long or short ears? Mental imagery and neuroimaging. *Trends in Neurosciences*, *17*, 292–294.

[Nakamura, Mikami and Kubota (1992] Nakamura, K., Mikami, A., & Kubota, K. (1992). Oscillatory neuronal activity related to visual short-term memory in monkey temporal pole. *Neuroreport*, *3*, 117–120.

[Oswald (1957] Oswald, I. (1957). Afterimages from retina and brain. *Quarterly Journal of Experimental Psychology*, *9*, 88–100.

[Over and Broerse (1972] Over, R. & Broerse, J. (1972). Imaged lines fail to induce contour masking. *Psychonomic Science*, *29*, 203–204.

[Paivio and Cohen (1979] Paivio, A. & Cohen, M. (1979). Eidetic imagery and cognitive abilities. *Journal of Mental Imagery*, *3*, 53–64.

[Pei, Vidyasagar, Volgushev and Creutzfeld (1994] Pei, X., Vidyasagar, T., Volgushev, M., & Creutzfeld, O. (1994). Receptive field analysis and orientation selectivity of postsynaptic potentials of simple cells in cat visual cortex. *Journal of Neuroscience*, *14*, 7130–7140.

[Perky (1910] Perky, C. (1910). An experimental study of imagination. *American Journal of Psychology*, *21*, 422–452.

[Peterson and Graham (1974] Peterson, M. & Graham, S. (1974). Visual detection and visual imagery. *Journale of Experimental Psychology*, *103*, 509–514.

[Pitts and McCulloch (1947] Pitts, W. & McCulloch, W. (1947). How we know universals: the perception of auditory and visual forms. *Bulletin of Mathematical Biophysics*, *9*, 9:127–147.

[Podgorny and Shepard (1978] Podgorny, P. & Shepard, R. (1978). Functional Representations common to visual perception and imagination. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 21–35.

[Press, Teukolsky, Vetterling and Flannery (1992] Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (1992). *Numerical Recipes in C, second edition*. New York, NY: Cambridge University Press.

[Purdy (1936] Purdy, D. (1936). Eidetic imagery and plasticity of perception. *Journal of General Psychology*, 437–455.

[Pylyshyn (1973] Pylyshyn, Z. (1973). What the mind's eye tells the mind's brain: a critique of mental imagery. *Psychological Bulletin*, *80*, 1–24.

[Pylyshyn (1986] Pylyshyn, Z. (1986). Mental imagery and functional architecture. In Z. Pylyshyn (Ed.), *Computation and cognition: toward a foundation for cognitive science*, chapter 8, pages 225–283. Cambridge, MA: MIT press.

[Rall (1981] Rall, W. (1981). Cable theory for dendritic neurons. In C. Koch & I. Segev (Eds.), *Methods in neuronal modeling, from synapses to networks*, volume 1, chapter 2, pages 9–62. Cambridge, Mass.: MIT Press.

[Reeves (1980] Reeves, A. (1980). Visual imagery in backward masking. *Perception and Psychophysics*, *28*, 118–124.

[Reeves (1981] Reeves, A. (1981). Visual imagery lowers sensitivity to hue-varying, but not to luminance-varying, visual stimuli. *Perception and Psychophysics*, *29*, 247–250.

[Reeves and Segal (1973] Reeves, A. & Segal, S. (1973). Effects of visual imagery in visual sensitivity and pupil diameter. *Perceptual and motor skills*, *36*, 1091–1098.

[Reppas, Niyogi, Dale, Sereno and RBH (1997] Reppas, J., Niyogi, S., Dale, A., Sereno, M., & RBH, T. (1997). Representation of motion boundaries in retinotopic human visual cortical areas. *Nature*, *388*, 175–179.

[Rhodes and O'Leary (1985] Rhodes, G. & O'Leary, A. (1985). Imagery effects on early visual processing. *Perception and Psychophysics*, *37*, 382–388.

[Richardson (1969] Richardson, A. (1969). *Mental Imagery*. NY, NY: Springer publishing Cy.

[Richardson and Harris (1986] Richardson, A. & Harris, L. (1986). Age trends in eidetikers. *Journal of Genetic Psychology*, *147*, 303–308.

[Roland, Eriksson, Stone-Elander and Widen (1987] Roland, P., Eriksson, L., Stone-Elander, S., & Widen, L. (1987). Does mental activity change the oxidative metabolism of the brain ? *Journal of Neuroscience*, *7*, 2373–2389.

[Roland and Friberg (1985] Roland, P. & Friberg, L. (1985). Localization of cortical areas activated by thinking. *Journal of Neurophysiology*, *53*, 1219–1243.

[Roland and Gulyas (1994] Roland, P. & Gulyas, B. (1994). Visual Imagery and visual representation. *Trends in Neurosciences*, *17*, 281–287.

[Roland and Gulyas (1995] Roland, P. & Gulyas, B. (1995). Visual memory, visual imagery, and visual recognition of large field patterns by the human brain: functional anatomy by positron emission tomography. *Cerebral Cortex*, *5*, 79–93.

[Sakai and Miyashita (1994] Sakai, K. & Miyashita, Y. (1994). Visual Imagery: an interaction between memory retrieval and focal attention. *Trends in Neurosciences*, *17*, 287–289.

[Schorr, Balzano and Smith (1978] Schorr, D., Balzano, G., & Smith, E. (1978). Selective interference between imagery and perception: Is it modality specific or relation specific? *Bulletin of the Psychonomic Society*, *12*, 419–422.

[Segal (1971] Segal, S. (1971). Processing of the stimulus in Imagery and Perception. In S. J. Segal (Ed.), *Imagery - Current Cognitive Approaches*, chapter 5, pages 69–100. New York and London: Academic Press.

[Segal and Fusella (1969] Segal, S. & Fusella, V. (1969). Effects of imaging on signal-to-noise ratio, with varying signal conditions. *British Journal of Psychology*, *60*, 459–464.

[Segal and Fusella (1970] Segal, S. & Fusella, V. (1970). Influence of imaged pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology*, *83*, 458–464.

[Segal and Glicksman (1967] Segal, S. & Glicksman, M. (1967). Relaxation and the Perky effect: the influence of body position on judgments of imagery. *American Journal of Psychology*, *80*, 257–262.

[Segal and Gordon (1969] Segal, S. & Gordon, P. (1969). The Perky effect revisited: blocking of visual signals by imagery. *Perceptual and Motor Skills*, *28*, 791–797.

[Sereno, Dale, Reppas, Kwong, Belliveau, Brady, Rosen and Tootell (1995] Sereno, M., Dale, A., Reppas, J., Kwong, K., Belliveau, J., Brady, T., Rosen, B., & Tootell, R. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*, 889–893.

[Sereno, McDonald and Allman (1994] Sereno, M., McDonald, C., & Allman, J. (1994). Analysis of retinotopic maps in extrastriate cortex. *Cerebral Cortex*, *6*, 601–620.

[Sheehan (1966] Sheehan, P. (1966). Accuracy and vividness of mental images. *Perceptual and Motor Skills*, *23*, 1011–1033.

[Shepard and Metzler (1971] Shepard, R. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701:703.

[Singer and Sheehan (1965] Singer, G. & Sheehan, P. (1965). The effect of demand characteristics on the figural after-effect with real and imaged inducing figures. *American Journal of Pasychology*, *78*, 96–101.

[Somers, Nelson and Sur (1995] Somers, D., Nelson, S., & Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *Journal of Neuroscience*, *15*, 5448–5465.

[Spitzer, Desimone and Moran (1988] Spitzer, H., Desimone, R., & Moran, R. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, *240*, 338–340.

[Stroop (1935] Stroop, J. (1935). Studies of interference in visual serial reactions. *Journal of Experimental Psychology*, *18*, 643–662.

[Titchener (1907] Titchener, E. (1907). *An outline of Psychology*. New York: MacMillan.

[Tootell, Dale, Sereno and Malach (1996] Tootell, R., Dale, A., Sereno, M., & Malach, R. (1996). New images from human visual cortex. *Trends in Neurosciences*, *19*, 481–489.

[Tootell, Hadjikhani, Vanduffel, Liu, Mendola, Sereno and Dale (1998] Tootell, R., Hadjikhani, N., Vanduffel, W., Liu, A., Mendola, J., Sereno, M., & Dale, A. (1998). Functional analysis of primary visual cortex (V1) in humans. *Proceedinds of the NAS USA*, *95*, 811–817.

[Ullman (1984] Ullman, S. (1984). Visual routines. *Cognition*, *18*, 97–159.

[Ullman (1995] Ullman, S. (1995). Sequence-seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex*, *1*, 1–11.

[van Essen and Felleman (1996] van Essen, D. & Felleman, D. (1996). On hierarchies: response to Hilgetag et al. *Science*, *271*, 777.

[von der Heydt, Peterhans and Baumgartner (1984] von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, *224*, 1260–1262.

[Wallace (1984] Wallace, B. (1984). Apparent equivalence between perception and imagery in the production of various visual illusions. *Memory and Cognition*, *12*, 156–162.

[Waterhouse, Ausim Azizi, Burne and Woodward (1990] Waterhouse, B., Ausim Azizi, S., Burne, R., & Woodward, D. (1990). Modulation of rat cortical area 17 neuronal responses to moving visual stimuli during norepinephrine and serotonin micoiontophoresis. *Brain Research*, *514*, 276–292.

[Watson (1929] Watson, J. (1929). *Psychology from the standpoint of a behaviorist (third edition)*. Philadelphia: JB Lippincott Company.

[Wilson and Cowan (1972] Wilson, H. & Cowan, J. (1972). Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons. *Biophysical Journal*, *12*, 1–24.

[Wilson and Cowan (1973] Wilson, H. & Cowan, J. (1973). A Mathematical Theory of the Functional Dynamics of Cortical and Thalamic Nervous Tissue. *Kybernetic*, *13*, 55–80.

[Woodward, Moises, Waterhouse, Yeh and Cheun (1991] Woodward, D., Moises, H., Waterhouse, B., Yeh, H., & Cheun, J. (1991). Modulatory Actions of Norepinephrine on Neural Circuits. In S. Kito & al. (Eds.), *Neuroreceptor Mechanisms in Brain*, pages 193–208. NY,NY: Plenum Press.

[Yao and Freeman (1990] Yao, Y. & Freeman, W. (1990). Model of biological pattern recognition with spatially chaotic dynamics. *Neural Networks*, *3*, 153–170.

[Zeki (1993] Zeki, S. (1993). *A vision of the brain*. Oxford, England: Blackwell Scientific Publications.

[Zelhart, Markley and Bieker (1985] Zelhart, P., Markley, R., & Bieker, L. (1985). Eidetic imagery in elderly persons. *Perceptual and Motor Skills*, *60*, 445–446.