COMPUTATIONAL AND EMPIRICAL INVESTIGATION OF NUMBER, TIME, AND MEMORY

A THESIS SUBMITTED TO THE UNIVERSITY OF HYDERABAD
IN PARTIAL FULFILLMENT OF THE REQUIREMENT
FOR THE AWARD OF THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN
COGNITIVE SCIENCE
BY
RAKESH SENGUPTA

CENTER FOR NEURAL AND COGNITIVE SCIENCES
SCHOOL OF MEDICAL SCIENCES
UNIVERSITY OF HYDERABAD
HYDERABAD- 500046
DECEMBER, 2014

CENTER FOR NEURAL AND COGNITIVE SCIENCES
SCHOOL OF MEDICAL SCIENCES
UNIVERSITY OF HYDERABAD

DECLARATION

I, Rakesh Sengupta., declare that the work embodied in this thesis titled, *'Computational and empirical investigation of number, time, and memory'*, is the result of bona fide research carried out by me under the supervision of Dr. S Bapiraju, School of Computer and Information Sciences, University of Hyderabad and Dr. Prajit K. Basu., Department of Philosophy, School of Humanities, University of Hyderabad.

I declare to the best of my knowledge that no part of this thesis was earlier submitted for the award of research degree of any University or institution of learning.

Date: **Rakesh Sengupta**

Place: Hyderabad **Enrollment No.: 08CCPC02**

CENTER FOR NEURAL AND COGNITIVE SCIENCES
SCHOOL OF MEDICAL SCIENCES
UNIVERSITY OF HYDERABAD

CERTIFICATE

This is to certify that the dissertation titled, *'Computational and empirical investigation of number, time, and memory'*, submitted by Mr. Rakesh Sengupta, as part of the Degree of Doctor of Philosophy in Cognitive Science is an original work carried out by him under my supervision. The thesis or a part thereof has not been submitted for any other degree or diploma at this or any other university or institution of learning.

Date:
Place: Hyderabad

Dr. S. Bapiraju,
Supervisor
School of Computer and Information Sciences

Dr. Prajit K Basu
Supervisor
Department of Philosophy

Head
Center for Neural and Cognitive Sciences

Dean
School of Medical Sciences

I would like to dedicate this thesis to all those who are
part-of-no-part, the ones who are so familiar yet never counted for
to all our heroes who die in squalor.

Acknowledgements

The problem of us extroverts is to limit the number of people we have to acknowledge in a finite prose. To be honest, this section would have been longer and possibly more fun to read than the rest of my thesis. However, for the sake of finitude I would restrict myself to the people who have been with me during the span of my Ph. D. and have suffered me till the end of it.

To begin with, Prof. S Bapiraju and Prof. Prajit Basu, my two supervisors, have been my Plato and Aristotle. They have tried to inspire qualities in me, both professional and personal, that I can only wish to achieve in a distant future. They have rescued me from darkness more than once. They also have been the first to criticize my errors. I am truly grateful and honored that I am their student.

I would like to thank Prof. David Melcher for providing me the opportunity under Indo-Trento project for advanced research to advance my professional skills in cognitive neuroscience. My scientific career in cognitive science owes much to my time spent working in his lab.

I would like to thank Dr. Joby Joseph and Prof. Matthew Belmonte for being generous with their ideas and friendship as well as, going over and above their call of duty in order to write recommendations and reviewing my work in spite of their busy schedule.

Ekta is my emotional foil, and thus often subject to my absurdities. She made the conscious decision of believing in me when most others would have had given up. I would thank her for her love, patience, and understanding, but then again, I have a lifetime to do that anyway.

My holy trinity - Anwesha, Alex, and Prasanth. We have stayed together through heaven and hell, followed each other across continents, and seen the unwatchables (yes, I mean our sessions with ridiculously bad movies). They have celebrated with me at the best of times and were my angels to the rescue at the worst of times.

Some friends are born, some friendships are achieved, Reshanne had my friendship thrust upon her. Thank you for not making me climb a thousand steps to see some ridiculous vicious monkeys more than once - and also for making me smile on days when nothing else could.

Jen and Andi, my sister and brother from different mothers - you are awesome. I love you both. Thanks for taking care of me and making me feel like I belonged in a place where I would not have otherwise.

My own little fireball, my best travel partner, my permanent student, sometimes my biggest joy and worry at the same time, my family - Tinni (Rusha).

Thank you Vivek, Jobeth, and Shobhit for listening to my absurd plans and then unexpectedly, voluntarily even going along with it. With friends like them, I did not need a therapist. Anime, Tamil movies, Breaking Bad, late night ginger tea / black coffee / shawarma are all that a soul needs.

Anupam has been the rock of stability for a large part of my Ph. D. - we have been in accidents together, almost got thrown out of the hostel together, almost managed to raise a dog (Pakkun) together. In spite of not being on the same continent for the large part of last six years, Jillet has managed to be an integral part of my life, with almost no effort on my part. Without them I can not imagine a life in Hyderabad, though they live so far away now.

I would now like to acknowledge the person who probably wants to be acknowledged the least amongst my friends. From watching and recommending Tamil movies to proof-reading a paper before an important talk, Cibani has been indispensable.

Prathibha, Anuragh, and Mohib are better friends than I deserve. They have and never will let me turn into a grumpy old man. Without their spontaneous moments of fun (mostly at my expense) all my days would have been a bit worse.

Anuj and Suchitra have helped me grow as a scientist and also made going to my department worthwhile.

I would also like to thank Dr. Priyedershi Jetley for his continuous support and his faith in me, Sarah for her tremendous baking skills, Jaap for taking care of me on many occasions, Rajarshi for tolerating my company, Philipp for his smile that can move mountains and his scientific guidance, Evelyn for wonderful evenings at Locos, Varsha for not allowing my lack of general sense get in the way of our friendship, Shwetha for her help in tormenting Rajarshi and not falling asleep when I droned on and on, Madhurima for never taking me seriously, Dr. Pasha and Hari Krishna for their intellectual and personal support that made my last year of PhD wonderful.

I would like to thank Venkat and Shilpi from my department for their wonderful discussions. I would like to thank the former co-ordinator my department, Dr. Vipin Srivastava, and the current Head of the Department, Dr. Ramesh Kumar Mishra for providing me with a platform for doing research in cognitive science.

Lastly, I would like to thank my parents for being so understanding about why I did not have a real job so far in spite of being in higher education for many years.

Abstract

Cognition is embodied in the sense that it is situated in a complex of dynamic property-relations between the brain and the world mediated through the body and its affordances. The neural networks of the brain give rise to emergent perceptual / cognitive phenomena as a dynamical system self-organizing under the influence of the world according to its internal state. In the current thesis I have explored one type of recurrent on-center off-surround network that searches for suitable equilibrium states characterized by strengths of external stimulation and the inhibition between the neural nodes. I have shown how this kind of network is useful in understanding visual spatial phenomena like enumeration, serial order visual short-term memory as well as for phenomena related to time's subjective expansion in temporal oddball paradigm. The localization and the network parameters of the neural substrate varies given the task. Phenomenally each of the cognitive domains mentioned above have also been explored through behavioral experiments conducted in order to test the predictions from the model.

Contents

Contents vi

List of Figures viii

1 Introduction 1
- 1.1 The questions for mind 1
- 1.2 The theories in question: Cognitivism 4
- 1.3 The underlying problems of cognitivism 11
- 1.4 Embodied representation and Frame problem 12
- 1.5 Towards an embodied approach to meaning and action 15
- 1.6 The Dynamical Systems approach and Embodied stance 20
- 1.7 Organization of the thesis 22

2 Working Memory, Numbers and Time: An Overview 23
- 2.1 Introduction 23
- 2.2 Vision and attention 24
- 2.3 Attention, memory and individuation 30
- 2.4 Numerosity, saliency and spatial attention 32
- 2.5 Attention, time, and perception 34
- 2.6 Statement of the problem 35

3 Computational Modelling Framework using On-center off-surround network 36
- 3.1 Network architecture 36
- 3.2 Modelling of Visual Sense of Number 41

3.3 Modelling of Effect of Spatial Cueing on Working Memory 57
3.4 Modelling of time's subjective expansion 63

4 Empirical investigation of visual sense of numbers 77
4.1 Introduction 77
4.2 Model 80
4.3 Method 82
4.4 Results 84

5 Empirical Investigation of the Effect of spatial cueing on serial recall 91
5.1 Introduction 91
5.2 Materials and Method 92
5.3 Results 96
5.4 Discussion 98

6 Empirical investigation of time's subjective expansion 101
6.1 Introduction 101
6.2 Model 102
6.3 Method 102
6.4 Results and analysis 104
6.5 Discussion 105

7 Discussion and Conclusions 106
7.1 Numbers, memory, and time 106
7.2 Embodiment, representation, and saliency map 109
7.3 Future directions 114

Appendix A: Mathematical Details 115

Appendix B: How embodied is time? 126

List of publications 140
References 142

List of Figures

1.1 A brief organization of standard philosophical stands in Cognitive Science. 11

2.1 Nicholas Wade, *Aristotle's Senses.* A partial portrait derived from an engraving in *The Historic Gallery of Portraits and Paintings* (Vol. 5. 1810, Vennor, Hood and Sharpe, London). Specification of the senses of seeing, hearing, taste and smell was straightforward, but problems were caused by skin senses, which he variously called touch or feeling. 25

2.2 The dual path of visual procesiing. Visual information transmitted through the optic nerves enters the visual cortex in occipital region primarily through lateral geniculate nucleus (LGN, not shown) and superior colliculus (SC). There are two pathways from visual cortex to higher brain regions like pre-frontal cortex (PFC) - 1) dorsal stream through posterior parietal cortex (PPC), 2) ventral stream through inferotemporal cortex (IT). Dorsal stream processes the location aspect of vision, whereas the ventral stream is more concerned with object recognition and identification. Dorsal stream is important for attentional deployment although ventral stream also receives attentional feedback. Reproduced from Itti and Koch [2001] 27

2.3 Feature Integration Theory by Treisman and Gelade [1980]. It shows that feature maps for color, orientation etc. for visual objects are collected separately in several feature maps before being integrated on a master map of locations throgh attentional selection and focus. 29

2.4 Primacy and Recency effect illustrated for word recall. 32

3.1 Illustration of the on-center off-surround model, showing nodes with excitatory and inhibitory connections. 37
3.2 Network dynamics for a small number of inputs (4) at different levels of β. Top panels show the initial activations and bottom panels show final activation on a 10 x 7 grid representing 70 nodes ($N = 70$) 40
3.3 Network dynamics for a large number of inputs (16) at different levels of β. Top panels show the initial activations and bottom panels show final activation on a 10 x 7 grid representing 70 nodes ($N = 70$) 40
3.4 Mean activation vs. Set size and β. The plot was derived at $\alpha = 2.4$ as an average of 100 simulations. The insets show the mean activation vs setsize plot for three particular β values - low: 0.01 (bottom inset), medium: 0.1 (middle inset), and high: 0.15 (top inset). The verical line in the insets denote the limits to the monotonic region. For instance we can see from top inset that the mean activation ($\bar{x}(n)$) is monotonically increasing up to set size 4 for high β, but for low β (bottom inset) we can see that $\bar{x}(n)$ is monotonically increasing with numerosity for set sizes greater than 15. 42
3.5 $P(\beta, setsize)$ plotted against $setsize$ for different β. The Modelling parameters are given in Table 3.1. The figure shows how probability of faithful representation decreases with decreasing β. 44
3.6 Probability of finding a given numerosity larger than the reference numerosity (16) plotted against numerosity ratios. 45
3.7 Simulation of human numerosity comparison data by averaging the probability values at different β (0.01-0.04). 46
3.8 Probability of response ‘smaller’ for different numerosities for reference numerosity 16. The plot shows that how at higher inhibition numerosity underestimation is possible. 47

3.9 We show pairwise discrimination probability between different numerosities 1-5. The probability is proportional to the areas of the circles at each point. The largest area is 1 and the smallest ones adjacent to top right corner being 0.03. As we move towards the top right corner the circle sizes decrease, i.e., the overall precision of discrimination suffers with increasing numerosity. This exhibits *Fechner's law*. And the discrimination probability between 1-2, 2-3,3-4 and 4-5 pairs reveal the *distance effect*, where for the same numerical difference smaller numerosities are more accurate than the higher ones. 49

3.10 A parametric display of the reaction time distribution as calculated from the model. The values are meant to be proportional to log(RT) rather than directly representations. The top inset shows the contour plot for the energy surface and bottom inset shows the prediction for the reaction times for set sizes $1 - 32$ at $\beta = 0.03$, $\beta = 0.05$, $\beta = 0.08$ and $\beta = 0.1$. 51

3.11 This figure shows the average RT values for two ranges of β: high-0.08-0.11 and low-0.03-0.06. The higher numerosity ranges hardly differ in RT between the two ranges of inhibition, however the low numerosity range differs quite significantly in RT between the two beta ranges. 52

3.12 Application of the saliency map model to serially presented stimuli. Serial presentation dynamics of the model with 70 nodes. $\alpha = 2.0$, $\beta = 0.1$. Eight stimuli are presented serially for 5 time steps each. The input was clamped to 0.33 level. 58

3.13 Predicted probability of recall for items presented at a fast rate (each input presented to the network for 4 time steps) or slow rate (each input presented to the network for 6 time steps) 59

3.14 Data from Waugh and Norman [1965] showing probability of recall for items presented once or four times per second 60

3.15 Predictions of the model for proportion of recall for different set sizes as a function of serial position. A serial order effect is found when the number of items exceeds 4-5 items. 61

3.16 Recall curve for spatially cued item. The recall curve shows enhanced recall probability for 5 th item, the one that is spatially cued 62
3.17 Example activation pattern for change detection task. In a network of 70 nodes 6 nodes were stimulated for 5 time steps and then after a gap of 10 time steps 5 of the original nodes along with a randomly chosen new node was stimulated. 64
3.18 Probability of change detection for different set sizes and different β. Probablity that both the new (j) and old (i) node will remain activated simultaneously after simulation ($P(i\&j)$) 65
3.19 A plot of $\alpha^2 - \alpha - 2\beta$ for different β values. Blue line for $\beta = 0.25$, green for $\beta = 0.3$ and magenta for $\beta = 0.35$. In the bottom right quadrant of the plot bound by the lines $\alpha^2 - \alpha - 2\beta = 0$ and $\alpha = 1$ we get the α values desired for WTA interaction, mainly $\alpha \geq 1$ and $\alpha^2 - \alpha - 2\beta \leq 0$. For $\beta = 0.25$ the range is $1 \leq \alpha \leq 1.37$, for $\beta = 0.3$, $1 \leq \alpha \leq 1.42$, for $\beta = 0.35$, $1 \leq \alpha \leq 1.47$. 67
3.20 Probability of winner take all interaction plotted against α for three different β values 0.25 (blue line), 0.30 (green line) and 0.35 (magenta line). For $\alpha > 1$, the WTA probability increases beyond chance level (0.5). The analytical limits on α obtained in Fig. 3.19 are shown in correspondingly colored lines. 68
3.21 Winner-take-all behavior shown at $\alpha = 1.05$, $\beta = 0.3$ for two nodes receiving inputs for 480 ms (red line) and for 660 ms (blue line). . 70
3.22 The figure plots the probability of response longer than standard (here 5 time steps) compared to the variable time intervals shown on x-axis. 71
3.23 The probability values for different variable intervals being judged larger than standard interval for three different conditions 1) variable interval input level (0.2) less than standard input level (0.3), 2) variable interval input level (0.3) same as standard input level (0.3), 3) variable input level (0.3) more than standard input level (0.2). 72
3.24 TSE values for different standard durations. 73

3.25 Experimental data by Tse et al. [2004] 74
3.26 TSE values for different standard durations at different levels of oddball input. Blue line has oddball input at 0.3, red line has oddball input at 0.4, green line has oddball input at 0.5. The different levels were interpreted as arising from different levels of habituation, i.e., if the oddball appears in a stream of 6 standard stimuli, its salience would be less than that appearing in a stream of 11 stimuli, and thus the former will have lower stimulus level. . 75

4.1 This figure shows the average RT correlate or RT_{model} values for two ranges of β: high-0.06-0.08 and low-0.04-0.06. The higher numerosity ranges show similar patterns of RT for the two ranges of inhibition, while the low numerosity range differs greatly in predicted RT between the two beta ranges. The error bands represent one standard deviation of the mean RT_{model} values at different set sizes. 81
4.2 Illustration of the structure of a trial. A display of black dots was presented for 100 ms and participants verbally reported the number of items as quickly and accurately as possible. The white flash was used as a mask to reduce the visual persistence of the dots in order to encourage participants to respond quickly rather than slowly counting each item individually from the after-image. 83
4.3 Enumeration error between the reported numerosity by the participant (estimate) and the actual presented numerosity for different set sizes. Negative numbers indicate that participants underestimated the numerosity. Error bars indicate standard error of the mean. 84
4.4 The time course of underestimation effects of the switch, starting on first trial after the switch from small to large (red squares) or large to small (blue diamonds). Amount of underestimation is shown also for the 2nd, 3rd and 4th consecutive numerosity trial of the same kind that continued after the switch. 85

4.5 Average reaction times for each numerosity plotted with error bars for standard/non-switched and switched trials. Non-switch trials refer to when a small numerosity trial followed a small numerosity trial or large numerosity trial followed a large numerosity trial (see Methods). 86
4.6 Enumeration switch cost (RT) for reaction time for large and small numerosity trials. 87
4.7 Reaction time as a function of number of subsequent trials following the switch, for small and large numerosity trials. 88
4.8 The fit between model predicted RT values and empirically obtained RT values for non-switch trials ($RT_{non-switch}(empirical) = a * RT_{non-switch}(model) + b$). Adjusted r^2 is 0.93 at 95% confidence interval. 89
4.9 A fit between model predicted RT values and empirically obtained RT values for switch trials using the same coefficients a and b obtained from the fit for non-switch trials. 90

5.1 The schematics of Experiment 1 93
5.2 The schematics of Experiment 2 94
5.3 The schematics of Experiment 3 95
5.4 Reaction time results for Experiment 1. The results show clearly that in terms of reaction times, for shapes the recall of spatially cued item is facilitated, but not for colors. 96
5.5 Accuracy results for Experiment 1. The results do not seem to show a speed accuracy trade-off for shape recall. Showing that they are better and faster at shape recall if the tested item's location was made spatially salient. 96
5.6 Reaction time results for Experiment 2. In this experiment we see that spatial cues do not have any noticeable effect on either shape or color recall for any set sizes. 97
5.7 Accuracy results for Experiment 2. The accuracy results seem to be consistent with the trend in RT's 97

5.8 Reaction time results for Experiment 1. The results show clearly that in terms of reaction times, the spatial cue probe are facilitatory for serial recall of shapes, but not for color recall. 98
5.9 Accuracy results for Experiment 3 98
5.10 Interaction between color and shape recall. In terms of reaction times, the results show a clear interaction between color and shape recall performances. 99

6.1 The design of the temporal oddball experiment 103
6.2 The average TSE values for all the subjects for all the conditions with standard error given in parenthesis. 104
6.3 TSE values calculated for each habituation condition for both the standard durations. The results confirm the model predictions . . 105

7.1 Embodied stance of Lakoff and Johnson [1980]. 111
7.2 The relation between different strands of embodied stance 112

Chapter 1

Introduction

To see a world in a grain of sand,
And a heaven in a wild flower,
Hold infinity in the palm of your
hand,
And eternity in an hour.

Songs of Innocence
William Blake (1803)

1.1 The questions for mind

> The one thing that you have that nobody else has is you. Your voice, your mind, your story, your vision. So write and draw and build and play and dance and live as only you can.
>
> -Neil Gaiman

William Blake, in his madness composed the poems which were later collected in the volumes called *Songs of Innocence* and *Songs of Experience.* Almost all the poems in *Songs of Innocence* have their opposite in *Songs of Experience* - 'The lamb' versus 'The tyger', 'The chimney sweeper' versus 'London'. His strangely dichotomous ways are seen also in the medium he chose to express in. His copper plate engravings were obsolete even in the period he wrote in.

However, the engravings were as much visual as haptic, his writings as much real as much fantasy. He was a man forever opposing the mechanistic view of Newton, believing in the infinite potential of humanity and the idea of holism beyond variations in appearances. More than one and a half centuries before the publication of *Songs of Innocence* Benedict de Spinoza was excommunicated from his community for his questions regarding the holism of Jewish theology. For his singular sin he was denied communication, communion and sharing the commons of existence (of culture, of language, of material) with his fellow human beings. William Blake severed ties with the commons through mythology and complex symbolism of his own invention in his fantasy while being characterized as the mad poet.

In their excommunication, they bring out in a dialectical fashion, the commons of human existence - the commons of our bio-genetic inheritance, the commons of our shared language and culture, the commons of our shared environment [Sengupta, 2013, 2014]. Vico [1984] wrote about how human cognition lies in our affective interests in the world, the affective interests depend upon our commons and shape our imagination (and fantasy) that hold our self together, akin to the Freudian ego.

Even before the 'mind' the classical exploration of self and psychology can be said to have started with Aristotle. Aristotle envisioned a rational teleological self, striving towards the just in society and polity[1]. The Aristotelian idea of self and the questions concerning them propagated till the middle ages through philosophers like Aquinas, Anselm, Boethius and William of Occam. This view of self was not challenged till enlightenment and the mechanical world view. It is interesting to note that in *Nicomachean Ethics* [Aristotle, 2007], explored the question of exchange and commons of currency and norms to establish how just prices are arrived at. In *Capital, Volume I*, Marx [2007] uses the same explorations to flesh out how material exploitation of human beings striving for their existence comes to be in a system that makes their existence possible.

What Marx opposed was the dualism that came with modernity and enlightenment - the separation of mental and physical labor that strives to establish a

---

[1] Although much of Aristotle's psychological ideas concern the soul, we take soul to be analogous to self.

disembodied consciousness. However for Marx [1970]

> The first premise of all human history is, of course, the existence of living human individuals. Thus the first fact to be established is the physical organization of these individuals and their consequent relation to the rest of nature. .... By producing their means of subsistence men are indirectly producing their actual material life. .... As individuals express their life, so they are. What they are, therefore, coincides with their production, both with what they produce and with how they produce.

For an example, let us consider the idea of a signature. The digitalized world has achieved what Gutenberg's printing press began - to copy without human agency. What it has done to the question of authenticity or the original is a matter the postmodern scholars have taken up. However, one cannot but notice that we had to keep the signature, even when the word digital signature is becoming increasingly more common place. Signatures can be forged - but at least it was forged by a human being with a will and agency. Our concepts of ethics and morality should apply to him/her. We can punish the one responsible for the forgery. Our target is to hold someone responsible - to make crime disembodied (and dis-embedded as well[2]). Alan Moore's *From Hell* scares us because the crimes are somehow embodied, resonating back and forth between centuries, resonating in tune with the whole society [Moore and Cambell, 2006; Sengupta, 2014]. Kafka pushes this fundamental idea of disembodied responsible self in his *Metamorphosis*, where the protagonist, even in a changed physical form is still responsible for his world - even trapped in it till he dies (by going back to his own room) [Kafka, 2008; Warjri and Sengupta, 2011]. However, as Hegel [1988] points out, we have to overcome the negation of our embodied mind.

> The life of the Spirit is not the life that shrinks from death and keeps itself untouched by devastation, but rather the life that endures it and

---

[2] A disembodied crime is also dis-embedded. The causal chain of crime is then unlinked from the totality of the concept of crime. It was very nicely illustrated in a web-comic where Superman tries to catch a burglar - who blames his employer for laying him off - the employer blames the Wall St. and so on, until Superman gets frustrated enough and punches the burglar.

> maintains itself in it. It wins truth only when, in utter disembodiment, it finds itself ... Spirit is this power only by looking negative in the face, and tarrying with it.

Indeed a Cartesian separation of mind and body is also a transition point when mind metonymically determines the self. Thus questions of self also become questions for a mind. How mind can become a mediator between objective reality and subjective notions[3]? How do we perceive the external world? What are the categories of perception? How is perception translated into action? How do we ascribe meaning to objects and events?

Psychology has evolved for over the course of a century through methods of mentalist introspection (works of Wundt [Wundt, 1912], James [James, 1950], and later Freud [Freud, 1895], and Gestalt school [Koffka, 1935]), psychophysics [Weber, 1978], behaviorism [Pavlov, 1927; Skinner, 1938; Watson, 1913], and finally to modern day cognitive psychology. Modern theories put two pathways for information processing at the core of cognitive processes (*bottom-up* sensory driven processing, and *top-down* knowledge-driven processing). This has led to the computer metaphor (Broadbent's model) for cognition [Styles, 2005]. Modern cognitive science and its questions concerning human mind also follow within this rich tradition. In the following I would try to show the methodological differences between different strands of cognitive science and philosophy of mind and their historical continuum.

1.2 The theories in question: Cognitivism

> It's this expandable capacity to represent reasons that we have that gives us a soul. But what's it made of? It's made of neurons. It's made of lots of tiny robots. And we can actually explain the structure and operation of that kind of soul, whereas an eternal, immortal, immaterial soul is just a metaphysical rug under which you sweep your embarrassment for not having any explanation.
>
> -Daniel C. Dennett

---

[3] Here we must remember the distinction between Subjective and Objective, as we know it, is a Post-Kantian understanding [Daston and Galison, 2010].

Let us consider the example of the phenomena of color vision. From the anatomical research we are aware of photosensitive and color sensitive cells that transmit differential signals to the visual areas of the brain with trichromacy [Wässle, 1999; Young, 1802] determined at receptor level and 'opponent' processes [Hurvich and Jameson, 1957] operating at the level of retinal ganglion cells and beyond.

We are also aware of the cultural aspect of color vision where knowing the name of a color is important for recognizing it [Brown and Lenneberg, 1954]. Even if we disregard stronger versions of Sapir-Whorf hypothesis [Sapir, 1983; Whorf, 1956], we know that categorical perception occurs in color perception on going from blue-green for example. This is not a gradual change like in shades of gray. Perceptual mechanisms coupled with coupled sensori-motor processes make some frequency ranges automatically "compressed". Thus the varying shades look like same color. Just as there is compression within each color range, there is expansion between them [Harnad, 2005].

The materialistic views in philosophy of mind continue predominantly in the form of **physicalism**. It encompasses a wide range of theoretical standpoints - from **identity theories** to **supervenience physicalism**. In the **type/token identity theories** it is assumed that brain states are identical to mental states. In **token identity**, for instance, the appearance of a particular pain state (mental state) would be identical to the corresponding particular brain state, however, **type-identity theories** assume that a kind of ensemble of token pain brain states can be classified separately from an ensemble of token brain states for other mental phenomena [Searle, 1992]. Within these kinds of reductive stands of physicalism, color vision should be identical to firing of corresponding color neurons in the brain. Non-reductive stand, within physicalism like supervenience is opposed to the identity theories. **Supervenience physicalism** considers mental states (e.g of color) to supervene on brain states and the properties of the brain states roughly determine mental states arising from them [Davidson, 1970]. The idea of *supervenience* however has a few noticeable problems. To talk about supervenience is to accept a de-facto abstraction/classification that seemingly operates over multiply realized physical brain states. However, the abstraction/-classification is a post-hoc one. Scientifically, one cannot predict the answer to the question, "Which mental state will supervene over a given brain state A?". Supervenience is useful only when we already have a de-facto classified mental state, which we can call upon as it were, to be always correlatively exemplified along with the said brain state. In case of identity theories the problem lies in one core assumption that the identity of physical and mental state leave no residue. However, Nagel [1974], Kripke [1980] and Searle [1992], have shown Qualia always remains as the residue, as even type and token identity theories cannot explain the subjective variances let alone the cultural effects elucidated by Sapir and Whorf.

Dualism on the other hand, is contrary to physicalist ideas. Dualism is as old as Descartes. Descartes was a **substance dualist** proposing mental having a different substance and thus subject to different rules from physical. Except for Eccles [1993] however, there are no neuroscientists who believe in substance dualism. The common form of dualism is encountered in contemporary times in the form of **property dualism/Dual-Attributes Theory**. Property dualists assume that mental phenomena like color vision are governed by rules independent of their neuro-physiological correlates, and can also be called propositional attitudes towards the world. Although the natural successor of property dualism is **functionalism**, there are some salient features that property dualism shares with **emergentism** as well. Emergentism considers mental phenomena to be an emergent property of the underlying neuro-physiological state. Thus although, the mental states can be said to be caused by physical brain states the former cannot be reduced to the latter. However, one crucial difference separates the property dualists from the emergentists framework. For property dualists, the mental phenomenon is not reliant on particular substrates, in case of emergentism it is necessary that the mental emerges out of the physical correlates. However, emergentism suffers from two major problems.

Multiple Realizability With its stress on non reductive nature of the mental, it fails to explain how the same emergent property can come to be multiply realized. *Multiple realizability* refers to the age old philosophical problem, - How can the same idea of pain be applicable to different human beings/ organisms?

Where is *emergence*? The sustainability of the idea of emergence depends on understanding mind/brain as a complex system, where one can not predict the system's behavior with arbitrary precision given the sensitivity to initial states, just as one can not predict the weather more than two days in the future. However, we can explain the behavioral pattern post-hoc from the initial parameters. This is the concept of causal emergence. In being so it is not emergent at all. According to Searle [1992]

> This conception of causal emergence, call it "emergent1" has to be distinguished from a much more adventurous conception, call it "emergent2". A feature F is emergent2 iff F is emergent, and F has causal powers that cannot be explained by the causal interactions of a, b, c. . . If consciousness were emergent2, then consciousness could cause things that could not be explained by the causal behavior of the neurons. The naive idea here is that consciousness gets squirted out by the behavior of the neurons in the brain, but once it has been squirted out, it then has a life of its own.
>
> It should be obvious ... that on my view consciousness is emergent1, but not emergent2. In fact, I cannot think of anything that is emergent2, and it seems unlikely that we will be able to find any features that are emergent2, because the existence of any such features would seem to violate even the weakest principle of the transitivity of causation.

This brings to question the very idea of emergence itself.

Functionalism seems at first sight capable of solving the multiple realizability problem. Functionalists focus on causal relationships between different mental states that can be realized on any substrate. Functionalists give preference to functions over structure, thus even if the same substrate can serve different functions and the same function can be served by multiple substrates, it is the causal relationship between these different functions that is important. Thus all kinds of mental phenomena can be said to be operations of certain rules on *mental representations* obtained from sensory motor modules. For instance in functionalism one can talk about rules that combine color with other features of objects without worrying about the underlying physical substrates in terms of a color module that interacts other modules of vision through some kind of central executive module. This kind of a black box approach obviously depends upon a notion of Fodorian *modularity* of mind [Fodor, 1983]. However, some of the current research does support modularity in the brain to a certain extent[4].

Functionalism falls within a broader tenets of the *information processing paradigm* and mind is envisioned as a symbolic system quite analogous to Universal Türing Machines relying on mental algorithms that operate on mental representations. This idea has drawn much fire from varied sources. One of the arguments advanced by Penrose [1994] states that human beings can not be described by formal axiomatic systems on the account that all such systems are Gödel susceptible (i.e., unable to check their own consistency [Gödel, 2002]) whereas human beings are not. However, as Putnam [1991] pointed out that if the algorithm for human mind has more lines than say US Congress Library, then it will not be possible for a human being to check for his/her own consistency making them functionally equivalent to Gödel susceptible systems, although they are not. However, Putnam [1991] has also stated that human beings are not Gödel susceptible and human mind can not be captured by functionalist account because human beings can use non-demonstrative reasoning. The debate is still open [Buechner, 2008].

The symbol systems approach was also criticized in the *Chinese room argument* and its variants by Searle [1980] and *twin-earth argument* by Putnam [1991]. The philosophical problem with functionalism is probably that it is too easily realizable on too many substrates as functionalism borrows very heavily from the *computer metaphor*. The traditional functionalists hold that cognition is in some sense *platform-independent*. A part of this argument is dubbed by Shapiro [2004] as the **separability thesis** (ST), which says that human like mind could very well exist in a non-human body. However, according to Shapiro [2004],

> ... psychological processes are incomplete without the body's contri-

---

[4] There has not been much evidence for *massive modularity* [Buller, 2005]

> butions. Vision for human beings is a process that includes features of the human body ... this means that a description of various perceptual capacities cannot maintain body-neutrality, and it also means that an organism with a non-human body will have non-human visual and auditory psychologies.

The other arguments against come from Lakoff and Johnson [1980] from the consideration of the role of the body in structuring human concepts.

One of the most vocal opponents of functionalist paradigm come from the framework of **Eliminative Materialism**. They differ from also reductive and non-reductive stands of physicalism in that they consider there is nothing to reduce in *qualia* at all. In other words, for Eliminative Materialists the subjective content for which functionalists formulate their rules for are nothing but *folk psychological terms* to be eliminated in completed neuroscience [Churchland, 1981, 1989]. It might be true that there is nothing special about qualia as Dennett [1988] has shown. However, the subjective content of consciousness like colour vision can not be denied (for instance, Abramov et al. [2012] has found significant difference in colour vision between and male and female humans). The subjective experience of specific colour ought to be explained through neurobiology (with or without retention of the concept called qualia). However, throwing away subjective experience baby with the qualia bathwater might not be the best course of action.

However, in terms of explanatory power and empirical evidence, only **connectionism** can be said to be a true contender against functionalism. Early neural network concepts found their way in literature in the late nineteenth century as soon as the existence of nerve cells were accepted [Freud, 1895; Spencer, 1872]. But the resurgence of interest in characterizing network of nerve cells through mathematics occurred through works of McCulloch and Pitts [1943], Hebb [1949] and Rosenblatt [1962]. Rosenblatt demonstrated mathematically that the perceptron convergence rule could adjust weights connecting an input layer and an output layer of simple neurons to allow the network to associate arbitrary binary patterns. But it was proven by Minsky and Papert [1988] that these two layer networks are very limited. But Rumelhart et al. [1986] showed postulation of "hidden" internal processing units along with networks trained using backpropagation can address key challenges to neural networks. Some key themes have emerged through the subsequent research in role competition in processing and learning, properties of distributed representations, and possibility of content addressable memory in networks with attractor states. The connectionists would try to account for color vision with statistical regularities in the environment that can be processed as distributed representation in a biologically plausible neural network.

However, there are some problems with the connectionist approach. Connectionist models either include properties that are not neurally plausible (like backward propagation) or omit other properties that neural systems appear to have. Another aspect of the symbolic vs. connectionist debate is that though some take the view that systamaticity of symbolic approaches are overstated and human behavior is better matched with connectionist models, but connectionist models are not powerful enough to account for some aspects of cognition best described by the symbolic [Thomas and McClelland, 2008].

On the other hand, the **behaviorists** would concentrate purely on the stimulus and response association, while undermining anything deeper. **Nativists** would argue on the basis of innate color perceptual abilities that vary with the genetic code.

A chart elaborating different strands in philosophy of mind is given in Fig. 1.1 [5].

---

[5] A brief explanation of the terms not given in the main text: **Idealism** maintains that everything is ultimately mental and our supposition of an external physical world is merely the result of an elaborate way of describing subjective mental experiences. **Neutral Monism** maintains that everything is ultimately made up of a neutral substance that is neither physical nor mental. But this substance can be described in physical or mental terms. **Realization Physicalism** maintains mental phenomena are realized by physical phenomena. Indeed, they can be realized by multiple kinds of physical phenomena. In **Anomalous Monism** all events are describable in physical terms; but some events are also describable in mental terms. The psychological explanations that use these mental terms are, however, not law-like (a - nomos, without law). **Substance Dualism** maintains that there are ultimately two kinds of stuff: mental stuff and physical stuff. Our minds, obviously, are made up of the former, not the latter. **Organismic DAT** maintains that the kinds of entities displaying mental attributes are physical organisms. **Epiphenomenalism** maintains that mental properties emerge from or are caused by physical phenomena, but that these mental properties play no causal role in physical reality. Other terminologies are described in the main body of the text.

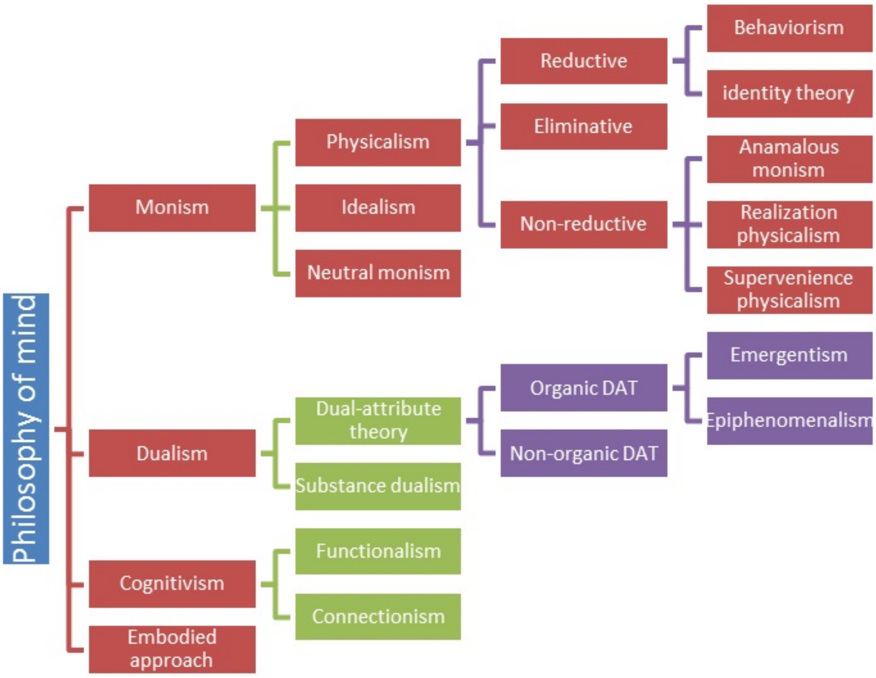


Figure 1.1: A brief organization of standard philosophical stands in Cognitive Science.

1.3 The underlying problems of cognitivism

> As the ego cogito, subjectivity is the consciousness that represents something, relates this representation back to itself, and so gathers with itself.
>
> -Martin Heidegger

The schools of thought mentioned before can be largely classified as cognitivism that started with Chomskian revolution. However,recent debates amongst Pinker and Ullman [2002] versus Ramscar [2002] (the past tense debate), Fodor and Pylyshyn [1988] versus Elman [1998] (the systematicity debate), Marcus et al. [1999] versus Seidenberg and Elman [1999] elucidate that all of these are somehow within paradigm debates. The fundamental similarity is that all of these approaches presuppose a cognitive sandwich architecture where the cognition fills in between perception and action [Calvo, 2005].

As we can see, the cognitivism paradigm falls short of a defensible account of qualia that is flexible enough to account for the findings of Sapir-Whorf to any satisfactory level. These failures are due to three fundamental problems.

1. How can the same object acquire different colors and remain the same object or more precisely, how do we understand when one event changes one property and not the rest (e.g. how can we change the size of an object while keeping the color constant?) [Dennett, 1996]. This is the formulation of the **Frame Problem**. The generality of the concept of a color sans the particularity of the object it is attached to [Putnam, 1991] requires a solution to the Frame Problem.

2. How does a range of continuous spectrum become associated with a given discrete category word of a color? How do words/symbols get their meaning? This is a formulation of the classical **Grounding Problem** [Harnad, 1990].

3. How is our perception of color oriented towards judgments and actions? To put it simply, how are we able to see and touch/rotate/pick up the same object? This is an expression of the **common code problem** [Hommel et al., 2001], where we want to understand what links perception to action.

As Calvo [2005] has shown that the filling in need not be just a Fodorian or a feed forward connectionist or the intentionality in the analytic tradition. The other approach that seems to promise a way out of the blind alley of cognitivism is the Dynamical Systems and Embodied Approach to Cognition. In the following sections I will try make the case for the embodied approach as a solution to the above problems.

1.4 Embodied representation and Frame problem

> The world is... the natural setting of, and field for, all my thoughts and all my explicit perceptions. Truth does not inhabit only the inner man, or more accurately, there is no inner man, man is in the world, and only in the world does he know himself.

-Maurice Merlaeu-Ponty

A logicist system does not infer anything unless it has an explicit rule to do so. Thus the commonsense law of inertia that dictates that if the light-bulb in a room is changed, the color of the walls will not change, is very hard to implement through the knowledge base of an expert system (AI) for world-scale systems. While the logical frame problem has been solved, the actual implementation remains challenging because the system engineer has to produce a correct set of axioms which lists the properties that do change for a given action, so that the system can be scaled up to handle real world situations. According to Pylyshyn [1987]

> Another way to put this problem is to say that Artificial Intelligence must face the problem of determining the relevance of facts it knows to some problem at hand. This, the problem of relevance, is what many believe lies at the heart of the frame problem, and which will continue to be a serious problem long after all the minor technical problems (e.g., concerning the need for "frame axioms") have been dealt with.

The axiom approach towards the solution for this problem may fail due to two factors

1. Writing a closed set of axioms for all given and novel situations is not possible [Fodor, 2000].
2. Even if it were possible to solve the previous problem, as Dreyfus [2008] pointed out, as the axioms would not be of help to decide which axioms would be relevant in the current *situation*.

Dreyfus [2008] points out that the frame problem arises due to a Cartesian ontology where one does not just see something, but sees it *as* something - there is always an extra step of deciding how to apply meaning to a given situation. On the other hand in the Heideggerian approach meaning is always ready-at-hand,

> To say a hammer has the function of being for hammering leaves out the defining relation of hammers to nails and other equipment, to the point of building things, and to our skills all of which Heidegger called readiness-to- hand and so attributing functions to brute facts couldnt capture the meaningful organization of the everyday world. - Dreyfus [2008]

As one simply responds to the situation at hand, there is never a problem of relevance to solve [Salay, 2009]. For Dreyfus, embodiment exists in "being-in-the-world" relation between cognitive agent and the world, what Dreyfus [2008] describes as 'background coping'. This relation however is a non-representational one. Dreyfus uses the dynamical systems approach of Freeman [2000] to point out how non-representational approaches can avoid the pitfalls of frame problem suffered by the disembodied approaches relying on representations.

> Thus Freeman contends that each new attractor does not represent, say, a carrot, or the smell of carrot, or even what to do with a carrot. Rather, the brains current state is the result of the sum of the animals past experiences with carrots. What in the physical input is directly picked up and resonated to when the rabbit sniffs, then, is the affords-eating, and the brain state is directly coupled with (or in Gibsons terms resonates to) the affordance offered by the current carrot. - Dreyfus [2008]

Although some think that Dreyfus concluded too much [Salay, 2009] in eschewing all representational approaches, however his emphasis on a more embodied approach is well justified. A common approach in the embodied stance follows an action-based model, where representation is derived from interaction, where Piaget's model of representation of manipulable objects in terms of organizations of potential actions [Piaget, 1954] can be translated directly into the interactive model. The truth value of a representation emerges naturally based on whether the interaction proceeds as indicated or not [Bickhard, 2008].

1.5 Towards an embodied approach to meaning and action

> Existence is random. Has no pattern save what we imagine after staring at it for too long. No meaning save what we choose to impose. This rudderless world is not shaped by vague metaphysical forces.
>
> -Alan Moore, *Watchmen*

Ring around the rosy,
pocket full of posy,
ashes, ashes all fall down.

This childrens rhyme is a metaphor for the bubonic plague that ravaged Europe in the dark ages. We have recently started using the term “differently able” for the physically handicapped persons. In the first case a dreadful history is re-contextualized in the form of a rhyme. In the second case an alternative conceptual schema (and thus also a different metaphor) is adopted in the place of an existing one. Our languages abound in figuration i.e., a major part of our vocabulary consists of figurative language and metaphors. These above examples provide certain clues as to how closely linked metaphors are to the conceptual structures in our minds and thus to cognition.

As we saw in the earlier section, the problem of meaning is also connected very closely to common code theory that suggests that our perception and action are more closely related than the theories of cognitivism allow for.

Recent trends in cognitive science suggest that it might be possible to work towards an account of meaning and action that does not overlook the biology. In this section I would like to make the case for a biology of meaning and action harnessing the views of Lakoff and Johnson, Vico, Merleau-Ponty and Aquinas in the following sections.

1.5.1 Meaning and metaphor

The question of meaning is a central one in terms of cognition. Although, meaning has been debated for a long time by philosophers of language, analytic philosophers, and linguists, following the cognitive revolution diverse fields like computer science, AI, physics, biology, etc. have tried to approach it from their respective vantage points. There is no proper consensus amongst such various different disciplines concerning the nature, construction, structure of meaning or its relation to other cognitive functions like memory. Lakoff and Johnson [1980] were pioneers to point out that meaning remains in unconscious mind as a potential property. Metaphors use this potentialities of meaning in unconscious memory to re-contextualize or re-categorize our experiences to produce novel connections that map across dissimilar domains. They further asserted that metaphors were not tropes of language, or aberration (Chomskian point of view), but a form of cognition.

Some famous examples used by Lakoff and Johnson [1980] to illustrate our cognitive structure through metaphors were, time is money, argument is war, etc. They also showed how our mental well-being is situated in bodily metaphors of high, low, up, down. Moreover these are the same terms that we use to define our position in society, growth etc. Lakoff and Johnson [1980] tried to show through these examples how our embodied mind works to understand the world. Even abstract metaphors are created via everyday human cognitive mechanisms that extend the structure of bodily experience while preserving the inferential organization [Lakoff and Nuñez, 2000].

1.5.2 Metaphors and scientific imagination

Kekulé claimed that the structure of benzene occurred to him in a dream.

> “One of the snakes had seized hold of its own tail and the form whirled mockingly before my eyes. As if by a flash of lightning I awoke... Let us learn to dream, Gentlemen.” [Koestler, 1964]

Although Kekulé’s work on benzene spanned ten years before it took its final shape, we can see the dream as an unconscious metaphoric process. An equally interesting account was of Poincaré inventing Fuschian functions. Einstein in one of his letters refers to an unconscious process where the play of fragmentary visual, auditory and kinesthetic sensations occur [Hadamard, 1945] - which can be inferred as an unconscious metaphoric process which is embodied.

Here embodied should not be understood as trivial as saying that human minds have a body as that can lead us to postulate different existence and thus different rules for mind and body (as in Descartess writings). Here it is important to posit the distinction between Descartes and Vico. Descartes separated mind from the natural world and tried to mathematize mind and the representations it works on (an implicit content of information theory). Whereas Vico [1984] talked of meaning being embodied in our total affective interest in the world. He said initially humans were without language and communicated through gestures and signs. Metaphors then were the primary ways of knowing and understanding the world. "Every metaphor is a fable in brief" [Modell, 2003].

1.5.3 Meaning to action

Both Chomsky and Fodor believed that meaning could be equated to formal symbolic logical mental code [Chomsky, 1975; Katz and Fodor, 1963]. But the belief in the objectivity of meaning (in a logical sense) requires one to accept the concepts such as correspondence [Austin, 1970] and representation to account for a constant and truthful relation between what is represented in the mind and what exists in the physical world [Fodor, 1987]. Although correspondence theory of truth has been rejected by contemporary philosophers, concept of representation still survives.

This kind of symbol processing view of mind reduces it to functional causal processes that do not need to depend on the actual brain, or the body for that matter. This is a case of disembodied cognition. The more biologically faithful approach of connectionism denies the existence of mind all together and conflates it to brain and central nervous system. Even this is disembodied to the extent that sensory-motor functions of the body are seen as peripheral to cognition [Clark, 2008]. A question for such an approach would be to ask whether a disembodied mind can indeed form and understand metaphors or meaning in general, as the production of meaning is not necessarily the same as processing of information. If metaphors are derived from the body and its relation to the world, then how can this relation be captured within a framework of disembodied mind? To carve out a space for mind (like Descartes), even functionally, out of the functions of the body might not even be possible.

In contrast to Descartes, Vico recognized a difference between self-knowledge and knowledge of social institutions in which we take part, on the one hand, and the knowledge of the outside world that exists irrespective of us, on the other [Vico, 1984]. This opens up the possibility of having a pluralistic epistemology that straddles the divide between first-, second-, and third- person perspectives.

Vico asserted that meaning is embodied in our total affective interest in the world. He talked of human minds having evolved linguistically through cultural revolution. Before language, metaphors were primary modes of knowing and understanding the world, as touched upon at the end of the last section. Thus metaphors were not figures of speech, but a vital means of understanding the world. Vico [1984] also stated,

> "...in all language the greater part of the expression relating to inanimate things are formed by metaphors from the human body and its parts and from the human senses and passions."

Thus metaphor also derived from the body and its affective experiences.

Meaning (metaphorical or otherwise) stems from the mind trying to interact with the environment. Meaning is sometimes equated with intentionality. Because relation of mind to body and thus extending to the world is said to be an intentional one.

> "The external forces, what we usually think of as 'environment', are themselves partly a consequence of the activities of the organism itself as it produces and consumes the conditions of its own existence. Organisms do not find the world in which they develop. They make it." -Lewontin [1991]

Such relations can not remain at the descriptive level of "about-ness".

The other strand of philosophical tradition follows from Thomas Aquinas who in 13th century delved into an intricate theory of Being as derived from Creation. Mind and body here were conceived in a much larger context of the predominant theory of the universe. The key concepts here were unity (by the mind's assertion of a boundary against the outside) and an intentional relation (between intellect and material objects) which shapes itself by changing its scales to accommodate the object through the process of knowing it. Intentionality as defined here is not "about-ness" [Searle, 1992]. Meaning is achieved through action in the world, and the self is altered by that action. Although Aquinas [2000] describes intentionality as directing action towards some future defined and chosen by the actor, intentionality is fundamentally an unconscious process, i.e., the self is not necessarily a conscious agent.

Merleau-Ponty's formulation of "the intentional arc" that completes an organic loop from action through the world and back into the brain, resembles Aquinas active intellect. The intentional arc names the tight connection between the agent and the world, perception and action. The skills acquired by the agent are not stored as representations but as dispositions to respond to directing situations in the world [Merleau-Ponty, 2012].

Neurobiologists like Freeman [1995] have endorsed the Aquinas' version of intentionality. In his terms intentionality means the "process of a brain in action having the properties of unity, wholeness, and intent." Unity refers to a state of integration by which a self distinguishes itself from non-self. Wholeness refers to a bounded process by which through stages a self actualizes its mature form, ultimately to die. Intent refers to a relation in which a self modifies itself in conformance with aspects of non-self. A living brain thus capable of actualizing these properties by purposive behavior is an intentional structure.

Gibson's work on visual perception brought us the concept of affordances [Gibson, 1979]. They may be thought of as invariant sensori-motor features. But more generally they are to be understood as all "action possibilities" latent in the environment, objectively measurable and independent of the individual's ability to recognize them, but always in relation to the actor and therefore dependent on their capabilities. An intentional arc is a complement to the affordances as we can see from the previous discussion. Rather one might be tempted to say that changes in the mind/brain in the act of knowing, is directly related to the affordances.

Keeping the above discussion in view, I have taken a Dynamical Systems and Embodied approach towards human cognition for the topic at hand. I will explore the dynamical approach and embodied stance in next sections.

1.6 The Dynamical Systems approach and Embodied stance

> The body is not a thing, it is a situation: it is our grasp on the world and our sketch of our project.
>
> -Simone de Beauvoir

Cognition operates in everyday life seamlessly integrating sensory inputs and motor functions, the flow of activities moving forward to create a complex maze of activities. In this section we will briefly outline the Dynamical systems approach as a way to formalize the embodied stance.

Cognition is embodied in an obvious sense as cognitive systems are housed in a physical and physiological body. It is also situated because it receives stimulus from and acts upon a structured environment. As a scientific stance it means not to neglect the linkage between cognitive processes and the sensory-motor surfaces, or the structured environment in which cognition takes place. They also have to take into account the role of individual experience in cognition and be cautious about abstracting data from concrete neuronal processes that are the basis for behavior in organisms [Schöner, 2008]. However, the tests of theoretical process models based on this stance sometimes face unsurmountable practical challenges. It is reflected in the limited success of the AI approach to autonomous robotics [Brooks, 1991]. Dynamical systems theory can provide a language to develop an embodied and situated theory of cognition. They can provide us with an alternate metaphor for cognition by making us think of underlying "forces" (vector-fields) from which observed behavior/pattern emerges. In nonlinear dynamical systems, the solutions may change qualitatively even when the underlying vector-fields change in only a graded way [Schöner, 2008][6].

Similarly, a control systems perspective can work to derive meaningful function from simple, embodied mechanism. It was best illustrated by Poggio and

---

[6] This image dates back to the notion of Gestalt field in Gestalt psychology [Köhler, 1939].

Reichhardt [1976] in their study regarding the orientation behavior of a common housefly. The movement of the flies as a part of their mating behavior can be explained by a motion detection system fed by luminance changes on the facet eye of the fly driving the flight motor, which generates a torque in the right direction. This stable flight is an attractor solution for the dynamical system of the neural circuitry linking the sensory surface to the flight motor. The initial analysis of this kind led to wrong predictions in subsequent experiments [Heisenberg and Wolf, 1988]. This happened as the simple nervous system of the fly computed an expected visual motion from its own motor commands (reafference) and treated detected visual motion matching the prediction differently than other extraneous motion signals related to movement of an object relative to the fly.

However, if cognition could be founded upon the notion of control then heater with a thermostat would be a cognitive system. Schöner [2008] suggests *flexibility* as another dimension to distinguish cognitive systems. By flexibility we mean a threshold criterion. A control system can be said to have reached preliminary cognitive system when its emergent behavior can no longer be uniquely determined by its sensory inputs, but depends to some extent on the choices generated by its internal states.

The summary of the principles of dynamical systems approach to behavioral patterns as given by Schöner and Kelso [1988]

1. Behavioral patterns are characterized by internal states. A state can be characterized by variables with specific values. The variables and the associated states are not limited to those of sensori-motor loops, but may originate from within the neural networks that control behavior.

2. The time evolution of the system and the state variables is generated by neural networks and the dynamical forces are provided by the sensory inputs, internal sensory feedback, physical dynamics and material property.

3. Attractor solutions (structured by asymptotically stable states) are likely to be observed in the long run. The nervous system being exclusively interconnected for a particular circuit or pattern other connections, sensory inputs and changes in the environment act as perturbations.

4. Instabilities (bifurcations) release the states from stability and give rise to flexibility.

Dynamical systems approach offers a more concept-based rather than model-based accounts of cognition. It captures the individual contributions to the neural dynamics in its *multi-functional* form and the mental states as *multi-causal* (there are a multiple ways a stable state might emerge and as well as disappear again). It is aimed at understanding constraints of learning and development, or how individual differences might manifest themselves in different contexts.

1.7 Organization of the thesis

The present thesis deals with a dynamical systems approach to certain important aspects of the visual system - visual sense of numbers, visual short-term memory, and time's subjective expansion for visual stimuli in oddball paradigm. In Chapter 2 we briefly introduce the literature regarding the theoretical and empirical studies on the above topics. In Chapter 3 we present a dynamic recurrent neural network architecture that forms the basis of saliency maps in visual-spatial domain, and show how it can capture most important aspects of visual sense of number and working memory and its involvement in temporal oddball paradigm. In Chapter 4 we give account of empirical experiments designed to test the predictions of the computational framework given in Chapter 3 for visual sense of numbers. In chapter 5 we describe a series of experiments designed to test the model predictions regarding spatial cuing and serial recall. In Chapter 6 we show how the same dynamical network architecture can explain and predict human behavioral performance of time's subjective expansion in temporal oddball phenomena along with a couple of experiments to verify the model. The conclusion of the thesis are given in Chapter 7. Appendix A gives the mathematical details of neural network. Appendix B is a small philosophical exploration of space and time in human mind.

Chapter 2

Working Memory, Numbers and Time: An Overview

> Clouds come floating into my life,
> no longer to carry rain or usher
> storm, but to add color to my
> sunset sky.
>
> *Stray Birds*
> Rabindranath Tagore

2.1 Introduction

> The relation between what we see and what we know is never settled. Each evening we see the sun set. We know that the earth is turning away from it. Yet the knowledge, the explanation, never quite fits the sight.
>
> -John Berger, *Ways of Seeing*

Vision is one amongst the five senses. Although the classification into five senses is rather arbitrary (even noticed by Plato). However, the Aristotelian distinction of five senses (Fig. 2.1) carried on, because of their localization into sense organs [Ross, 1931]. It was not until the mechanical world view of enlightenment that the classification into five senses sparked renewed interests. For instance, Galileo considered vision to be troublesome because, it was the only sense that defied mechanical-causal explanations. Vision was the only sense which was not immediate and also lacking a physical force that could be empirically described with the science of that age [Piccolino and Wade, 2008].

We have come a long way from Galileo's times. However, even in contemporary cognitive and psychological discourse, vision has a special place because it seems to dominate the other senses, as evidenced by the recovery process of experiential blindness [Nöe, 2004]). In recent literature, it also appears that vision has played a major role in human evolution. Homo sapients remained the only survivors of contemporary hominid groups because of invention of projectile weaponry, unlike their evolutionary cousins [Shea and Sisk, 2010]. Use of projectile weaponry demands a more advanced visuo-spatial cognition [Webster and Ungerleider, 1999] as evidenced by superior parietal lobes in human beings as opposed to other primate species [Hyvärinen, 2012].

Visual sense is roughly a combination of several different abilities at once. It involves: extremely fine categorization of objects [Harnad, 2005], extracting detailed features of the environment in the immediate vicinity [Marr, 1976, 1982], attending to objects in space and over time [Kahneman et al., 1992; Treisman, 1988], coarse estimation of distances and shapes of far away objects in the environment [Holway and Boring, 1941; Ono, 1966], detailed as well as extremely flexible representation of letters and geometric objects [McCandliss et al., 2003], crude representation and categorization of non-foveal objects [Alvarez, 2011], amongst others. Not only does vision involve doing all of the above in real time, it also involves these representations and estimates to be kept in a limited short term memory system, while being interacting with the Long term memory [Awh and Jonides, 2001; Awh et al., 2006; Corbetta et al., 1998; Desimone, 1996]. Although Pylyshyn [1999] has questioned the cognitive penetrability of vision, visual representation and attention form very important parts of human cognitive experience.

2.2 Vision and attention

> Everyone knows what attention is. It is the taking possession by the mind in clear and vivid form, of one out of what seem several simultaneously

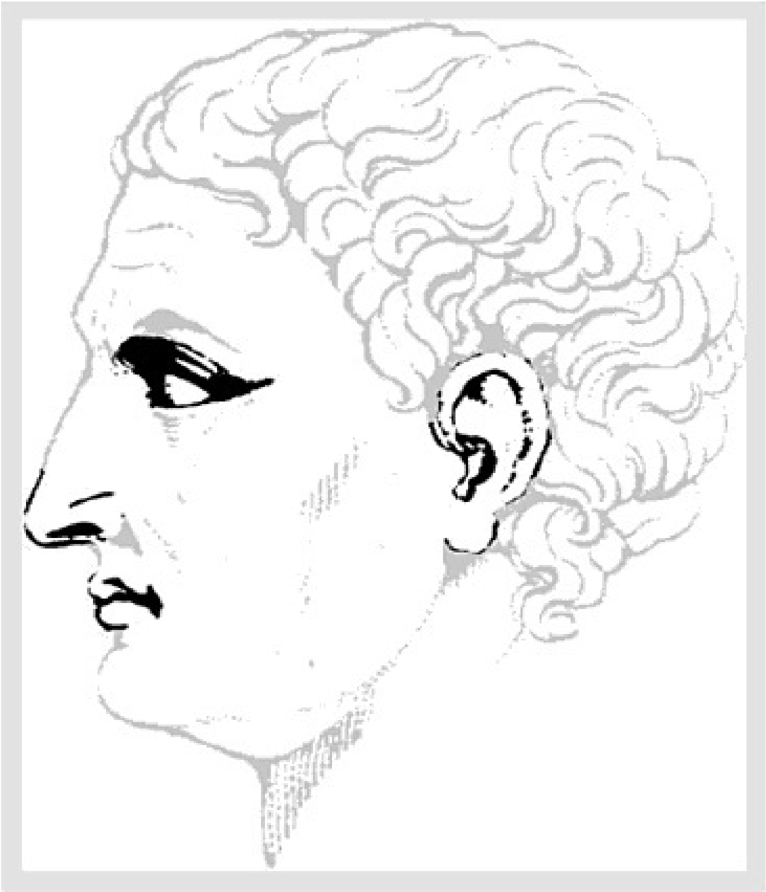


Figure 2.1: Nicholas Wade, *Aristotle's Senses.* A partial portrait derived from an engraving in *The Historic Gallery of Portraits and Paintings* (Vol. 5. 1810, Vennor, Hood and Sharpe, London). Specification of the senses of seeing, hearing, taste and smell was straightforward, but problems were caused by skin senses, which he variously called touch or feeling.

> possible objects or trains of thought...It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state.
>
> -William James

2.2.1 Stages of visual processing

The sensory information enters the primary visual cortex through lateral geniculate nucleus (LGN). From visual cortex information is processed paralelly across two pathways [Milner and Goodale, 2008] leading to pre-frontal cortex (PFC) - 1) dorsal pathway via posterior parietal cortex (PPC) 2) ventral pathway via inferotemporal cortex (IT). Lesion studies [Farah, 1990] show that damage to IT leads to *visual agnosia* (inability to recognize objects by sight). On the other hand damage to PPC leads to *optic ataxia* (the ability to voluntarily shift gaze to new location of the same object is severely diminished, although patients have no trouble in recognizing it) and *optic apraxia* (related to optic ataxia, where patients are unable to reach to the location of peceived objects) [Robertson, 2003]. These and similar evidences [Deubel et al., 1998] has led to the understanding that the ventral stream processes the information regarding identity ('what') of visual objects and the dorsal stream processes spatial information ('where') regarding the same (See Fig. 2.2). However Milner and Goodale [1995] have opposed the idea of a single external location processing in favor of the idea that the dorsal stream is involved in several spatial-motor representations operating in parallel supporting several dfferent motor actions [Milner and Goodale, 2008]. It is also known that the streams are not uni-directional, but have feedback re-entrant pathways almost al all levels from visual cortex upto PFC [Itti and Koch, 2001].

However the during these pathways visual information undergoes several important routines of selection, competition and binding [Desimone and Duncan, 1995; Styles, 2005]. All these processes are mediated through visual attention [Desimone and Duncan, 1995]. In the following we will try to give a brief overviw of cognitive psychological account of visual attention.

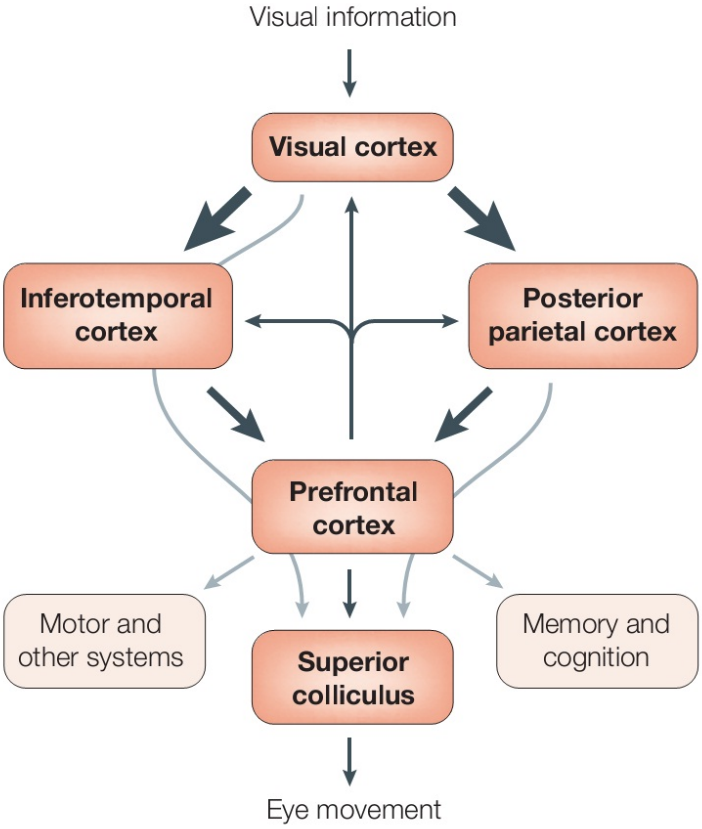


Figure 2.2: The dual path of visual procesiing. Visual information transmitted through the optic nerves enters the visual cortex in occipital region primarily through lateral geniculate nucleus (LGN, not shown) and superior colliculus (SC). There are two pathways from visual cortex to higher brain regions like prefrontal cortex (PFC) - 1) dorsal stream through posterior parietal cortex (PPC), 2) ventral stream through inferotemporal cortex (IT). Dorsal stream processes the location aspect of vision, whereas the ventral stream is more concerned with object recognition and identification. Dorsal stream is important for attentional deployment although ventral stream also receives attentional feedback. Reproduced from Itti and Koch [2001]

2.2.2 Visual attention

Eyes take different cues for converting the 2-D projection of the 3-D world on the retina into a proper realistic picture of the world. These cues come from convergence of the eye, stereopsis [Hubel and Wiesel, 1970; Poggio and Poggio, 1984], familiar size (perceptual constancy of size), texture density gradient, linear perspective, shape and shading, motion parallax etc[1] [Styles, 2005]. So, we select/filter the relevant information from the visual sensory input in order to construct visual representations. Selective attention is the name given to the bundle of processes that achieve the desired selection through different formal mechanisms [Logan, 2004].

Visual attention can be either *endogenous* (controlled by observer) or *exogenous* (sudden movement that captures attention). Posner [1980]'s experiments in this field reveal that visual attention can be covertly oriented in absence of overt eye movements. He also showed that exogenous system automatically shifts attention and cannot be ignored. The specific subset of selective attention that deals with attending to non-foveal locations is refered to as *spatial attention*[2]. It has also been shown through electrophysiological studies of monkey cortex that neurons from different cortical areas show differential selection to features [Fellaman and Van Essen, 1991]. Moreover all these different properties of visual cognition is processed through different specialized cells and this gives rise to the binding problem [Robertson, 2003; Styles, 2005]. Treisman and Gelade [1980] proposed a feature integration theory of focal attention (FIT) to overcome it (See Fig. 2.3.

All these different aspects of perception and attention are brought together in FIT. Different features are assumed to be coded into feature maps. Separate features can be combined by memory driven (top-down) process, or focal attention selecting a location on a master map. Some recent works by Awh and Jonides [2001]; Bisley and Goldberg [2006] and others have shown that it is possible that a spatial map based architecture might exist in the parietal cortex. Robertson

---

[1] Neisser [1967] and Gregory [1998] thus support the *constructivist approach to cognition* which asserts that 'we perceive the world as an active, constructive process'.

[2] The temporal dimension of visual processing is explored within the episodic model of attention [Sperling and Weichselgartner, 1995] which has evolved into elaborate attention-gating models.

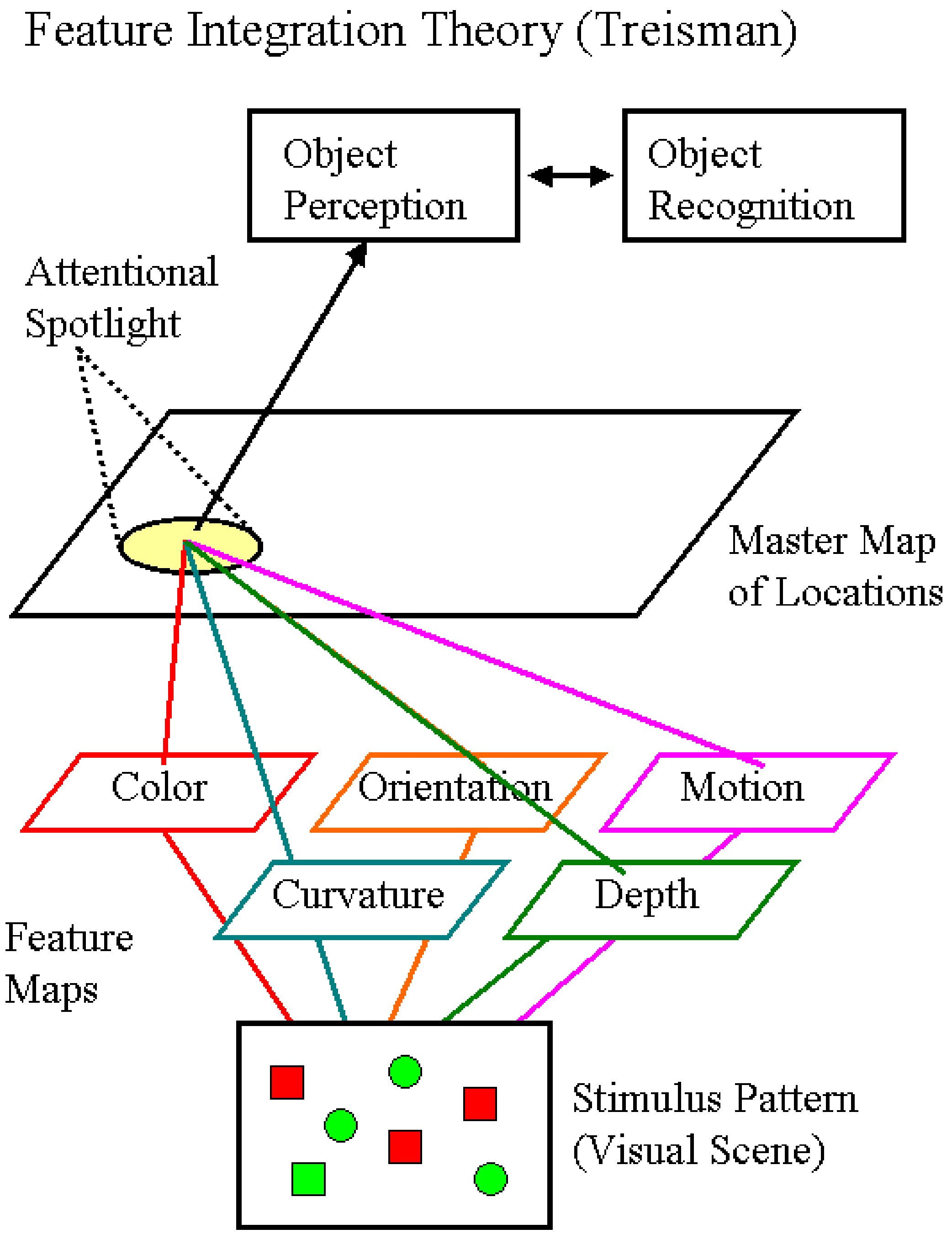


Figure 2.3: Feature Integration Theory by Treisman and Gelade [1980]. It shows that feature maps for color, orientation etc. for visual objects are collected separately in several feature maps before being integrated on a master map of locations throgh attentional selection and focus.

.

[2003]; Robertson et al. [1997] have shown from their work on patients with *Balint's syndrome* [3]

> It is not that objects require perception of space to rise above the threshold of perceptual awareness – both features and objects can be perceived without spatial information provided by parietal input – but rather that awareness of the relations between features represented in different cortical feature maps and the spatial relations between objects requires explicit spatial information. -Robertson [2003]

An alternative account to binding is given by biased competition model (BCM) [Desimone and Duncan, 1995] that depends on continuous interaction between competing neurons with different preferred features. According to BCM feature conjuction should become difficult with more number of distractors added to a display. However it is noted by Robertson [2003] that features and conjuctions activate parietal cortex differentially even after controlling for difficulty. Thus, the evidence supports more modular interaction between dorsal location maps with ventral feature maps like the ones proposed by FIT.

In a changing natural scene the efficiently encoded objects through spacial attention are kept in working memory as the attention to them is sustained [Awh et al., 2006]. In fact, spacial attention and spatial working memory share common neural pathways in the brain [Awh et al., 2006; Corbetta et al., 2002]. In the following we will try to disambiguate the two faculties.

2.3 Attention, memory and individuation

> "Memory's images, once they are fixed in words, are erased," Polo said. "Perhaps I am afraid of losing Venice all at once, if I speak of it, or perhaps, speaking of other cities, I have already lost it, little by little."
>
> Italo Calvino, *Invisible Cities*

---

[3] Severe damage to parietal lobes in both hemispheres leading to *simultagnosia* (inability to perceive more than one object), optic ataxia and optic apraxia.

One of the important functions of vision is to provide a representation of the world that can be kept in our short term memory to guide action. Visual processing of objects in space involves both selective spatial attention [Melcher and Piazza, 2011; Roggeman et al., 2010] and visual working memory [Awh et al., 2006]. It is well known that visual short term memory and spatial attention share similar mechanisms and neural substrates [Awh et al., 2006; Corbetta et al., 1998, 2002]. Studies on primates and adult humans have indeed identified common neural substrates that are activated for visual spatial attention and visual short term memory [Awh and Jonides, 2001; Bisley and Goldberg, 2006]. Most common tasks used to probe visual spatial attention involve spatial cueing for target detection [Posner, 1980], visual search [Müller and Rabbitt, 1989], and change detection [Fernandez-Duque and Thornton, 2000]. These paradigms generally elucidate that spatial cuing when valid (i.e., the cue gives truthful information regarding the location of the target), facilitates effective encoding and detection of the target stimulus. This facilitation is measurable through reaction time and accuracy. Invalid cues, however, are disadvantageous to target detection in terms of RT and accuracies. On the other hand, most common visual short term memory tasks involve recall of items presented serially in the same spatial location [Amiez and Petrides, 2007], or change detection for items presented simultaneously on different locations on the screen [Todd and Marois, 2004; Vogel and Machizawa, 2004; Xu and Chun, 2005]. Serial presentation paradigms generally tend to follow the standard serial recall curves for *primacy* [B B Murdock, 1962] and *recency* effect [Waugh and Norman, 1965]. Primacy and recency effects refer to better recall of early and late items in a sequence of words/objects presented serially (See Fig. 2.4). It is generally thought to reveal the limited capacity system of short-term memory [Miller, 1956; Todd and Marois, 2004].

In simultaneous presentation paradigms with simple color or shape stimuli, participants can usually detect changes accurately where there are four or less items [Vogel and Machizawa, 2004; Xu and Chun, 2005]. However, there have been few studies that probed the effect of spatial cueing on serial recall.

This brings us to the point that spatial attention encodes objects in the visual space and the encoded objects are kept in visual working memory [Awh and Jonides, 2001; Roggeman et al., 2010]. Some researchers [Bisley and Goldberg,

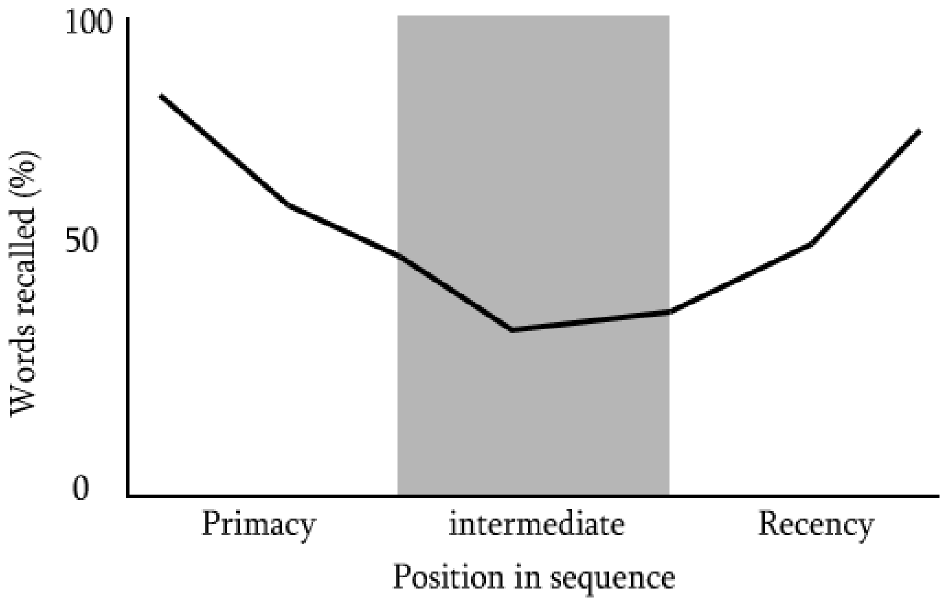


Figure 2.4: Primacy and Recency effect illustrated for word recall.

2006; Melcher and Piazza, 2011; Roggeman et al., 2010] have suggested possible spatial topographic maps of the world in lateral intraparietal (LIP) region of the brain, might mediate both the mechanisms. These maps underlie individuation process for objects in visual domain [Melcher and Piazza, 2011; Piazza et al., 2011], and so also facilitate the sense of 'how many', along with 'what' and 'where'.

2.4 Numerosity, saliency and spatial attention

> Real museums are places where Time is transformed into Space.
>
> -Orhan Pamuk, *Meuseum of Innocence*

One of the important theoretical insights in the field of visuo-spatial attention over last few years has been regarding the visual sense of number [Burr and Ross, 2008a,b]. The number sense allows us to estimate the number of objects in front of us and this sense has a developmental counterpart [Xu and Chun, 2005]. Some studies on cotton-top tamarins [Hauser et al., 2003] point towards the possible evolutionary origins of the visual number. It has also been shown that number sense is possible even in societies where language does not have words for distinguishing higher numbers [Gordon, 2004]. Visual number sense also seems to distinguish between a range of smaller numbers (varied range, generally less than 6) or subitizing range [Kaufman et al., 1949] and the range of higher numbers. Numbers in subitizing range are characterized by their rapid and confident enumeration with very high degree of accuracy by human beings. For higher numbers humans use one of the two strategies - 1) precise sequential counting, and 2) estimation, a simultaneous enumeration. Counting is precise but is much slower than either subitizing or estimation. Estimation, although is faster than counting, is much less precise [Whalen et al., 1999]. The separation of these two ranges has obviously produced questions regarding the mechanisms underlying the processes of subitizing and estimation. The debates center around the question whether subitizing and estimation share a common mechanism and thus possibly a common substrate or not.

To even begin to understand how to disambiguate the answer we have to address certain complexities. In the visual domain enumeration depends at least in part on object individuation, a visuo-spatial mechanism that allows us to locate and track objects in space [Melcher and Piazza, 2011; Piazza et al., 2011]. The process thus involves factors like color, texture, etc that have nothing to do with numerosity *per se*. Variable areas of objects indicate that numerosity can not have a simple solution on the basis of cumulative area [Dehaene and Changeux, 1993]. There has been attempts to understand subitizing as estimation mechanism operating at low numbers [Dehaene and Changeux, 1993; Gallistel and Gelman, 1992]. However, the different Weber fractions[4] over the two ranges suggest otherwise [Revkin et al., 2008]. Trick and Pylyshyn [1994] have suggested subitizing might employ pre-attentive mechanisms unlike estimation. However, Burr et al. [2010] have systematically manipulated both spatial and temporal attention to show that such manipulations indeed affect subitizing performance, which led them to suggest possible common pre-attentive mechanism for both

---

[4] Weber fraction or Weber ratio refers to the minimum relative change in stimulus intensity in order for the stimulus level to be be perceived as different from a reference stimulus intensity.

subitizing and estimation with possible additional attentive mechanisms necessary for enumeration within subitizing range.

In order to explore possible computational strategies for enumeration Dehaene and Changeux [1993] developed a reinforcement based supervised learning approach with proposed extension towards self-organized model to explain possible learning mechanisms in infants. However this model restricts enumeration only up to five items. As another possible solution Stoianov and Zorzi [2012] have trained 'deep' networks to use pixel by pixel information of images through unsupervised learning. They show that numerosity detectors can emerge in the highest level of the generative network. Their model could account satisfactorily for numerosity comparison task data in monkeys and human adults within estimation range. These numerosity detectors may also have physical basis in 'number neurons' suggested by Nieder et al. [2002] and Roitman et al. [2007]. Both the models mentioned above have complex structures that allow for learning of numerosity detectors. Another interesting model was developed by Grossberg and Repin [2003], based on on-center off-surround architecture, involving an interaction between spatial number map and semantic categories to explain error rates and reaction times in human numerosity comparison data. In the present work we wanted to work out what properties should a network of numerosity detectors have in order to explore enumeration across both small and large numbers. In the present work we wanted to work out what properties should a network of numerosity detectors have in order to explore enumeration across both small and large numbers. Interestingly, it appears that after one postulates the numerosity detectors as neural nodes that receive some kind of normalized input from lower domains, the earlier confounds like area, texture etc do not affect the computation as they have already been pre-processed.

2.5 Attention, time, and perception

A fundamental part of our cognitive experience is the idea of subjective time, whether it is to reminisce past events or to make a decision regarding which route will be faster to take in rush hour traffic, or even to enjoy music. In the visual domain, time is necessary to keep track of order, movement etc., for different stimuli. It has been seen that subjective time experience is dependent on attentional focus, especially in duration judgments of persistent stimuli. Tse et al. [2004] have shown that the subjective expansion of time for novel stimuli in an oddball paradigm is mediated by attentional bottleneck and time's subjective expansion (TSE) does not happen for stimuli too transient (<120 ms) for attentional mechanisms. However, they subscribe to the idea that a subjective time is continuously in tune with objective time, and it is just that the attentional focus captures more information at oddball trials and thus giving rise to TSE. Tse et al. [2004] make a distinction between Time Ordered Errors and Times Subjective Expansion, following Adaptation Gating Model by Hellström [1985]. Tse et al. [2004] claim TOE and TSE though both attentional in nature, TOE occurs along other stimulus dimensions than time. Counter based theories [Thomas and Weaver, 1975; Treisman, 1963] as noted by Tse et al. [2004] can give an understanding of Oddball induced expansion in subjective time. It is suggested that TSE occurs due to attentional boost to information accrual leading to the effect observed. This is also supported by the fact that the stimulus needs to be shown for greater than 120 ms in order for TSE effect to be shown, as it needs the attentional mechanisms.

2.6 Statement of the problem

In the present thesis, I intended to outline a possible on-center off-surround neural network architecture that can account for enumeration, visual short-term memory, and time's subjective expansioncan by virtue of its integration of magnitude and order related measures. The network will be detailed in next chapter. The predictions and corresponding empirical verifications in the above domains of various cognitive phenomena are given in the chapters 4-6.

Chapter 3

Computational Modelling Framework using On-center off-surround network

> Most of an organism most of the time is developing from one pattern to another not from homogeneity into a pattern.
>
> *The chemical basis of morphogenesis*
> Alan M. Turing

3.1 Network architecture

3.1.1 Dynamics

We began by modelling critical features of a recurrent on-center off-surround network reported in Usher and Cohen [1999] and used in Roggeman et al. [2010]. It is essentially a saliency map model based on the nonlinear leaky competeing accumulation models (LCA) that have been used to account for performance in multiple-alternative choice paradigms [Bogacz et al., 2007]. These models capture the recurrent on-center and off-surround nature of activity observed in neural

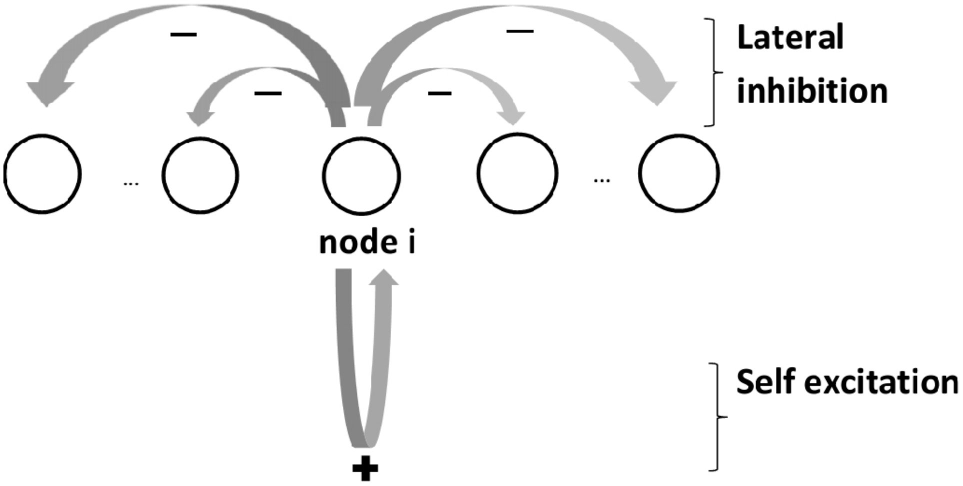


Figure 3.1: Illustration of the on-center off-surround model, showing nodes with excitatory and inhibitory connections.

systems [Grossberg, 1973; von der Malsburg and Buhmann, 1992] particularly in the visual modality.

The network consists of a single layer of completely interconnected nodes (Fig. 3.1). Each node corresponds to a neuronal assembly encoding an object or location of an object (or particular features) depending upon the cognitive phenomenon being modeled. The three main parameters that define the type of network are α (strength of self-excitation for each node), β (strength of lateral inhibition between nodes) and λ (decay constant for the passive decay term).

The differential equation governing the time-evolution of the network of N nodes is given by

$$\frac{dx_i}{dt} = -\lambda x_i + \alpha F(x_i) - \beta \sum_{j=1, j \neq i}^{N} F(x_j) + I_i + noise \tag{3.1}$$

$x_i(t)$ is the activation of node i at time t. I_i represents the intensity of external input ($\forall i, 0 \leq I_i \leq 1$). In our simulation I_i is a unit step function, i.e., it has the value 1 for certain number of time steps for the particular node i and 0 for rest of the time steps. Input is only presented for a finite amount of time, typically much less than total time of simulation. $F(x)$ is the activation function given by the formula,

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \frac{x}{1+x} & \text{for } x > 0 \end{cases} \tag{3.2}$$

The total number of nodes receiving input is referred to as set size in rest of the article. In order to remain close to the cortical neuronal dynamics we chose to keep the decay parameter $\lambda = 1$.

We modeled the dynamics according to the discrete form of the Eq. 3.1. The activation of the nodes are updated at each step according to the following equation,

$$x_i(t) = \alpha F(x_i(t-1)) - \beta \sum_{j=1, j \neq i}^{N} F(x_j(t-1)) + I_i + noise \tag{3.3}$$

Following the example of Roggeman et al. [2010], we used a network of 70 nodes ($N = 70$) for the present work as higher values of N do not change the results qualitatively. All the simulations were run at fixed $\alpha = 2.3$[1]. The input was presented for 5 time steps and the simulations were run for a total of 50 time steps. A normal distribution of mean 0 and standard deviation 0.03 is used to sample the *noise* at each time step. As can be seen in Eq. 5 in Appendix A, that in the absence of external input the network settles to steady state dynamics characterized by its steady state activation $x(n)$, where n is the number of nodes in steady state activity. This quantity is uniquely determined by the network parameters α, β, and n (but not on the size of the network, i.e, N). So we use mean activation ($\bar{x}(n) = \sum x_i/N$)[2] as the main output of the network for the remainder of the chapter. Table 3.1 lists the major parameters used for simulation throughout the study.

---

[1] The excitation parameter was chosen in line with Usher and Cohen [1999]. The excitation parameter was varied systematically between 2.0 and 2.4 to check for stability. The convergence results are shown below.

[2] Here x_i is the final activity of the node i at the end of simulation. It is either 0 or close to the value given in Eq. 5 in Appendix A. Or written in another way, $\bar{x}(n) = nx(n)/N$ in absence of noise, where $x(n)$ is the steady state activation values of the nodes as given by Eq. 5.

Parameter	Value
N	70
α	2.5
β range	0.01-0.15
Set size range	1-64
Duration of stimulus presentation (No. of time steps)	5
Total duration of simulation in time steps	50
Number of simulations run to compute the average	100

Table 3.1: Simulation parameters

3.1.2 Behavior

The recurrent nature of the network with its lateral inhibitory connection between nodes and self-excitations create a push-pull effect that drives the dynamics to a stable equilibrium. For instance if the network has high inhibition only relatively few of the nodes may remain active at equilibrium, no matter how many nodes were excited initially. In the low inhibition condition however even small perturbations created by noise can get amplified due to self-excitation in that particular node. Thus even if a small number of nodes were excited initially in a low inhibition network, there might be a large number of them in the final equilibrium state. The inhibition also controls the mean activation levels that can be reached in equilibrium. Fig. 3.2 and 3.3 illustrate the activation patterns for different set sizes at different inhibition levels. So higher inhibition acts like more focused (local) attention with capacity limits and lower inhibition acts like diffused (global) attention.

The parameters for the simulations were chosen after careful exploration of the parametric space given in Appendix A. It suffices to say that the network behaviour is robust over a range of simulation parameters and the results not mere trick of parameter choice.

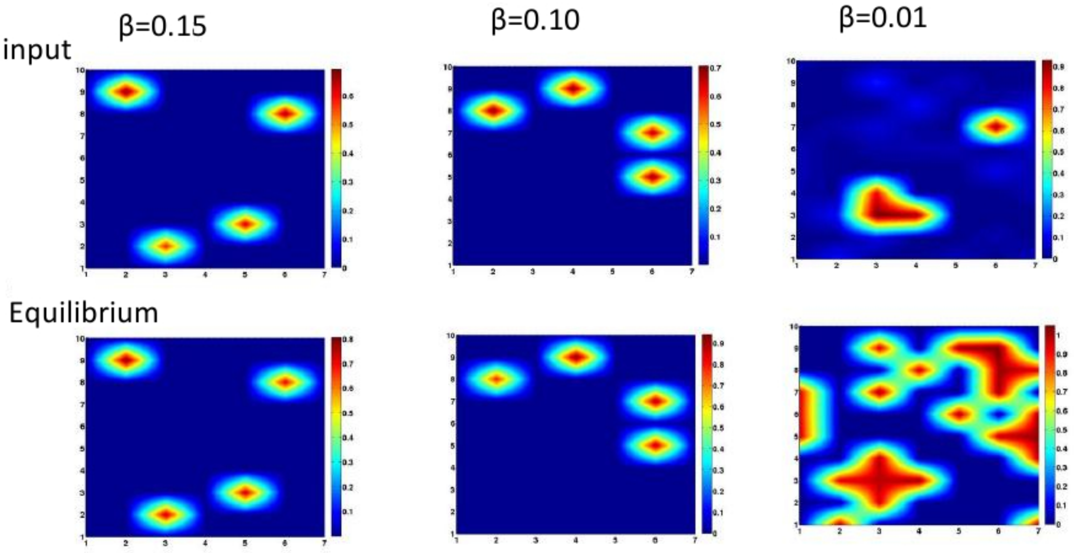


Figure 3.2: Network dynamics for a small number of inputs (4) at different levels of β. Top panels show the initial activations and bottom panels show final activation on a 10 x 7 grid representing 70 nodes ($N = 70$)

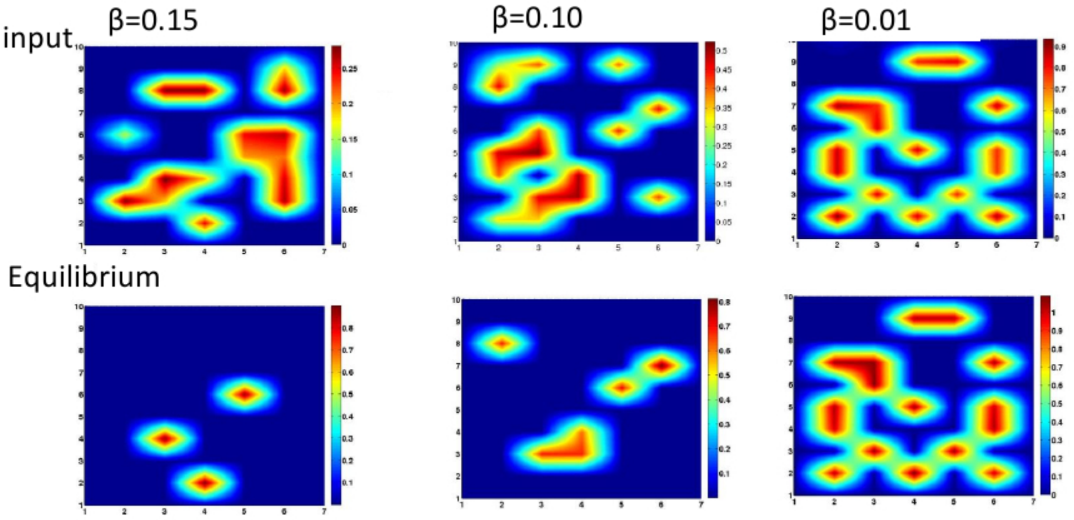


Figure 3.3: Network dynamics for a large number of inputs (16) at different levels of β. Top panels show the initial activations and bottom panels show final activation on a 10 x 7 grid representing 70 nodes ($N = 70$)

3.2 Modelling of Visual Sense of Number

3.2.1 Subitizing and estimation from network output at different inhibition levels

Considering that the transient input is converted into steady state activity for the current network, we wanted to see the sensitivity of the network output, i.e, the mean activation ($\bar{x}(n)$) with the parametric variation of β and set size. We considered that the parameters suitable for numerosity estimation will be the ones that give a monotonically increasing variation of the mean activation with the increase in numerosity of the inputs. The plot of mean steady state activation of the network as a function of the set size and the inhibition parameter is given in Fig. 3.4. The input was clamped at level 0.33for this and subsequent simulations. Other simulation parameters are given in Table 3.1.

It is evident that at medium inhibition the total activation of the network is sensitive to a lower range values of numerosity (see the lower plot in inset of Fig. 3.4) while low inhibition seems to function better for higher numerosities (range of 15-40, see the upper plot in inset of Fig. 3.4). Moreover, the plot shows that at higher values of β the total activation of the network is fairly insensitive to higher numerosity and low β does not have monotonically increasing relationship with lower numerosities. The finding, that different inhibition parameters work best for representing different numerosities, confirms the previous report of Roggeman et al. [2010].

3.2.2 Truthfulness of representations

Although lower β-values are more sensitive to higher numerosities, the precision of the representation is compromised at lower inhibition values. In order to illustrate the point, we can construct a crude measure for the probability of faithfulness of representation ($P(\beta, setsize)$) as

$$P(\beta, setsize) = 1 - d_{i/o} \tag{3.4}$$

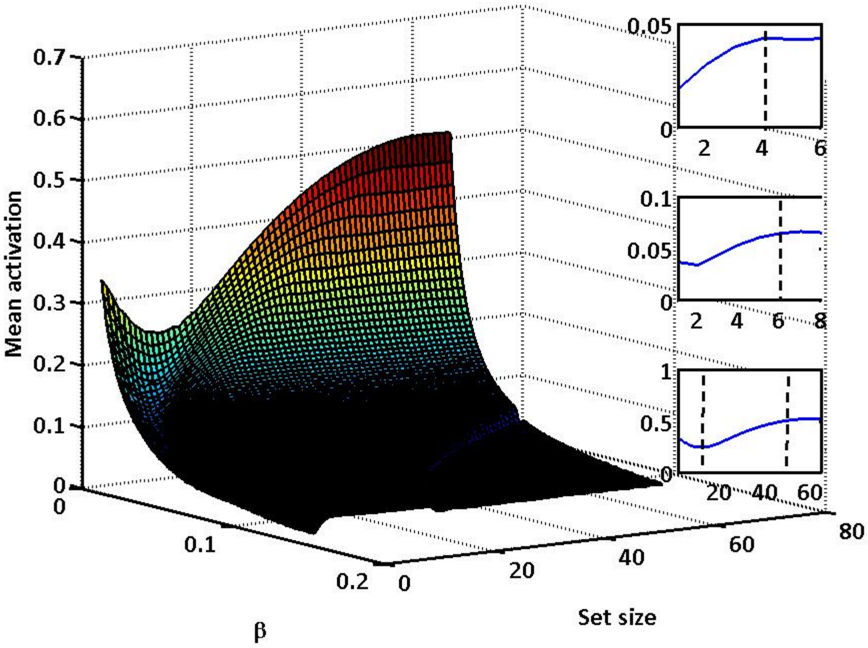


Figure 3.4: Mean activation vs. Set size and β. The plot was derived at $\alpha = 2.4$ as an average of 100 simulations. The insets show the mean activation vs setsize plot for three particular β values - low: 0.01 (bottom inset), medium: 0.1 (middle inset), and high: 0.15 (top inset). The verical line in the insets denote the limits to the monotonic region. For instance we can see from top inset that the mean activation ($\bar{x}(n)$) is monotonically increasing up to set size 4 for high β, but for low β (bottom inset) we can see that $\bar{x}(n)$ is monotonically increasing with numerosity for set sizes greater than 15.

where $d_{i/o}$ is the mean Hamming distance[3] between input activation pattern and output activation pattern over N nodes parametrized over inhibition parameter β and $setsize$. The simulation parameters are the same as in Table 3.1. Fig. 3.5 illustrates that at lower inhibition the maximum value of $P(\beta, setsize)$ is around 0.6, although the lower inhibition is better for estimation of higher numerosities (see Fig. 3.4). At lower β enumeration should be more prone to errors and inaccuracies due to more diffused representation. This finding is in line with numerous behavioral studies indicating that precision is reduced for estimation of large numerosities [Burr et al., 2010; Kaufman et al., 1949].

It is iteresting to note that, from a purely representational point of view, Fig. 3.5 differentiates three types of phenomena - 1) at high inhibition ($\beta =$ 0.15), focused representation of single object, 2) at medium inhibiton ($\beta \sim 0.1$), specialized subitizing and 3) at low inhibition, estimation. All these behaviors emergent from the self-organized behavior of the network.

3.2.3 Numerosity Comparison

In order to simulate human numerosity comparison data, we decided to calculate the probability that a given numerosity will be judged to be 'larger' than a given reference numerosity (16, in our case). The input was clamped at level 0.33. The probability of responding 'larger' is calculated by running 100 simulations for numerosity values 10-24 for different β values (the parameters of the network are the same as Table 3.1), as the proportion of times the mean activation of the network for a given numerosity is larger than the mean activation of the network for the reference numerosity by threshold value (δ =0.01). In other words the probability that numerosity m will be perceived 'larger' than a predefined reference numerosity ref is given by,

$$Prob_m(response\ larger) = \frac{\#of\ times\ \bar{x}_{ref}(n) - \bar{x}_m(n) > \delta}{\#of\ simulations} \tag{3.5}$$

where $\bar{x}_{ref}(n)$ and $\bar{x}_m(n)$ are mean activation values of the network when they are presented with numerosities ref and m respectively.

---

[3] Distance between two binary vectors that is equal to the number of bits that do not match between the vectors. For example, Hamming distance between 010101 and 011111 is 2.

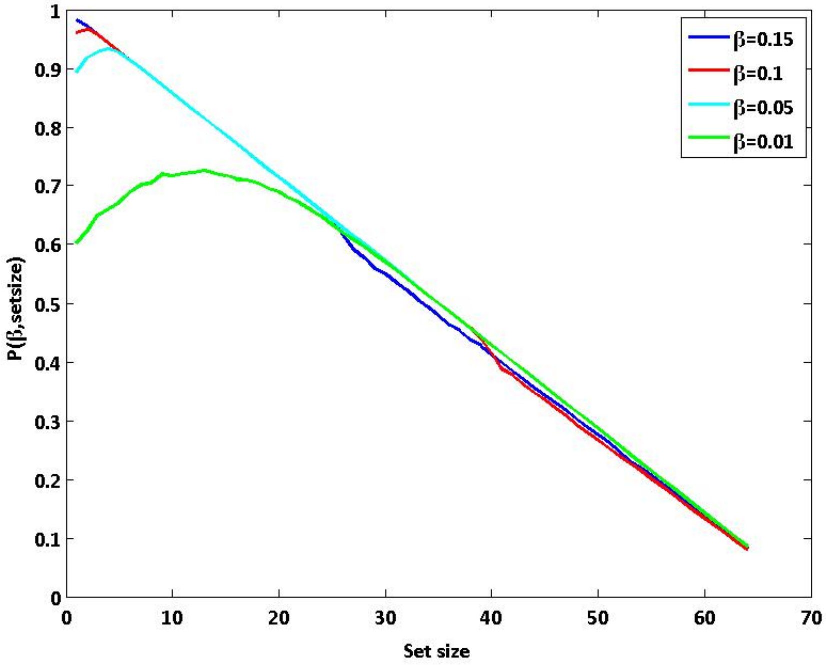


Figure 3.5: $P(\beta, setsize)$ plotted against $setsize$ for different β. The Modelling parameters are given in Table 3.1. The figure shows how probability of faithful representation decreases with decreasing β.

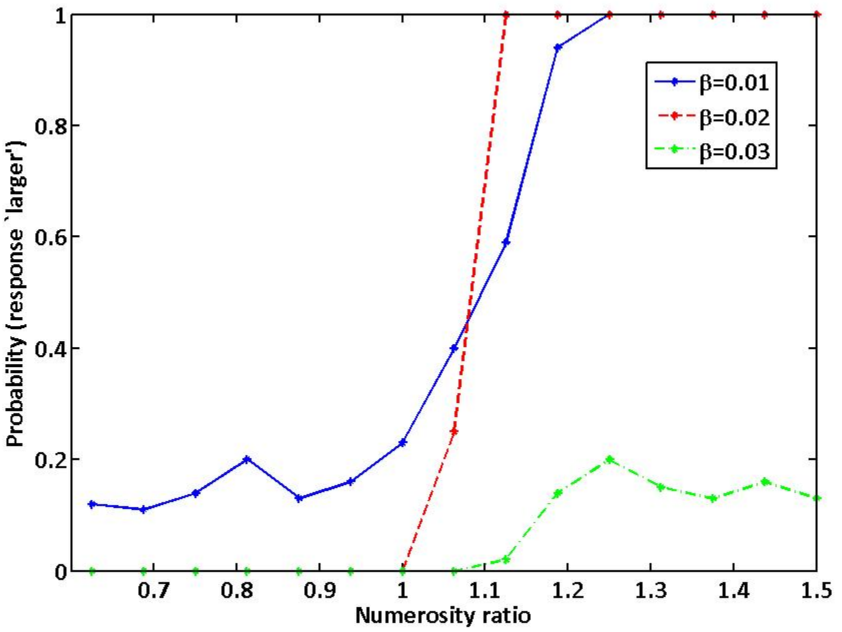


Figure 3.6: Probability of finding a given numerosity larger than the reference numerosity (16) plotted against numerosity ratios.

As shown in Fig. 3.6 performance differs depending on the β values. However, if we calculate the probability as an average over the different β values (0.01, 0.015, and 0.02), and perform a sigmoid fit against the numerosity ratio, we get a Weber fraction of 0.14 calculated at 75% performance value ($r^2 = 0.93$ at 95% confidence level), which is close to human data collected by Piazza et al. [2004] (see Fig. 3.7). Interestingly, at higher β values the probability of response 'larger' decreases and probability of response 'smaller' (i.e., 1−probability of response 'larger') increases (see Fig. 3.8) showing that it might be possible that when comparing numerosities at higher β, one might see underestimation.

For smaller numerosities, we wanted to see if our model shows the classic *distance effect* and follows *Fechner's law* [Dehaene and Changeux, 1993]. Distance effect refers to the finding that numerical distance being equal, the precision for discrimination judgment is always better at lower numerosity and Fechner's law refers to the universal phenomenon of degrading performance in discrimination judgments with increasing numerosity. Similar to Eq. 3.5 we calculate

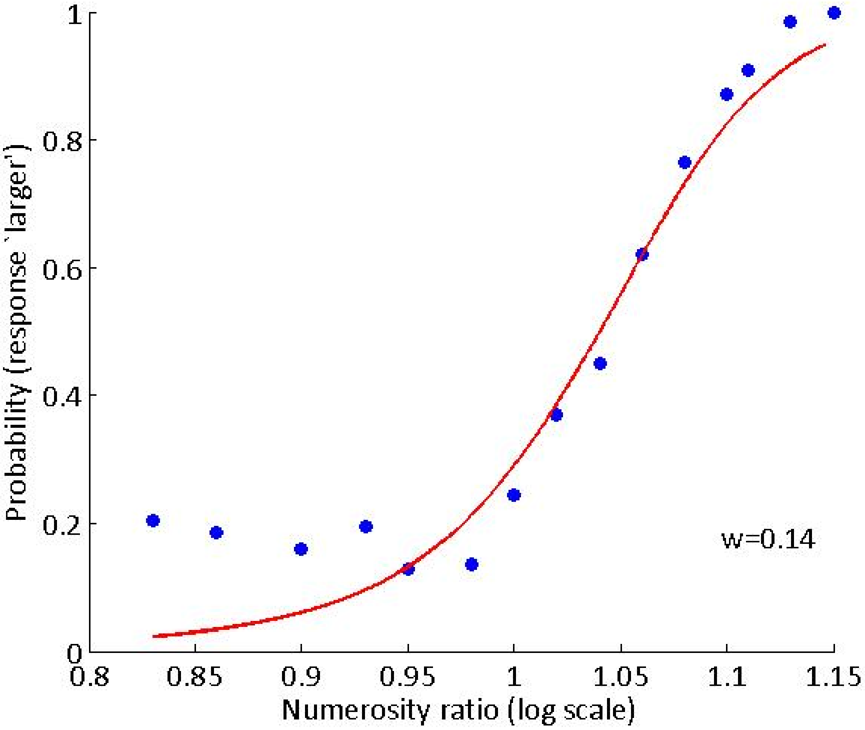


Figure 3.7: Simulation of human numerosity comparison data by averaging the probability values at different β (0.01-0.04).

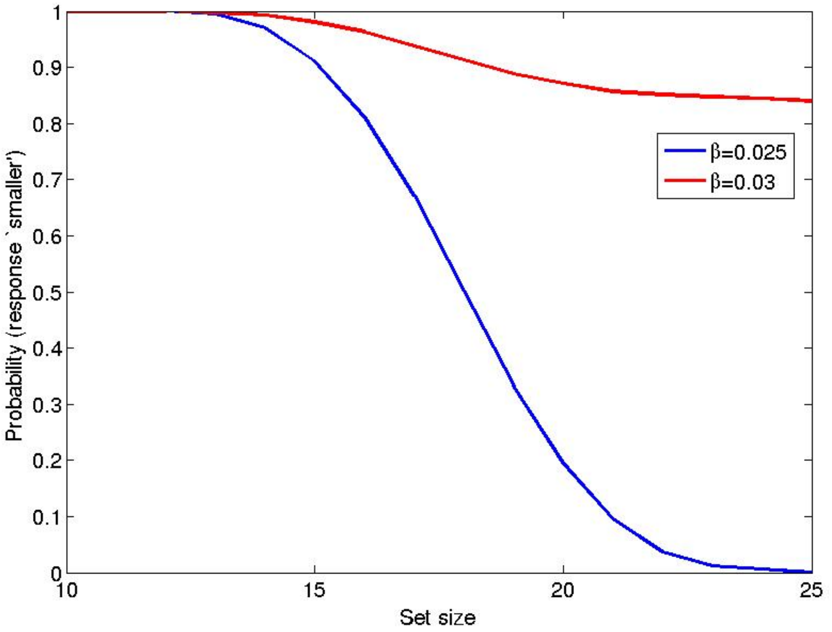


Figure 3.8: Probability of response 'smaller' for different numerosities for reference numerosity 16. The plot shows that how at higher inhibition numerosity underestimation is possible.

the discrimination probability of each numerosity $i = 1 - 5$ against the reference numerosity $ref = 1 - 5$ as,

$$Prob_i(response\ different) = \frac{\#of\ times\ |\bar{x}_{ref}(n) - \bar{x}_i(n)| > \delta}{\#of\ simulations} \tag{3.6}$$

where the symbols have their usual meaning. As we are operating in the subitizing range we chose to calculate the probability as an average of $\beta = 0.06$, 0.08, 0.1, 0.12, $0.14 - 0.15$ ($\delta = 0.01$) for this simulation. Fig. 3.9 represents the $Prob_i(response\ different)$ matrix for each reference numerosity (1-5) and compared numerosity (1-5), the area of the circles representing the magnitude of the probability (largest circles at top left and bottom right have value 1). Observing the probability values adjacent to the diagonal line, we can see that the overall precision decreases with increasing numerosity (i.e., from 1 to 5). Whereas for same numerical distance (like 2-3, 3-4, 4-5 pairs), we see that the probability for discrimination is higher at lower numerosity. Thus the Fig. 3.9 illustrates that the *distance effect* and *Fechner's law* emerge from the model.

3.2.4 Empirical evidence for inhibition parameter

In line with the results of Roggeman et al. [2010], the behavior of this on-center off-surround model depends on the inhibition between nodes (Fig. 3.1). In the case of high inhibition (top inset of Fig. 3.4), the pattern resembles that reported previously in fMRI studies of visual working memory, with activation increasing up to around 4 items [Kawasaki et al., 2008; Todd and Marois, 2004, 2005]. Varying the inhibition parameter (β), results in different patterns of activation. The graph for medium inhibition (middle inset of Fig. 3.4) successfully models the activity of fMRI data for enumeration tasks Roggeman et al. [2010]. Here the activation flattens out at about 6 items. For low inhibition we have the prediction as a concave curve shown in bottom inset of Fig. 3.4. The model's prediction of concave curve is well supported by data published in Roggeman et al. [2010]. Their data was collected for Lateral Intraparietal (LIP) area of the brain, and thus providing us with possible clue towards the localization of

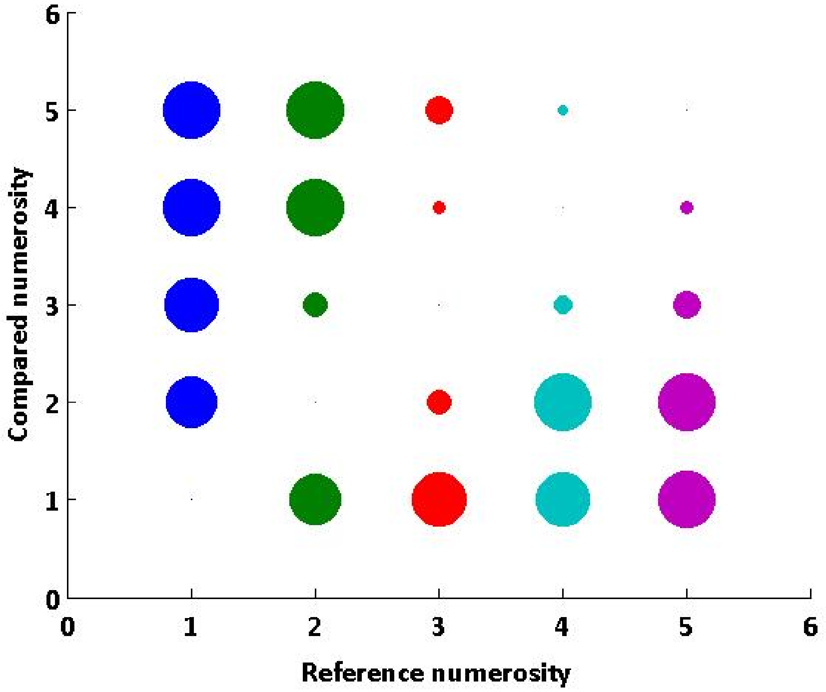


Figure 3.9: We show pairwise discrimination probability between different numerosities 1-5. The probability is proportional to the areas of the circles at each point. The largest area is 1 and the smallest ones adjacent to top right corner being 0.03. As we move towards the top right corner the circle sizes decrease, i.e., the overall precision of discrimination suffers with increasing numerosity. This exhibits *Fechner's law*. And the discrimination probability between 1-2, 2-3,3-4 and 4-5 pairs reveal the *distance effect*, where for the same numerical difference smaller numerosities are more accurate than the higher ones.

such kind of numerosity network. The location is fairly consistent with other numerosity related findings (e.g, [Roitman et al., 2007]).

3.2.5 Estimation of reaction times from the model

In order to explore the parametric space of the network we modeled an energy function on the following lines (see appendix B)

$$H = \sum_i H_i \propto - \sum_i \int \left(1 - \alpha \left(\frac{F(x_i)}{x_i} \right)^2 \right) \dot{x_i}^2 dt \tag{3.7}$$

$\dot{x}_i = \frac{dx_i}{dt}$ and H_i is the energy for a particular node i.

The reaction time distribution is assumed to correlate with the total amount of allowed fluctuation energy at a given state (as shown in Appendix B), i.e.,

$$RT \sim \sum_i \int \left(1 - \alpha \left(\frac{F(x_i)}{x_i} \right)^2 \right) \dot{x}_i^2 dt \tag{3.8}$$

The simulated values for RT were obtained as an average of 100 simulations. The Modelling parameters are given in Table 3.1. Fig. 3.10 shows the RT values for a given inhibition parameter and set size for a network with 70 nodes. It seems to be consistent with the general idea that subitizing (enumeration of small numbers) is faster than estimation (enumeration of large numbers.

However it also offers some very interesting features when we take our cue from the mean activation value of the network. From Fig. 3.4 we can see that higher inhibition ($\beta \sim 0.1$) is better for representing low numerosity ranges including the subitizing range, while lower inhibition ($\beta \sim 0.03$) is more suited for higher numerosity estimation. Fig. 3.10 (bottom inset) shows the estimated RT values for the different levels of inhibition (low or high). It shows a nearly linear increase in reaction times for numerosities in subitizing range . However, the reaction times for the high numerosity ranges remain very similar for a wide range of β value. This effect has been well known and demonstrated by Kaufman et al. [1949]. However the RT values for enumeration in low numerosity range at low β tend to be higher. So if a network tries to enumerate smaller numerosity at low inhibition, then the reaction time should be higher than reaction time for

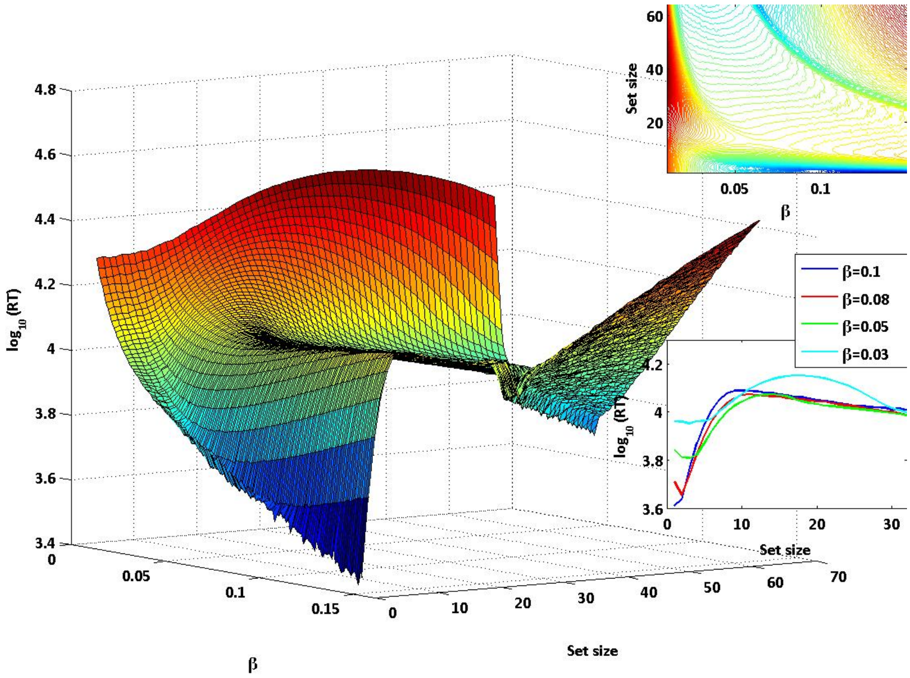


Figure 3.10: A parametric display of the reaction time distribution as calculated from the model. The values are meant to be proportional to log(RT) rather than directly representations. The top inset shows the contour plot for the energy surface and bottom inset shows the prediction for the reaction times for set sizes $1-32$ at $\beta=0.03$, $\beta=0.05$, $\beta=0.08$ and $\beta=0.1$.

enumerating the same at higher inhibition.

3.2.6 Model discussion

In the current work we presented a single layer recurrent neural network with complete interconnectivity between nodes that compete with each other through mechanisms of self-excitation and lateral inhibition on being presented with finite transient normalized input. Once the transient input is taken away, the network settles down to a steady state activity ($x(n)$) governed by the strength of lateral inhibition (β) between nodes. The network's behavior can be explored by varying β and observing values of $x(n)$ for different sizes of input (set sizes).

We have used mean activation ($\bar{x}(n)$) as the main output of the network in order to parameterize it against different β and set sizes. The condition that mean activation of the network should increase monotonically with set size gave

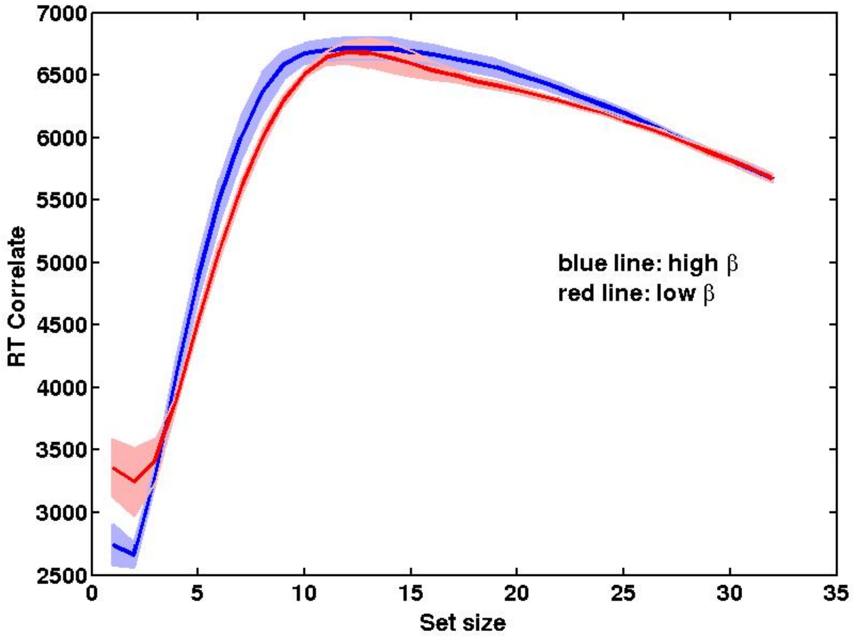


Figure 3.11: This figure shows the average RT values for two ranges of β: high-0.08-0.11 and low-0.03-0.06. The higher numerosity ranges hardly differ in RT between the two ranges of inhibition, however the low numerosity range differs quite significantly in RT between the two beta ranges.

us two regimes of set sizes governed by two different β values - higher β for smaller numbers and lower β for higher numbers (see Fig. 3.4) [4]. However, we also saw from the measure for the probability of faithfulness of representation ($P(\beta, setsize)$), that although lower β is needed for enumeration of higher numbers, the representation becomes more coarse and thus prone to errors. Moreover, we saw that although $P(\beta, setsize)$ is almost linear for higher numerosities, it is near constant for lower numerosities, thus indicating possible direction towards the reason that precision levels for enumeration differ significantly in subitizing and the estimation range (see Fig. 3.5).

We have also tried to extract the measure for the probability that a particular numerosity will be called 'larger' than a given reference numerosity (16, in our case) on the basis of the fraction of times the mean activation $\bar{x}(n)$ for that particular network is larger than that of the reference numerosity. Empirical data observed by [Piazza et al., 2004] shows that for reference numerosity 16, numerosity comparison data could be fitted to a sigmoid curve corresponding to Weber fraction (w) of 0.15. We saw that one particular β could not explain the spread of the data in this case, rather we had to calculate it as an average of different β values. We obtained a good fit (w=0.14, $r^2 = 0.93$ at 95% confidence level) to human numerosity comparison data collected by Piazza et al. [2004] (see Fig. 3.7). At higher β, the model could successfully capture (see Fig. 3.9) *distance effect* (for equal numeric distance, the same-different judgment performance is always better for smaller numerosities) and *Fechner's law* (precision of numerosity discrimination degrades with increasing numerosity).

Next we saw how the mean activation values for different set sizes are predictive of fMRI response data. The data from Roggeman et al. [2010] also shows that there is a different level of inhibition β involved in numerosity comparison as opposed to explicit enumeration. The β level found in their comparison experiment is line with the β values taken for Modelling numerosity comparison data. From these experiments LIP area emerges as a possible site for the proposed network.

---

[4] Here we have used the additive variant of on-center off-surround activity rather than the shunting one (where excitation terms are multiplied by $B-x$ and inhibitory terms are multiplied by $x - C$, where B and C are excitation and inhibition parameters, see Grossberg, 1973), as the additive function gives us the monotonic property of the mean activation of the network with numerosity unlike the shunting variant.

3.2.7 Comparison with other models

There have been limited attempts at a unified computational approach towards numerosity. This is due to the fact that human performances in subitizing and estimation ranges are clearly distinct - both in terms of performance measures like Weber fractions [Burr et al., 2010], precision [Revkin et al., 2008] as well as reaction time [Kaufman et al., 1949]. The differences in empirical data for the two numerosity ranges suggested the possibility of different enumeration mechanisms involved in subitizing and estimation. Dehaene and Changeux [1993] suggested a model based on supervised learning through ordered numerosity detectors and were able to demonstrate the *distance effect* and *Fechner's law* for numerosities 1-5. On the other hand Stoianov and Zorzi [2012] trained hierarchical generative networks to use the statistical properties of images in order to develop numerosity detectors in an unsupervised manner at the highest level of the network. Their work was able to simulate adult human numerosity comparison data really well. Our model falls somewhere in-between these two extremes. Our network is unsupervised to the extent that there is no reinforcement involved. However, the network has no learning component like the previous two networks. We can say that our network demonstrates the self-organization process required for the formation of numerosity detectors that arrive through learning in a hierarchical network proposed by Stoianov and Zorzi [2012].

We proposed a network that responds to numerosity as a normalized input and we used the output steady state activity to characterize the network response. Interestingly, it removes a very important limitation to the networks proposed by Stoianov and Zorzi [2012] and [Dehaene and Changeux, 1993]. The former does not address the subitizing range whereas the latter does not work well in the estimation range. Thus the earlier models seem to assume a break between estimation and subitizing ranges, respectively and model that accordingly.

The model developed by Grossberg and Repin [2003], based on a similar on-center off- surround architecture, was used to account for data in numerosity comparison tasks and provided an explanation of how the mapping from spatial to actual linguistic number categories might take place. In the current work, the question was what mechanisms might underlie the spatial map itself, whose activation may then be read off to produce actual responses through other higher order networks (similar to the one used in Grossberg and Repin, 2003). In agreement with Grossberg and Repin [2003] and also Roggeman et al. [2010], our results show how capacity limits can emerge out of inhibitory interactions between nodes in a spatial saliency map.

In our case, the constraint that network mean activation should monotonically increase with numerosity allows the two regimes to emerge naturally on the basis of different recurrent inhibition within the same network. Task dependent fMRI response correlates well with the predicted pattern for mean activation (e.g. in Roggeman et al. [2010]), and thus gives a possible physical basis for inhibition parameter as well as indicates a possible neural substrate in the networks of lateral intraparietal cortex (LIP) common to visual-spatial attention as well as visual short-term memory networks[Gottlieb and Goldberg, 1999; Vogel and Machizawa, 2004]. The network also seems to provide an interesting idea that numerosity comparison data emerges as an average response across a range of recurrent inhibition thus matching the success of Stoianov and Zorzi's network. The model also allows for variable precision between the different ranges of numerosity as evidenced by Fig. 3.5 and also able to account for *distance effect* and *Fechner's law* similar to Dehaene and Changeux [1993] (see Fig. 3.9).

The key feature of the present network is that it is able to give a comprehensive account of a host of observed phenomena within both subitizing and estimation range through a single layer network with one tuning parameter (β). Because of the simplicity, such network could be theoretically analyzed through Energy Function formulation (see Appendix A), which is a novel contribution of this work. The advantage of our network is that such a generic network can not only explain previous experimental data but also gave novel predictions. It can also be easily coupled with other components in order to explore learning and multi-modal interactions.

3.2.8 Possible mechanisms for modulating the inhibition parameter

An important feature of the model proposed here is that the network gives task-dependent response profile that are contingent upon varying the strength of lateral inhibition β. So it is natural to ask the question regarding possible mechanism for changing inhibition. Task dependent changes might be motivated by top-down influence. However, modulation of inhibition within the same task leads to the possibility that there might be an internal mechanism for β modulation. A simple additional summator node that receives input from all the nodes and gives feedback to the entire network based on a threshold might provide a mechanism for modulating the recurrent inhibition from within the network.

3.2.9 Possible implications of Modelling reaction time

Following the success of the network in predicting behavioral performances in numerosity comparison, we tried a different approach In order to extract new testable predictions from the model we formulated a possible expression for reaction time or it's close correlate. We took averages of the RT values at two different inhibition ranges (high: 0.08-0.11, low: 0.03-0.06) in order to elucidate what happens to RT when one tries to enumerate in the low numerosity range with low β or enumerate in the high numerosity range with high β. The model predicts an asymmetry or hysteresis in the reaction times for enumeration of small and large numerosities depending on whether they were preceded by a set of small or large numerosities. We reason that the network will generally be in a low β state if it is engaged in higher numerosity enumeration tasks and in a state of higher β if it is engaged in repeated enumeration tasks with smaller numerosity. Thus if the network is presented with a low numerosity enumeration task followed by a chain of high numerosity estimation tasks, the reaction times should be higher, whereas no such significant increase in reaction time should be observed in the other direction (i.e., high numerosity estimation task followed by a set of low numerosity enumeration task).

To get a clear prediction, computed the average RT values for low β range (0.03-0.06) and high β range (0.08-0.11). Fig. 3.11 clearly shows that task switch between low numerosity enumeration and high numerosity enumeration should yield asymmetric switch cost depending the switch was low to high or high to low enumeration. Moreover, we saw from Fig. 3.8 that enumeration of high numerosity from a higher β state increases probability of underestimation. Thus we arrive at two interesting predictions for an experiment, where a series of enumeration trials where varying lengths of low numerosity enumeration blocks are interspersed by varying lengths of high numerosity enumeration - a) The low numerosity enumeration followed directly by high numerosity estimation trials will have higher RT compared to standard low numerosity enumeration trials. b) The high numerosity trials directly followed by low numerosity enumeration trials will suffer from more underestimation effect than standard high numerosity enumeration trials. These predictions were verified in an empirical experiment detailed in Chapter 4.

3.3 Modelling of Effect of Spatial Cueing on Working Memory

3.3.1 Serial recall

This model was first developed in order to account for visual short-term memory. So we tested it for some classic visuo-spatial experiments. The first test involved serial presentation of stimuli in a visual working memory task. In line with previous studies showing capacity limits of 7 +-2 items [Miller, 1956], activating an increasing number of units results in a decrease in the average activation for each node (Figure 3.12). As the number of items exceeds capacity, earlier items would no longer be highly activated. In order to obtain a serial recall curve, we performed 500 such simulated trials (Figure 3.13). The inhibition parameter β was fixed at 0.1. All the inputs ere clamped at level 0.33 [5]. Other parameters are given in Table 3.1. Serial position curves were obtained by measuring the fraction of trials in which an item corresponding to a specific position remained active at

---

[5] Clamping here is analogous to electro-physiological practices where a constant d.c. like current with a certain amplitude is applied to the neuron/synapse for a finite time. Here the input amplitude is 0.33.

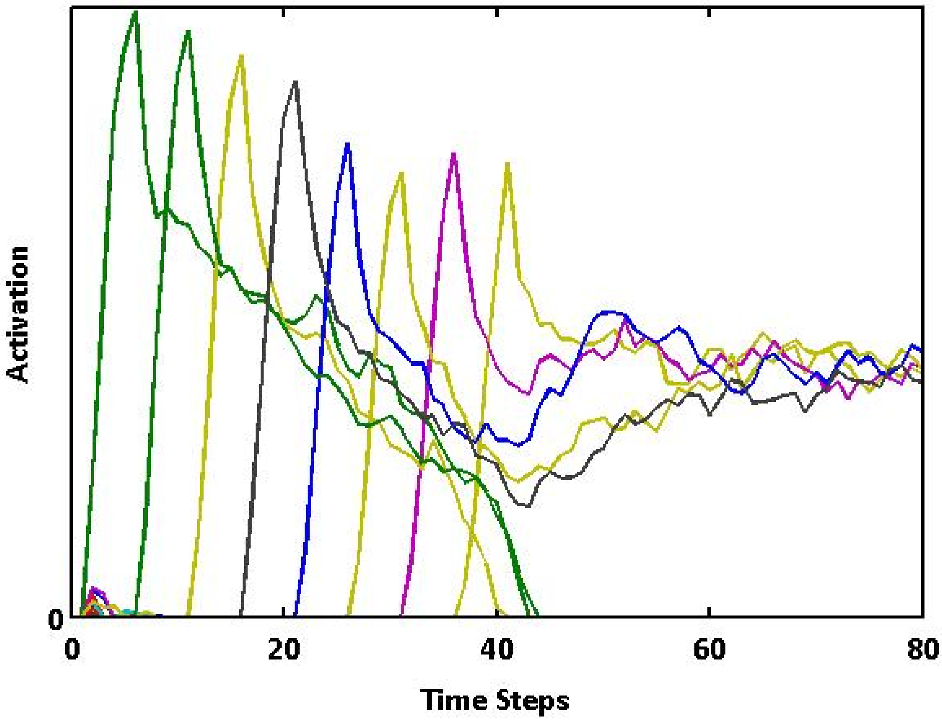


Figure 3.12: Application of the saliency map model to serially presented stimuli. Serial presentation dynamics of the model with 70 nodes. $\alpha = 2.0$, $\beta = 0.1$. Eight stimuli are presented serially for 5 time steps each. The input was clamped to 0.33 level.

the end of the simulation. Proportion of correct recall was estimated based on the simplified assumption is that if an item was still active in memory when the recall cue for it was given, then recall should be successful. Although we chose a particular criterion (0.2) for the minimum activation needed, the exact value was not critical to the overall pattern of results. As can be seen in Figure 3.13, the recency curve shared many characteristics found in experimental data, including dependence on the rate at which items were presented (Figure 3.14). In addition, the predicted performance was influenced by setsize, as expected from behavioral studies. Specifically, the model predicted complete recall for up to 4 items but worse performance for larger set sizes (Figure 3.15).

Interestingly, we also modeled the recall curve for the case when one of the

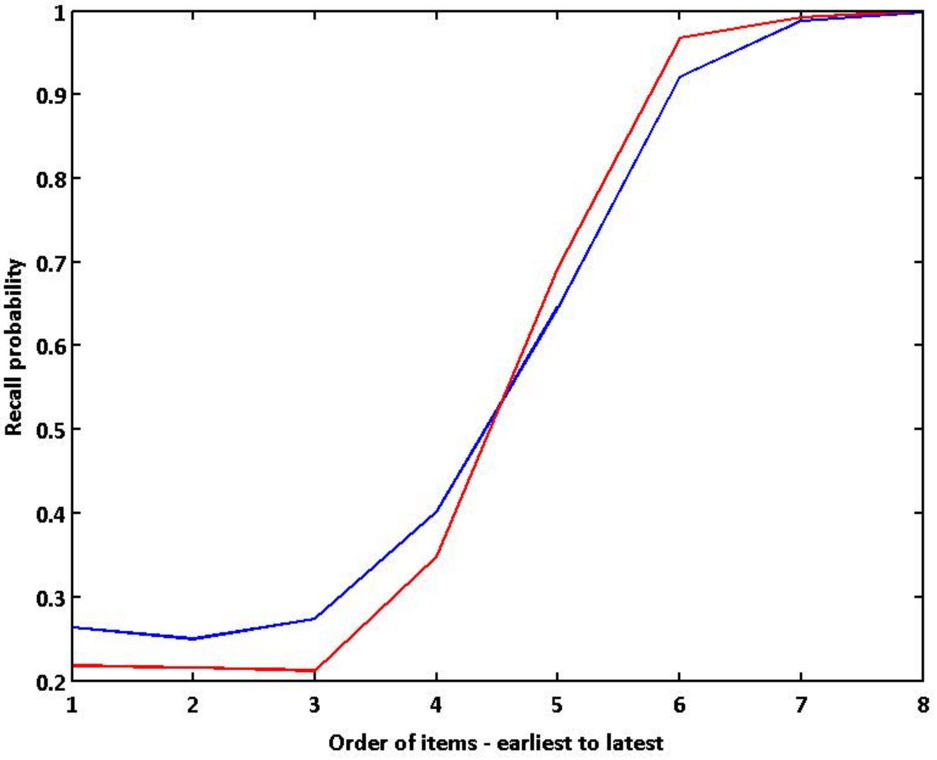


Figure 3.13: Predicted probability of recall for items presented at a fast rate (each input presented to the network for 4 time steps) or slow rate (each input presented to the network for 6 time steps)

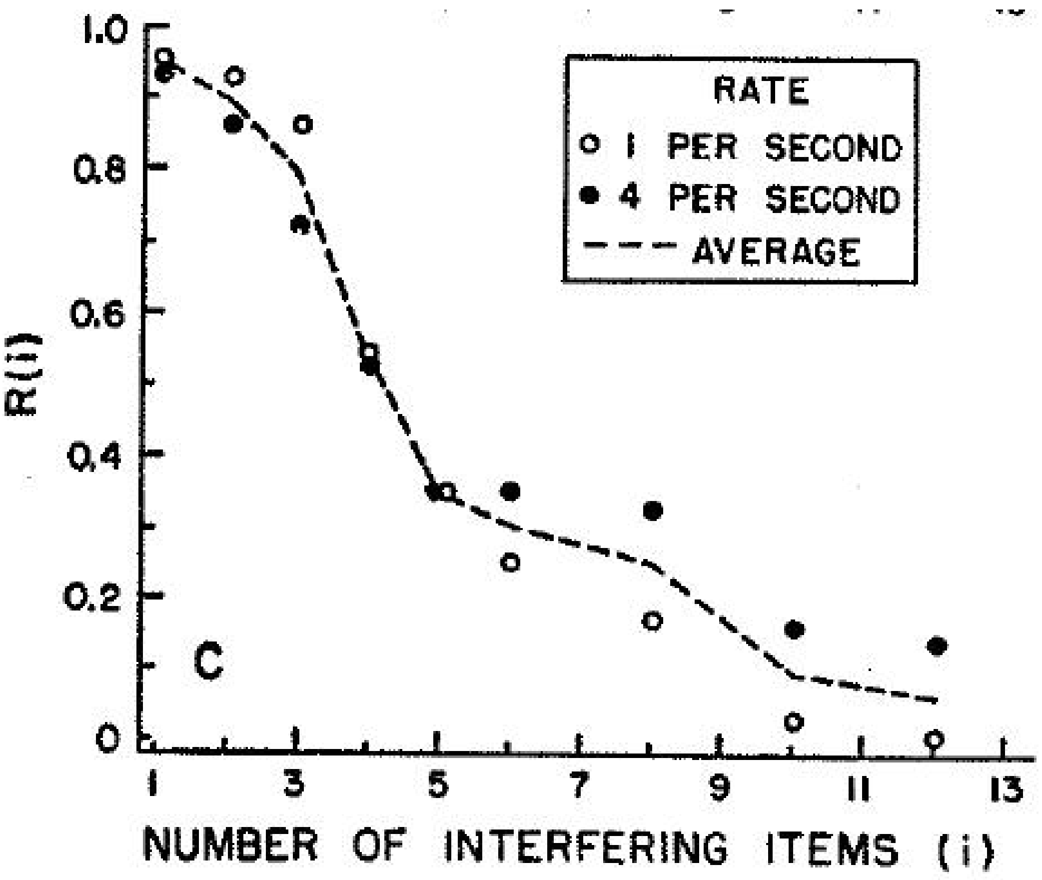


Figure 3.14: Data from Waugh and Norman [1965] showing probability of recall for items presented once or four times per second

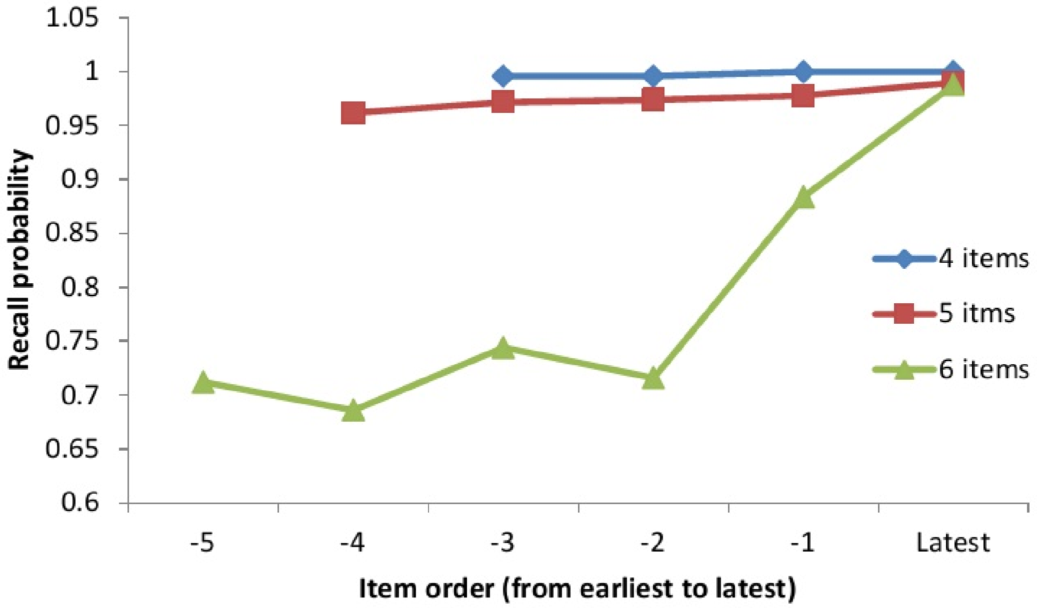


Figure 3.15: Predictions of the model for proportion of recall for different set sizes as a function of serial position. A serial order effect is found when the number of items exceeds 4-5 items.

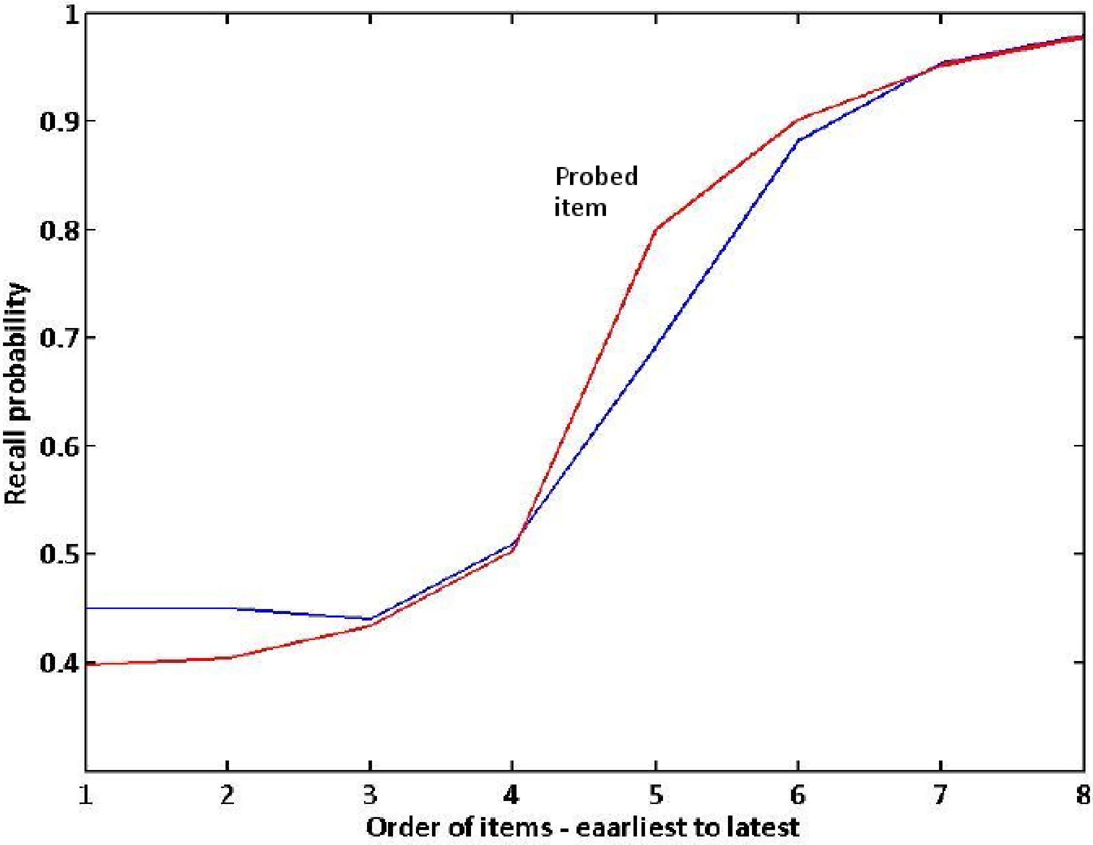


Figure 3.16: Recall curve for spatially cued item. The recall curve shows enhanced recall probability for 5 th item, the one that is spatially cued

items appears more salient than the others (e.g., with an accompanying spatial cue). We achieved this by clamping one unit at 0.66 while others are at 0.33. We can see from the results in Fig. 3.16 (where the 5th item was made more salient) that the cued item has greater recall probability than uncued cases. Thus the model predicts easier recall for spatially probed items.

3.3.2 Summary and other predictions

In the above we saw how a recurrent architecture can dynamically explain capacity limits in working memory. The idea here is that the capacity limits in working memory are flexible as well as the items are encoded with variable precision. This idea of working memory has also been proposed by others [van den Berg et al., 2012], as opposed to slot models for working memory developed by Zhang and Luck [2008].

Interestingly the idea of a flexible variable precision working memory also allows for characterisation of the serial recall curve and recency effect. The model also predicts that increasing saliency of one of the items with spatial cue should boost its recall probability. We tested this hypothesis in Chapter 5.

Furthermore the model makes some critical predictions for change detection tasks. In standard change detection experiment subjects are displayed a set of objects and then change some of them in a subsequent display. The task is to detect whether a feature change has taken place in a probed location or not. To model this I have taken a simplified case where only one object is changed. Following my old model, I have tried to capture this by successive stimulation of certain nodes following a brief interval. In the second stimulation one of the nodes is changed at random. An example time course of the neural activation is shown in Figure 3.17.

Now considering nodes i and j have been switched, we run 500 simulated runs and calculate in what fraction of the runs both nodes i and j are active together above the threshold 0.2. This probability $P(i\&j)$ is the probaility for detecting change. The results for different set sizes and different inhibition parameter (β) values is shown in Figure 3.18.

It seems to indicate peripheral/diffused attention might be better at change detection than focused attention. Moreover it also gives the idea for capacity limit for change detection. In future we intend to explore this phenomenon in more detail.

3.4 Modelling of time's subjective expansion

In Appendix A we have shown that there are constraints on the stability of the network response, i.e., the condition that allows for the activation of two nodes

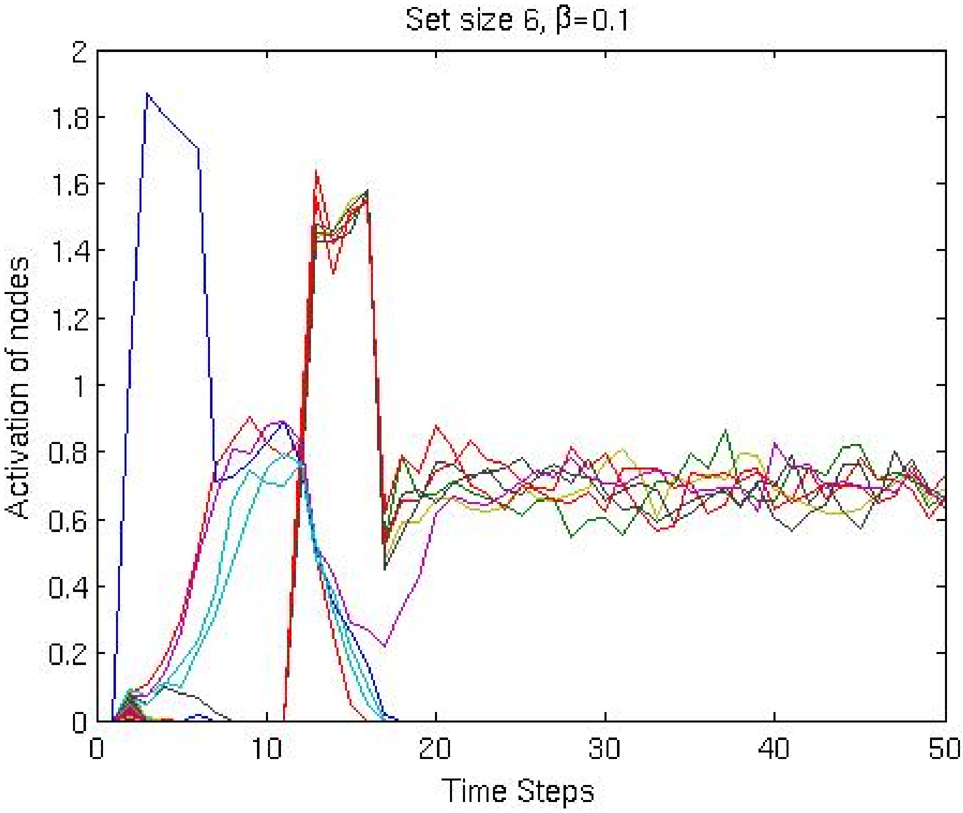


Figure 3.17: Example activation pattern for change detection task. In a network of 70 nodes 6 nodes were stimulated for 5 time steps and then after a gap of 10 time steps 5 of the original nodes along with a randomly chosen new node was stimulated.

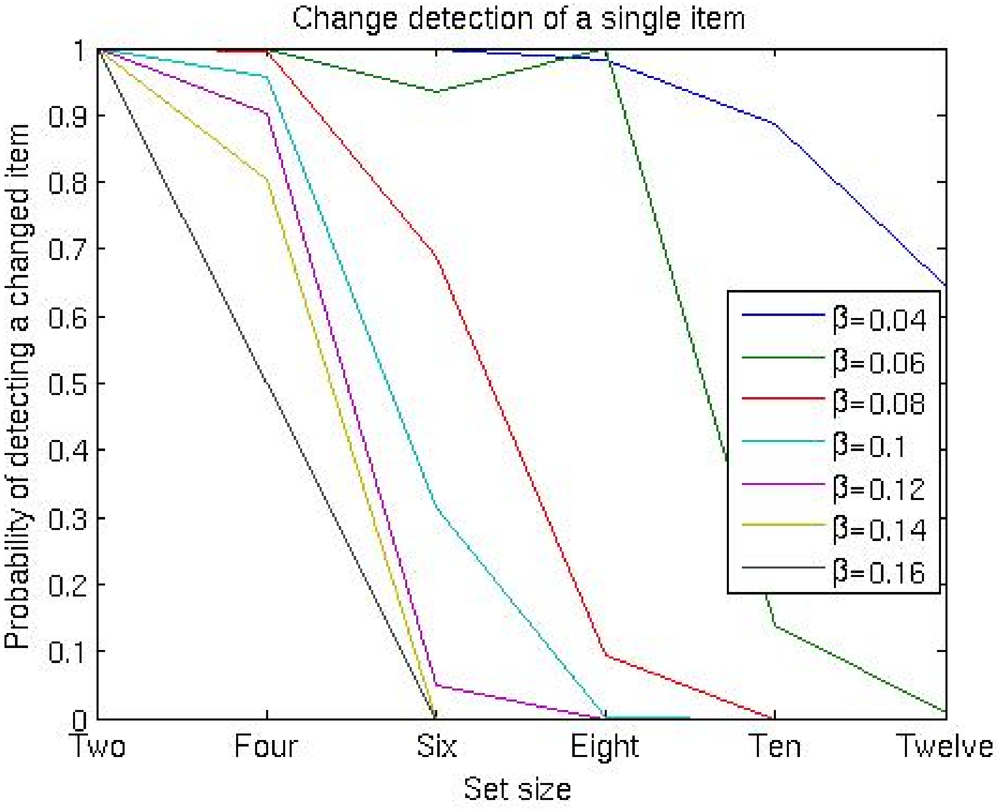


Figure 3.18: Probability of change detection for different set sizes and different β. Probablity that both the new (j) and old (i) node will remain activated simultaneously after simulation ($P(i\&j)$)

to not diverge during the simulation. The constraint was given by

$$\frac{\alpha + \beta}{(\alpha - (n-1)\beta)^2} < 1 \tag{3.9}$$

where n is the number of nodes active after the simulation is run. We also obtained another similar stability condition after consideration of an energy function (see Appendix A), mainly,

$$\frac{\alpha}{(\alpha - (n-1)\beta)^2} < 1 \tag{3.10}$$

However, for a very important behavior of the network, the winner-take-all dynamics requires activation differences between nodoes to be amplified until only one node emerges as winner. Now if $n = 1$, as in standard WTA interaction, Eq. 3.10 transforms into

$$\alpha - 1 > 0 \tag{3.11}$$

This gives the lower limit of α for WTA. The upper limit will be set by Eq. 3.9, i.e. when $n = 2$ reaches stability, i.e.,

$$\alpha^2 - \alpha - 2\beta > 0 \tag{3.12}$$

Thus WTA behavior will be supported by a range of α that satisfies the conditions $\alpha \geq 1$ and $\alpha^2 - \alpha - 2\beta \leq 0$. This is shown in Fig. 3.19. It shows how the α ranges should be calculated for different β avlues. For $\beta = 0.25$ the range is $1 \leq \alpha \leq 1.37$, for $\beta = 0.3$, $1 \leq \alpha \leq 1.42$, for $\beta = 0.35$, $1 \leq \alpha \leq 1.47$.

We also confirmed the parameters with actual simulation. We chose a network of 10 nodes and gave inputs to 2 nodes for each simulation. The input level was clamped at 0.3. The probability of WTA interaction was calculated as the fraction of time out of 1000 simulations, that only one node survives. We varied the α between 0.5 and 1.5. Both stimuli were presented for 255 time steps each and the total duration of simulation was 2500 time steps. Noise was sampled from normal distribution of mean 0 and standard deviation 0.1. The results are shown in Fig. 3.20. The analytic limits obtained in Fig. 3.19 are closely confirmed by the simulation as well.

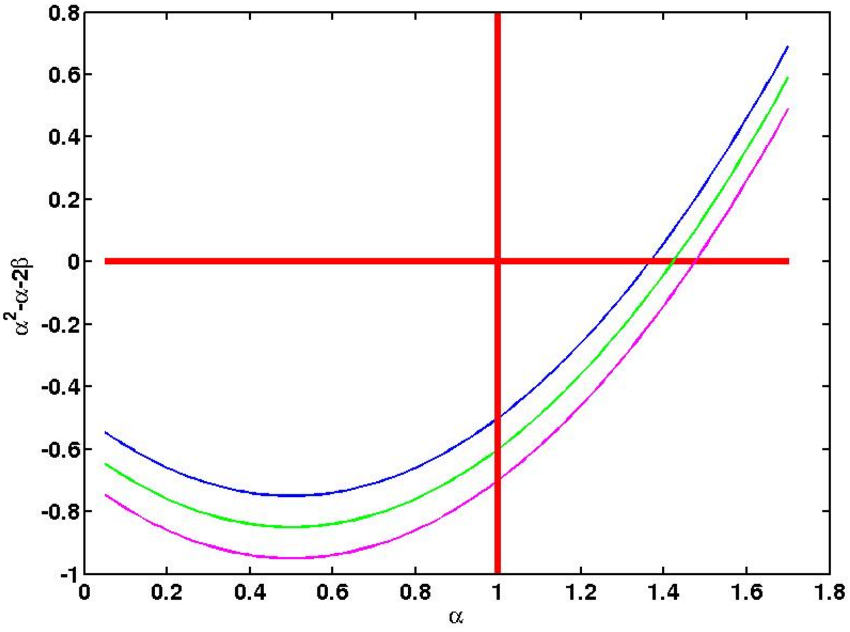


Figure 3.19: A plot of $\alpha^2 - \alpha - 2\beta$ for different β values. Blue line for $\beta = 0.25$, green for $\beta = 0.3$ and magenta for $\beta = 0.35$. In the bottom right quadrant of the plot bound by the lines $\alpha^2 - \alpha - 2\beta = 0$ and $\alpha = 1$ we get the α values desired for WTA interaction, mainly $\alpha \geq 1$ and $\alpha^2 - \alpha - 2\beta \leq 0$. For $\beta = 0.25$ the range is $1 \leq \alpha \leq 1.37$, for $\beta = 0.3$, $1 \leq \alpha \leq 1.42$, for $\beta = 0.35$, $1 \leq \alpha \leq 1.47$.

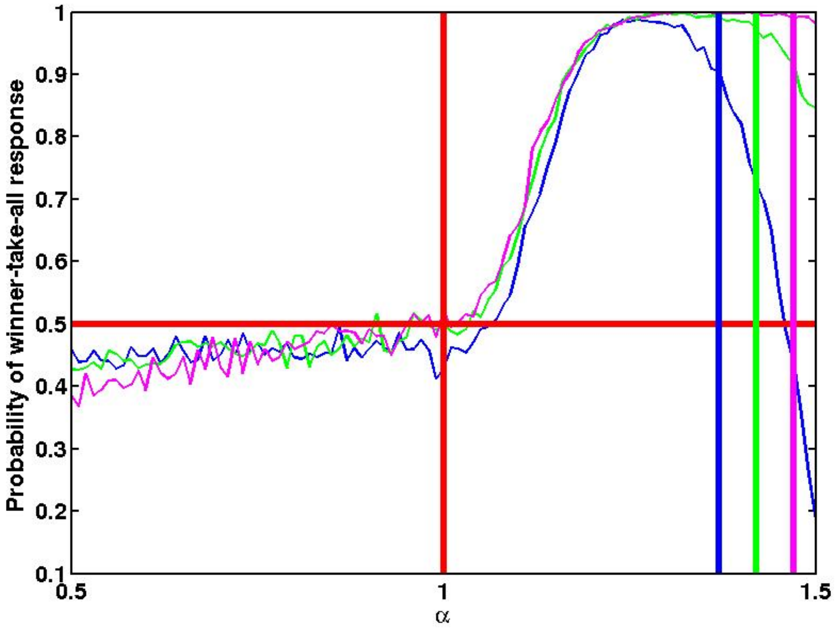


Figure 3.20: Probability of winner take all interaction plotted against α for three different β values 0.25 (blue line), 0.30 (green line) and 0.35 (magenta line). For $\alpha > 1$, the WTA probability increases beyond chance level (0.5). The analytical limits on α obtained in Fig. 3.19 are shown in correspondingly colored lines.

Parameter	Value
N	10
α	1.05
β	0.3
No. of nodes receiving input	2
Total duration of simulation in time steps	2500

Table 3.2: Simulation parameters

3.4.1 Model parameters

In line with Usher and Cohen [1999] and our previous analysis for the winner-take-all kind of interaction, we chose the parameters given in Table 3.2. We worked with a network of 10 nodes, of which only 2 will be excited (e.g., one node was given input of 1020 time steps and the other receives input for varied interval of 30:45:1200 steps). A normal distribution of mean 0 and standard deviation 0.1 is used to sample the *noise* at each time step.

The network shows winner-take-all behavior as shown in Fig. 3.21. Here both the inputs are clamped at 0.3 level (In Usher and Cohen [1999] levels 0.1 were used).

3.4.2 Oddball judgement

In a standard oddball task for temporal judgements involve showing a participant a series of frequent standard stimuli followed by infrequent oddball stimulus. A few standard frequent stimuli might also be shown after oddball presentation. The participant makes a judgement whether the oddball stimulus was longer in duration than the standard or not. Our model assumes that a normalized input is given to two nodes of varying durations simultaneously as the duration judgements are post-event judgements in temporal oddball, it is not such an unreasonable assumption.

Now we ran 100 simulations for each duration pairs (30 vs 660 ms steps, 75 vs. 600 ms steps,... 1200 vs. 660 ms steps) for the two nodes and calculated the probability of response that one duration will be judged longer than the other by counting the fraction of times the node with the variable input is the winner of

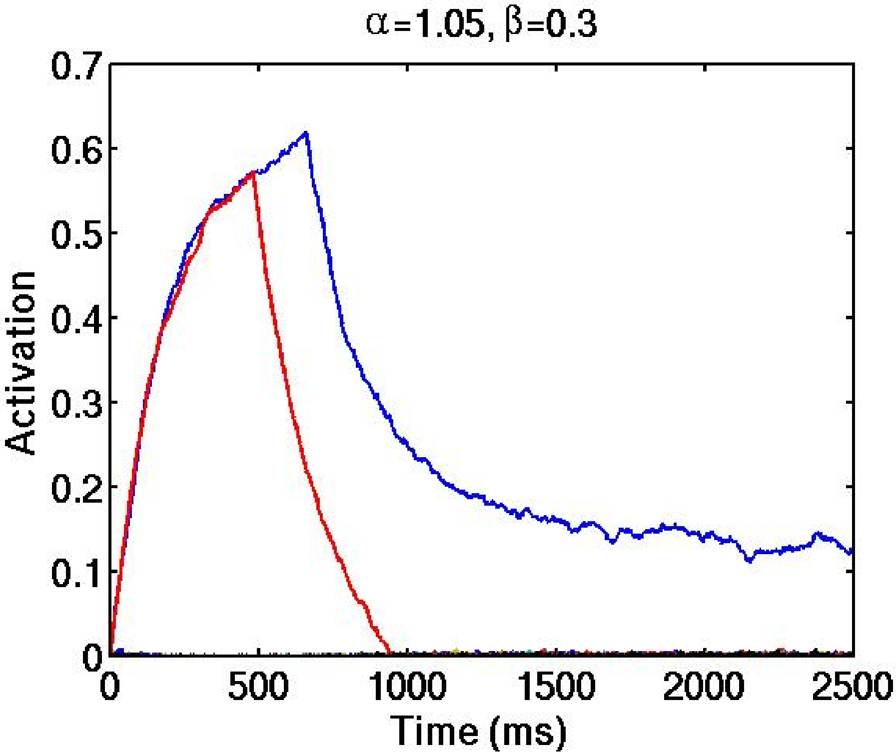


Figure 3.21: Winner-take-all behavior shown at $\alpha = 1.05$, $\beta = 0.3$ for two nodes receiving inputs for 480 ms (red line) and for 660 ms (blue line).

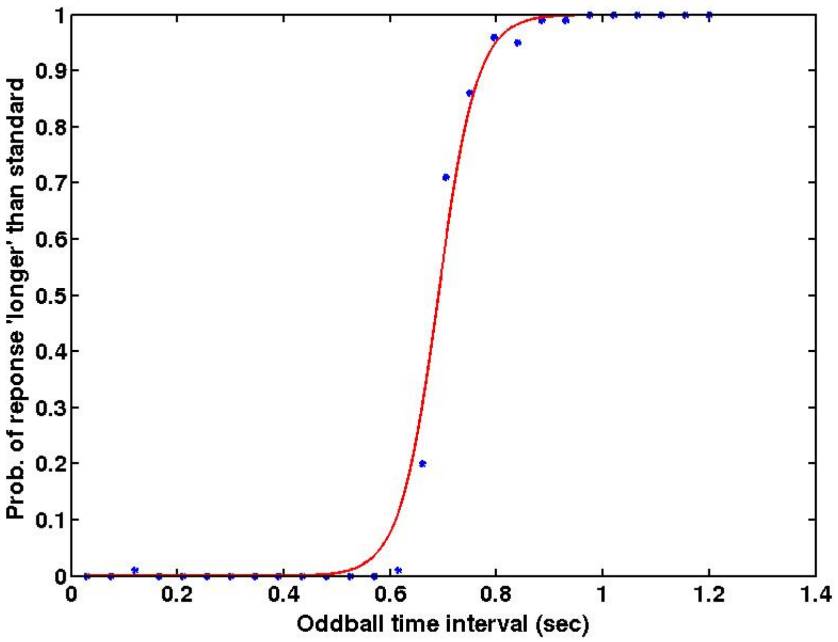


Figure 3.22: The figure plots the probability of response longer than standard (here 5 time steps) compared to the variable time intervals shown on x-axis.

WTA interaction. The result is shown in Fig. 3.22. Here the point of subjective equality is around 680 ms after a logistic fit. So there is not much time's subjective expansion.

However, in reality it is rare that both the units to be excited with the same level of input. It is more likely that the salience of the inputs differ. So we checked for three different conditions - 1) variable interval input level (0.2) less than standard input level (0.3), 2) variable interval input level (0.3) same as standard input level (0.3), 3) variable input level (0.3) more than standard input level (0.2). As we can see that the point of subjective equality can shift in either direction given the salience of the inputs.

The standard stimuli is more frequent and thus prone to habituation and should receive lesser level input than the oddball stimuli. So the temporal oddball should correspond to the case 3 considered above. This seems to be an important finding. This shows that the dynamic OCOS paradigm can be adapted for time judgements. It is also observed the β modulates the slope of the logistic curve (it

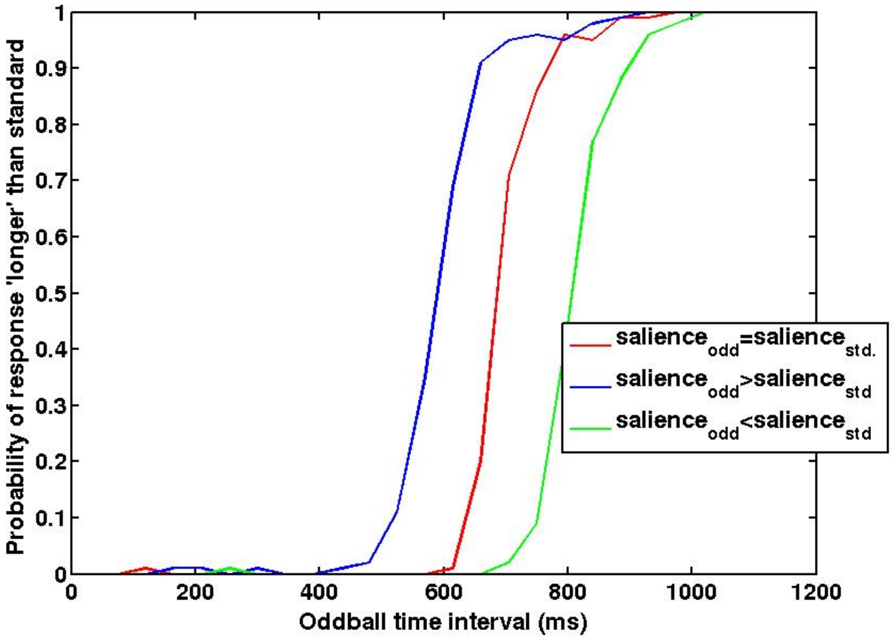


Figure 3.23: The probability values for different variable intervals being judged larger than standard interval for three different conditions 1) variable interval input level (0.2) less than standard input level (0.3), 2) variable interval input level (0.3) same as standard input level (0.3), 3) variable input level (0.3) more than standard input level (0.2).

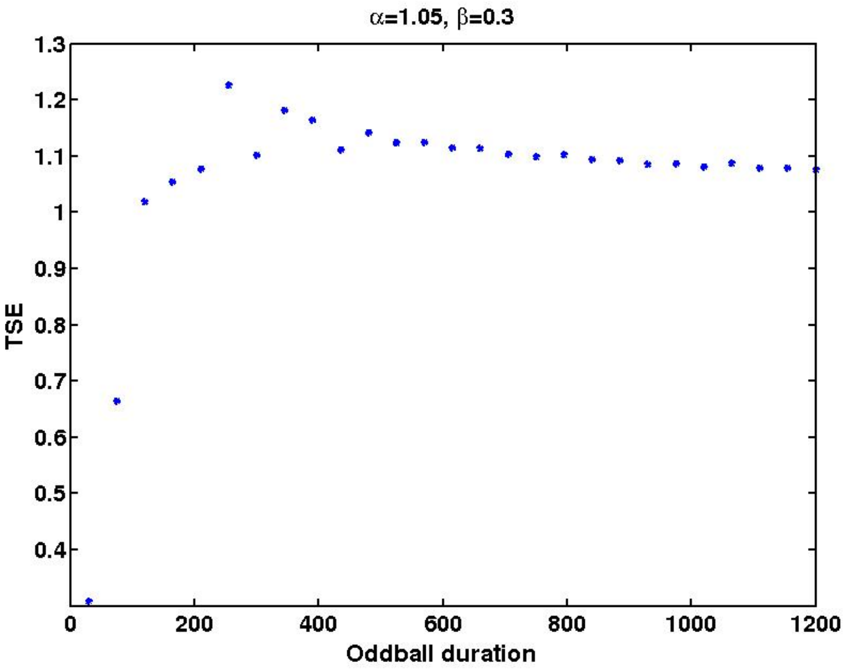


Figure 3.24: TSE values for different standard durations.

becomes steeper at higher β).

3.4.3 Modelling TSE and predictions

Now we calculated the values for the probability that oddball duration would be perceived as longer for each of the standard durations in the set (30:45:1200) (to be given input of 0.2). The oddball durations are also (30:45:1200) (to be given input 0.3). We calculated the points of subjective equality (PSE) through sigmoid fits and time's subjective expansion by dividing the standard durations by corresponding PSE's. The result is given in Fig. 3.24.

Compare this with experimental data given by Tse et al. [2004] in Fig. 3.25. Interestingly, in the model also the TSE value of 1 is crossed around 115 ms mark like the experimental data, where it is around 120 ms. Overall, there seems to be very close link between the data and the model. Although, quantitatively TSE experimental values appear more, but they can be modified further in the model through increasing the salience (or here the clamping level) of the oddball nodes.

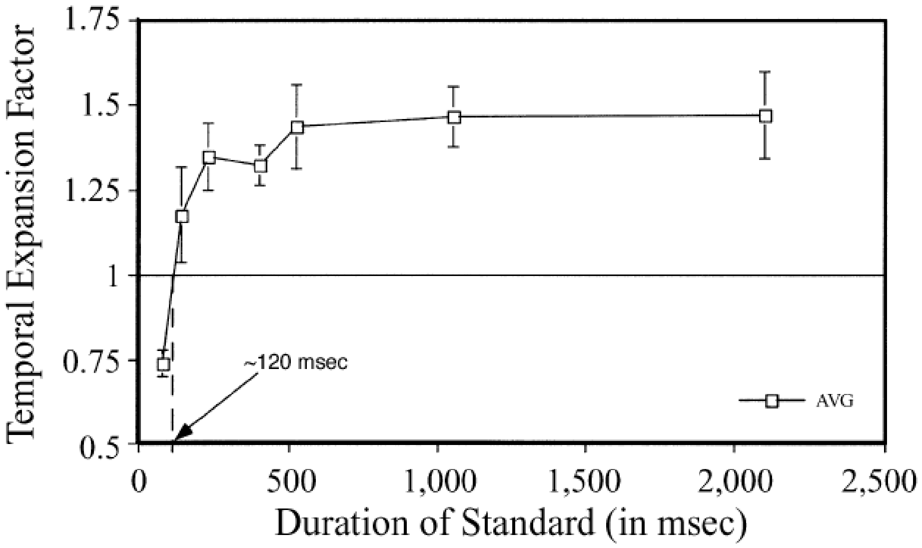


Figure 3.25: Experimental data by Tse et al. [2004]

Assuming that more salient oddball will provide greater level of input than less salient one. We tested different oddball stimulus levels (0.3, 0.4, and 0.5) to derive the TSE curve. The result shown in Fig. 3.26 shows that if we interpret the oddball stimulus levels to correspond to the amount of habituation, then we can predict the effect of habituation on time perception in an interesting manner. The results seem to suggest that the effect of habituation is largest at 200-400 ms range for standard stimuli, whereas not so much for higher or lower duration standard stimuli.

3.4.4 Discussion

In the context of perceptual decision making, a winner-take-all mechanism is a prime candidate for a biologically plausible neuro-computational approach [Wong and Wang, 2006]. In case of temporal oddball paradigm, we observed that the experiment follows a two-alternative forced choice (2AFC) task. The 2AFC task led to the idea that time's subjective expansion factor observed in temporal oddball paradigm might arise from a flexible decision boundary arising from a neural WTA mechanism. It turned out indeed a WTA model was able to explain most salient effects in experiments conducted by Tse et al. [2004]. The model also was

---

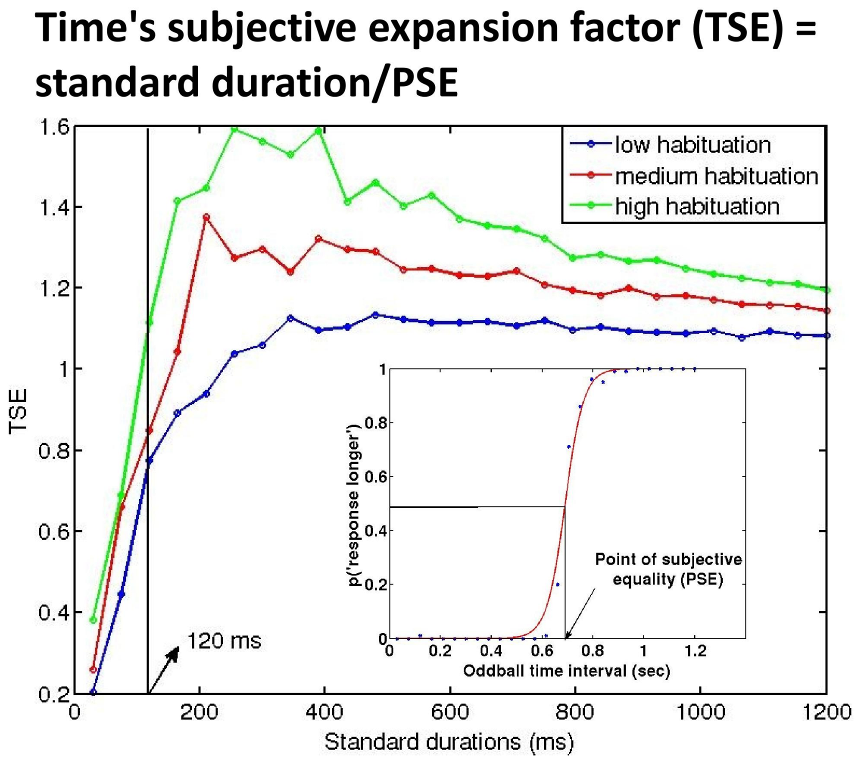


Figure 3.26: TSE values for different standard durations at different levels of oddball input. Blue line has oddball input at 0.3, red line has oddball input at 0.4, green line has oddball input at 0.5. The different levels were interpreted as arising from different levels of habituation, i.e., if the oddball appears in a stream of 6 standard stimuli, its salience would be less than that appearing in a stream of 11 stimuli, and thus the former will have lower stimulus level.

effective in coming up with new testable predictions which we verified in Chapter 6.

Chapter 4

Empirical investigation of visual sense of numbers

> Confined in the ship, from which
> it is impossible to escape, the
> madman is confined to the
> thousand branches of the river,
> the thousand paths of the sea, to
> this great uncertainty external to
> everything. He is a prisoner in the
> midst of the most free, the most
> open of roads: chained solidly to
> an infinite crossroads.
>
> *Madness and Civilization*
> MICHELE FOUCAULT

4.1 Introduction

Numerical cognition is grounded in the perceptual capacity to individuate or estimate the number of objects in a set [Halberda and Feigenson, 2008; Piazza et al., 2011, 2007]. Enumeration of multiple visual objects gives rise to two behaviorally distinct phenomena - subitizing and estimation - depending on the numerosity of the objects presented on a display [Kaufman et al., 1949]. A rapid, precise and confident enumeration strategy called subitizing works for smaller numerosity (generally less than 5), while a less accurate and slower enumeration strategy of estimation seems to operate at higher ranges of numerosity. The distinction between the ranges has been shown through psychophysical measures like Weber fractions [Burr et al., 2010] and precision measures [Whalen et al., 1999].

It has been of considerable interest in the last few years to understand the possible mechanisms for enumeration [Trick and Pylyshyn, 1994]. Some have proposed that the subitizing range might reflect capacity limits appearing in the visuo-spatial object individuation domain [Melcher and Piazza, 2011; Piazza et al., 2011], while estimation reflects coarse ensemble statistics measures [Ariely, 2001; Burr and Ross, 2008b; Feigenson et al., 2004]. Alternatively, it has been suggested that subitizing reflects a single numerical estimation mechanism operating at different numerical resolutions [Gallistel and Gelman, 1992; Revkin et al., 2008]. However, recent studies have shown that precision for the two ranges do not scale evenly [Revkin et al., 2008]. This has led some to suggest the possibility of separate mechanisms for the two enumeration strategies [Burr et al., 2010].

A recent study by [Burr et al., 2010] suggests that subitizing and estimation share common pre-attentive mechanisms, but subitizing makes use of additional attentional resources and thus are subject to capacity limits. One hypothesis is that subitizing reflects the activity of attentional priority (saliency) maps that represent at most a few salient objects [Knops et al., 2014; Melcher and Piazza, 2011; Piazza et al., 2011; Sengupta et al., 2014b]. By definition, a saliency map does not represent all of the items in a large display, rather only the most relevant (and attended) objects. Thus, the link between subitizing and attention could be mediated by such a mechanism. There are several advantages to a priority/saliency map model explanation for capacity limits (for reviews, see Franconeri et al. [2013]; Melcher and Piazza [2011]; Sengupta et al. [2014b]), including the fact that there is extensive neurophysiological and neuroimaging data supporting the existence of such maps (for reviews, see: Gottlieb and Goldberg [1999]; Knops et al. [2014]) and that capacity limits emerge naturally out of such maps [Franconeri et al., 2013; Melcher and Piazza, 2011].

A number of studies have indicated posterior parietal cortex as the locus of object individuation, working memory and enumeration (for a review, see: Piazza et al. [2007]). It has been suggested that numerosity detectors in the lateral intraparietal (LIP) region might be possible candidates for our perception of numbers [Piazza et al., 2004; Roggeman et al., 2010; Roitman et al., 2007]. These detectors might develop through self-organized or reinforcement learning during development [Dehaene and Changeux, 1993; Grossberg and Repin, 2003; Stoianov and Zorzi, 2012]. It is interesting to note that the LIP area is shared between visual spatial attention and visual short-term memory networks [Awh and Jonides, 2001]. Depending on the task (working memory task or enumeration) different activation patterns are seen in posterior parietal cortex using fMRI [Kawasaki et al., 2008; Knops et al., 2014; Todd and Marois, 2004, 2005].

In order to investigate the possible mechanisms underlying rapid enumeration, we previously modeled a recurrent on-center off-surround (OCOS) network that receives a normalized input based on locations occupied in the visual field and outputs a steady state mean activation [Sengupta et al., 2014a,b]. The model successfully captured a wide range of behavioral phenomena observed in enumeration research, including the distance effect (precision for numerosity discrimination for a given numerical distance is better for smaller numerosities than higher), Fechner's law (discrimination performance degrades for higher numerosities), as well as human numerosity comparison data collected by Piazza et al. [2004]. In our model, subitization and estimation ranges emerge naturally with parametric variation of the recurrent inhibition strength of the network (see also Roggeman et al. [2010] for a similar approach). Moreover, the model also predicted that reaction times for enumeration would show hysteresis if the subjects were forced to change enumeration strategy by making them enumerate large numbers following several trials of small numerosity and vice versa. This novel prediction (explained in more detail in the next section) hinges on the idea that enumeration and estimation might be carried out by the same flexible network acting under a different level of inhibition between nodes. In the current chapter we describe the model and its predictions, then present the experimental evidence for the model's predictions.

4.2 Model

In our earlier Chapter 3 we had derived an energy function in order to explore the stability of this kind of network. Reaction times (RT) were derived based on the assumption that RT correlates with the highest possible allowed fluctuation in energy. The expression for RT for enumeration tasks is given as

$$RT \sim \sum_i \int \left(1 - \alpha \left(\frac{F(x_i)}{x_i}\right)^2\right) \dot{x}_i^2 dt \tag{4.1}$$

The quantity on the right hand side (see model description above for a definition of the terms) is proportional to the actual RT, but not equal to it and thus we refer to it as RT correlate or RT_{model}. Considering that high inhibition is better for enumerating smaller numerosity and low inhibition is better for accounting for large numerosity, we asked the question: what happens to RT when you enumerate smaller numerosity at low inhibition or you enumerate large numerosity at higher inhibition? In order to test the model predictions regarding such a switching cost we took an average of RT correlates obtained through simulation for higher inhibition ranges ($\beta = 0.06 - 0.08$) and lower inhibition ranges ($\beta = 0.04 - 0.06$). As shown in Fig 4.1, the model predicts a particular pattern of hysteresis. If one enumerates lower numerosity with low inhibition (a condition that might be manipulated experimentally by having a low numerosity enumeration task follow a series of high numerosity enumeration tasks) then the RTs should increase compared to standard RT for that range. In contrast, there is almost no RT cost for higher numerosity being enumerated at higher inhibition. Thus we decided to test the prediction of the model with an enumeration task where a variable number of small numerosity enumeration (1-6 items) trials would follow a variable number of large numerosity estimation (24-34 items) trials and vice versa.

Regarding errors, the model would predict greater underestimation at higher inhibition. This leads to the prediction that a high numerosity trial following low numerosity trials would suffer greater underestimation compared to a high numerosity trial following high numerosity trial.

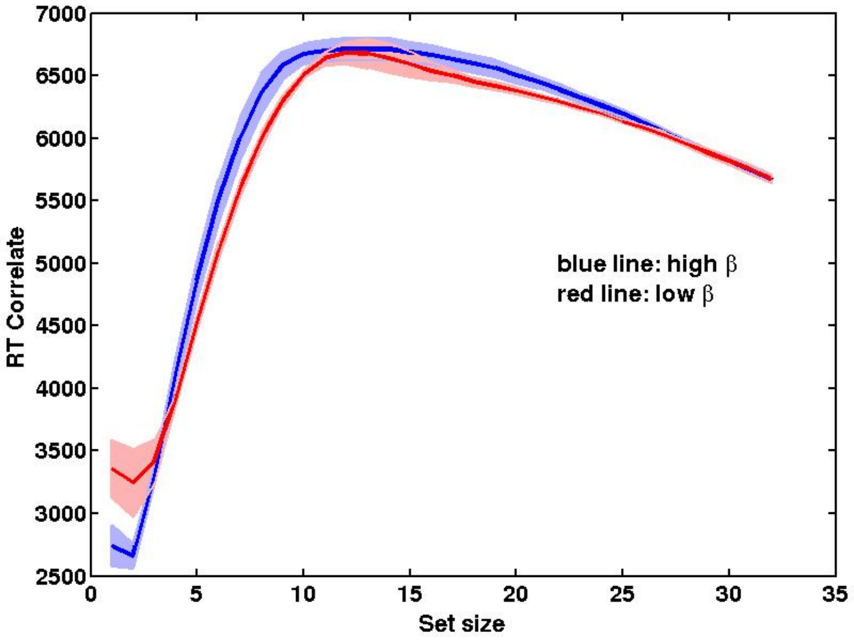


Figure 4.1: This figure shows the average RT correlate or RT_{model} values for two ranges of β: high-0.06-0.08 and low-0.04-0.06. The higher numerosity ranges show similar patterns of RT for the two ranges of inhibition, while the low numerosity range differs greatly in predicted RT between the two beta ranges. The error bands represent one standard deviation of the mean RT_{model} values at different set sizes.

4.3 Method

4.3.1 Participants

A total of 21 adult volunteers participated in this study (age range: 24-32). Of this group, 12 were tested at the Center for Mind/Brain Sciences, University of Trento, Italy and 9 additional participants were tested at the Center for Neural and Cognitive Sciences, University of Hyderabad, India. The participants included 10 male and 11 female participants. Informed consent was obtained from all participants in accordance with the local ethical committee guidelines.

4.3.2 Stimuli

Stimuli were black dots (20 x 20 pixels) presented on a gray screen background within a virtual rectangle (917 x 683 pixels) on a (1280 x 1024 pixels) monitor. There were two sets of numerosities: small (1,2,3,4,5,6) and large (24,26,28,30,32,34). The exact locations of the dots were randomized within the virtual rectangle on each trial.

4.3.3 Design

Participants were seated in a dimly lit room approximately 57 cm from a 45.2 cm x 36.1 cm computer monitor. Following a short practice session (20 trials), participants began the experiment by pressing a keyboard button. After a brief delay (1 second), a set of dots was presented on the screen for 100 ms followed by a flash. The flash was presented to reduce the visual persistence of the dots. After each trial, there was a 2 second interval prior to starting the next trial.

Participants were instructed to quickly and accurately verbally report the exact numerosity of dots presented on a computer screen. The schematics of the experiment are given in Fig. 4.2.

There were two types of trials: small numerosity or large numerosity trials. Within these two types of trials, the exact numerosity varied across trials, as described above. The experiment was designed so that there were sequences of trials in which the same type (small or large numerosity) was repeated. Each trial

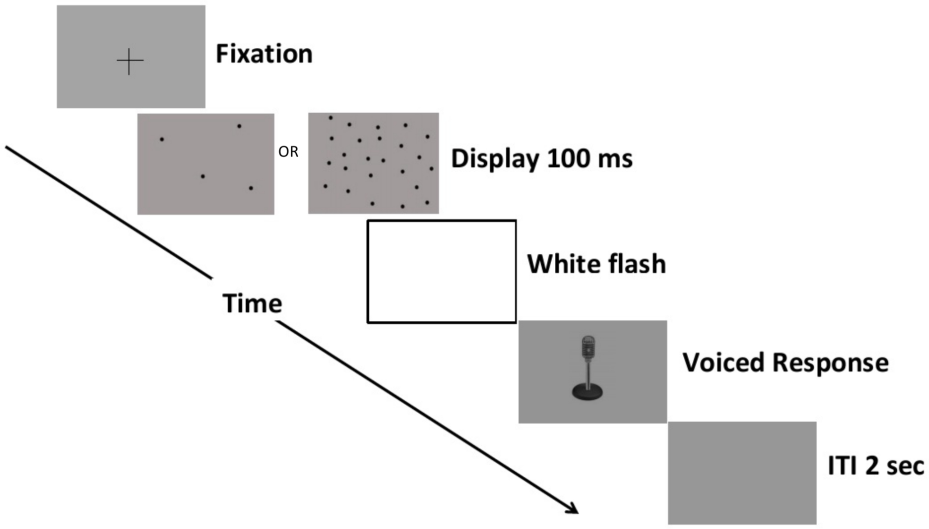


Figure 4.2: Illustration of the structure of a trial. A display of black dots was presented for 100 ms and participants verbally reported the number of items as quickly and accurately as possible. The white flash was used as a mask to reduce the visual persistence of the dots in order to encourage participants to respond quickly rather than slowly counting each item individually from the after-image.

was associated with a Markov-like counter with probabilities of - 0.10, 0.15, 0.22, 0.32, 0.46, 0.68, or 1.00. For instance, if the same kind of trial was shown for 5 times in a row, then the switching probability (probability that the next trial would be from the other set of numerosities) would become 0.46. As a result, the number of trials in a sequence (of either large or small numerosity) ranged from 1 to 7 (average of 4). The total number of trials was 200 per subject and the total duration of the experiment was around 10 minutes. Reaction time (RT is the voice onset time starting from when the flash terminates) was collected for each trial and the actual enumeration response was transcribed manually from the voiced response for each trial. Response error was computed by taking the difference between the reported and actual numerosities.

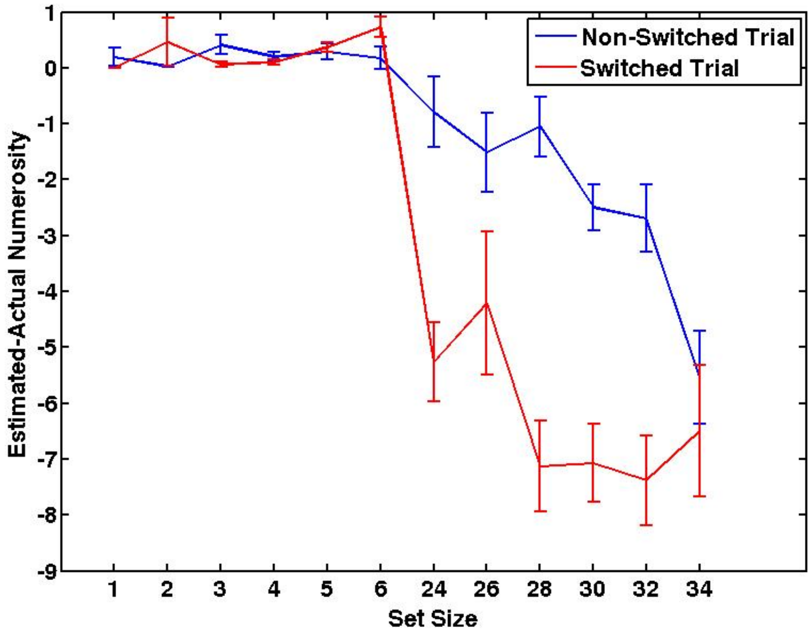


Figure 4.3: Enumeration error between the reported numerosity by the participant (estimate) and the actual presented numerosity for different set sizes. Negative numbers indicate that participants underestimated the numerosity. Error bars indicate standard error of the mean.

4.4 Results

4.4.1 Accuracy and underestimation

Consistent with previous findings in the literature, participants were highly accurate (83%) and fast (mean RT= 717 ms, SE = 19 ms) for items presented within the subitizing range and considerably slower (mean RT= 1156 ms, SE= 14 ms) and less accurate for items presented within the estimation range. As predicted by the model, there was a strong tendency, within the large number range, to underestimate numerosity on switch trials compared to non-switch trials (Fig. 4.3). The specificity of this effect, occurring only for the switch from small to big trials, is suggestive of a hysteresis effect from the small trials.

Further evidence for a hysteresis from small numerosity trials comes from the duration of the switch cost. As shown in Fig. 4.4, participants were increasingly

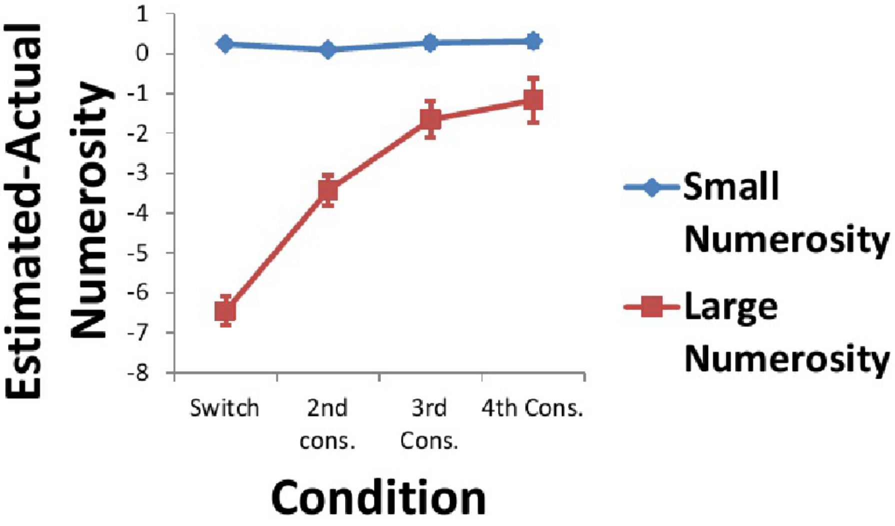


Figure 4.4: The time course of underestimation effects of the switch, starting on first trial after the switch from small to large (red squares) or large to small (blue diamonds). Amount of underestimation is shown also for the 2nd, 3rd and 4th consecutive numerosity trial of the same kind that continued after the switch.

more accurate with each trial following the switch from small to large (SL) numerosity. In other words, the initial switch cost for SL trials was reduced across trials and settled into a stable pattern of accuracy. In contrast, enumeration of small set sizes remained stable.

4.4.2 Reaction time and switch cost

As shown in Fig. 4.5, the reaction times for enumeration show standard features found by Kaufman et al. [1949], in that reaction time remained relatively fast for 1-4 items, then increased in a somewhat linear fashion (here, up to around 6 items). Interestingly, the average RT for 6 and 24 items was similar, suggesting that RT remained relatively constant as a function of set size after 6 items.

As predicted, there was a significant RT cost for switch trials when going from large to small numerosities. The main effects of set size, $F(1,20) = 276.34$, $p < 0.001$, and switch condition, $F(1,20) = 16.75$, $p < 0.001$ were confirmed in a 2-factor repeated measures ANOVA. There was also a significant interaction

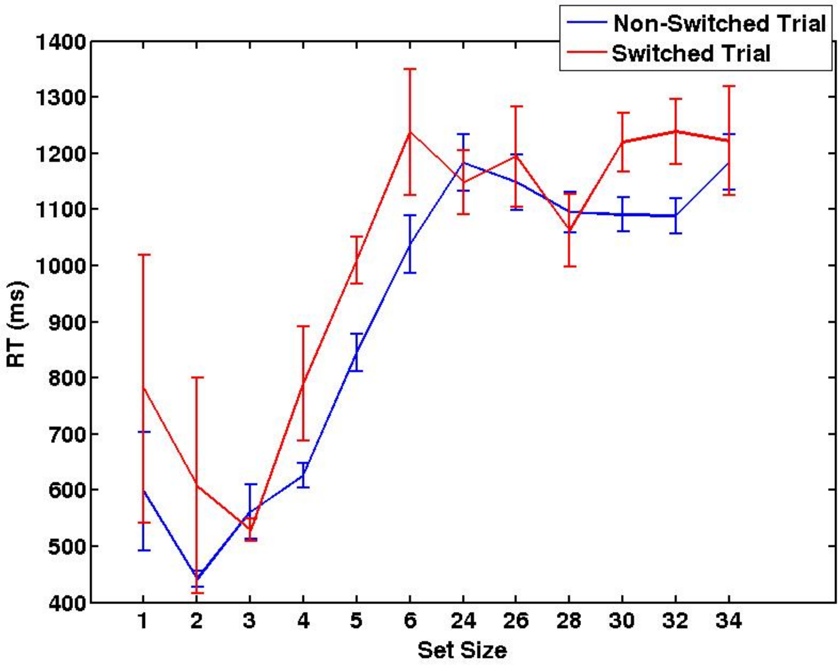


Figure 4.5: Average reaction times for each numerosity plotted with error bars for standard/non-switched and switched trials. Non-switch trials refer to when a small numerosity trial followed a small numerosity trial or large numerosity trial followed a large numerosity trial (see Methods).

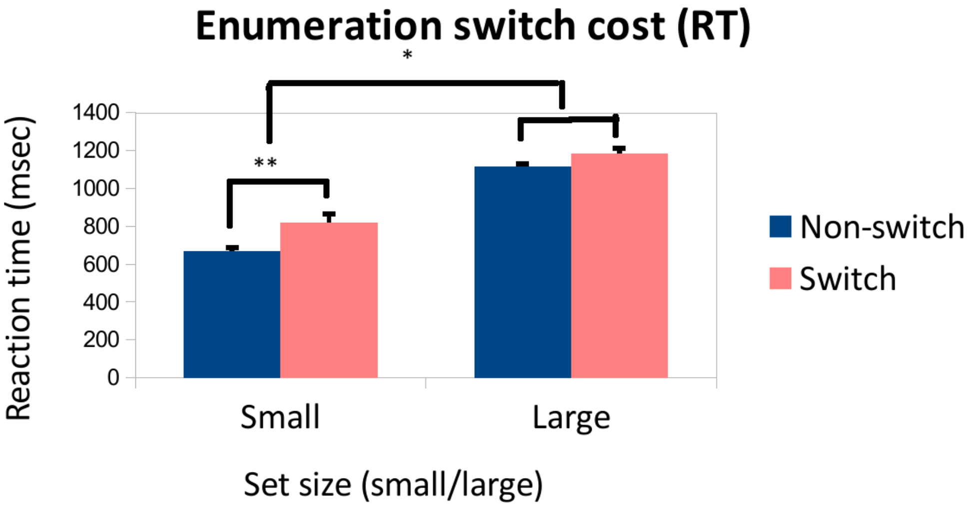


Figure 4.6: Enumeration switch cost (RT) for reaction time for large and small numerosity trials.

between these two factors, $F(1, 20) = 4.06$, $p < 0.05$. The results show a strong main effect for switch condition, $F(1, 20) = 276.34$, $p < 0.001$ as well as a significant interaction between set size and switch condition, $F(1, 20) = 4.06$, $p < 0.05$. The interaction appeared to be driven by the strong effect of switch only on small numerosity trials (Fig. 4.6). Given the interaction term we performed a post-hoc comparisons of means (with Bonferroni correction) between switch categories within each set size class. We found an effect size of ~ 150 ms, $F(1, 20) = 13.48$, $p < 0.001$ for large to small switch in terms of RT, whereas the cost for a small to large switch was not significant, $F(1, 20) = 3.44$,$p < 0.1$. The results are shown in Fig. 4.6.

Also of interest was the duration of the switch cost. We separated the trials for small and large numerosity according to their distance from a corresponding switched trial, e.g., a small numerosity trial following a switched small numerosity trial (the last trial in the series L-S-S) would be called the 2nd consecutive small numerosity trial, the second large numerosity trial following a switched large numerosity trial (the last trial in the series S-L-L-L) will be the third consecutive large numerosity trial etc. As indicated in Figure 4.7, reaction times decreased when small numerosities were repeated multiple times ($F(1, 20) = 4.87$, $p < 0.05$ between switched S trial and the third consecutive S or Switch+4 trial), whereas

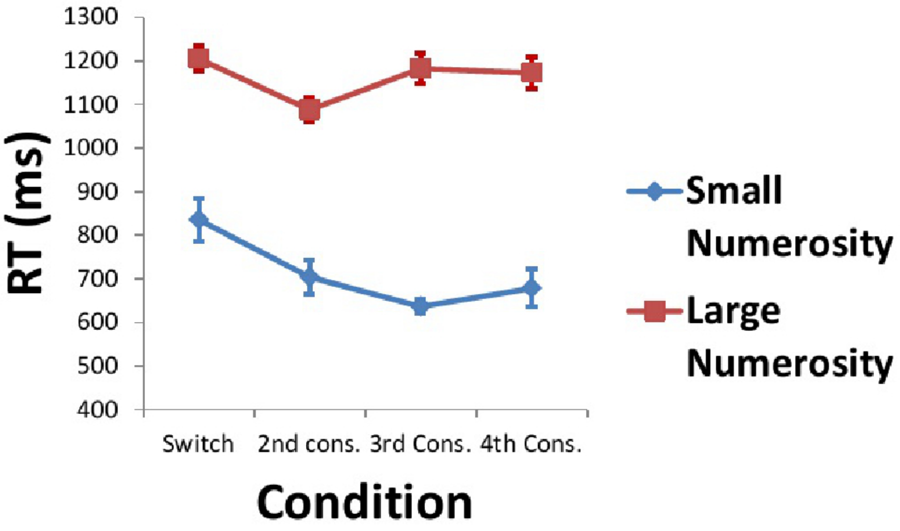


Figure 4.7: Reaction time as a function of number of subsequent trials following the switch, for small and large numerosity trials.

no such difference was found between switch trials and switch+4 trials with large numerosities, $F(1, 20) = 0.36$, $p < 1$. The pattern of results after the switch is consistent with the network settling back into a stable state of low energy.

In order to check the model predictions against the empirical data, we considered two main predictions with regard to RTs. Firstly, mean RTs for numerosity judgments in the non-switch condition for set sizes 1-6 is predicted to be correlated with the mean RT estimate corresponding to a high inhibition value for the β in the model. Secondly, mean RTs for numerosity estimation in the non-switch condition for large set sizes would be correlated with the RT estimates from the model operating with a low inhibition parameter value β. Whereas in the switch condition, the RT values are expected to be correlated with high inhibition RT estimates for large numerosities and with low inhibition RT estimates for small numerosities. We first estimated a linear fit between the empirical values for the average RTs in non-switch trials against the corresponding values obtained from the model.

$$RT_{non-switch}(empirical) = a * RT_{non-switch}(model) + b \tag{4.2}$$

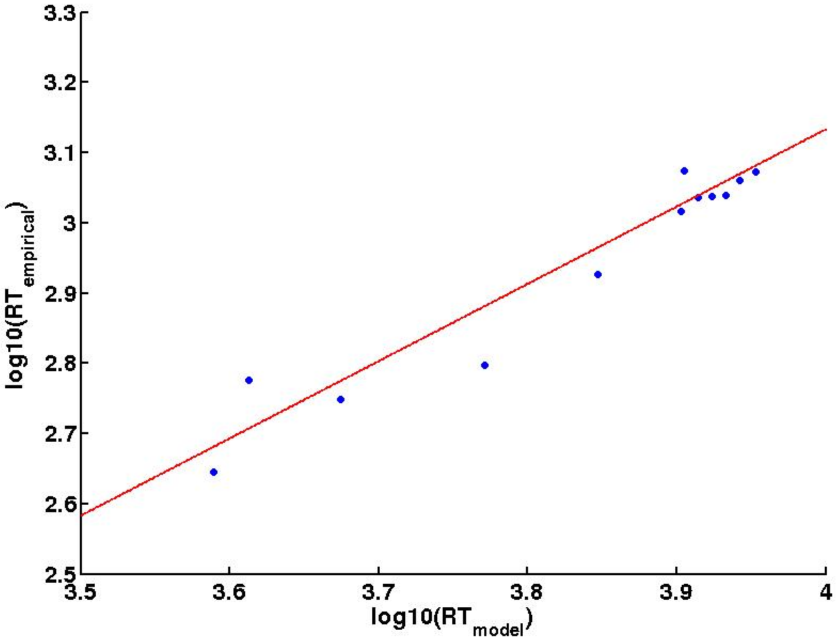


Figure 4.8: The fit between model predicted RT values and empirically obtained RT values for non-switch trials ($RT_{non-switch}(empirical) = a * RT_{non-switch}(model) + b$). Adjusted r^2 is 0.93 at 95% confidence interval.

There was a very good fit between the model and the data (adjusted $r^2 = 0.93$ at 95% confidence interval; see Fig. 4.8). The coefficient values obtained were $a = 1.097$ and $b = -1.267$.

We used the same coefficients to check the goodness of correlation between $RT_{switch}(empirical)$ and $RT_{switch}(model)$. Using the same coefficients a and b, we found a correlation coefficient of 0.9 at 95% confidence interval for the fit between model-derived and empirical RT on switch trials (see Fig. 4.9).

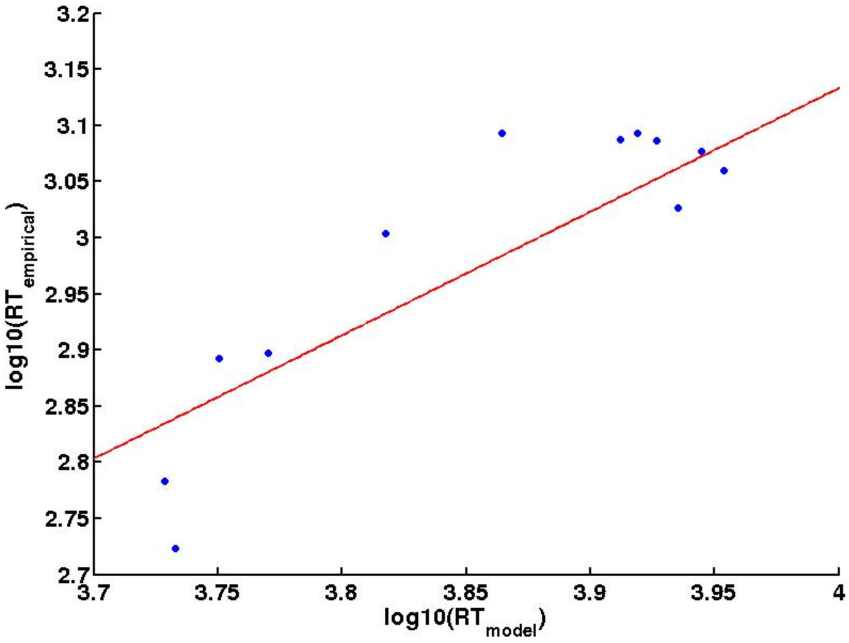


Figure 4.9: A fit between model predicted RT values and empirically obtained RT values for switch trials using the same coefficients a and b obtained from the fit for non-switch trials.

Chapter 5

Empirical Investigation of the Effect of spatial cueing on serial recall

5.1 Introduction

In this chapter we explore whether spatial cueing can affect serial recall. The motivation for this comes from the modelling results demonstarted in Chapter 3, especially Fig. 3.13 & 3.16, which shows that the model explains not only the *primacy effect* visible in the concave pattern for the serial recall curve, but also that a boost in spatial saliency would facilitate serial recall.

We know that spatial cueing guides selective attention to certain locations, facilitating effective encoding of the items in those said locations. Thus, we can assert computationally and theoretically that, spatial cueing should improve serial recall of items presented in different locations, measurable in reduced RTs and improved accuracies if all or some of the items were probed with valid cues. In contrast in the case where spatial cueing does not carry any relevant information towards the location of the items presented (invalid cues), recall facilitation may not be observed. We also wanted to probe deeper into the stimulus feature itself by looking at colored squares or grayscale shapes. We know from previous studies, that color and shapes are processed differently [Corbetta et al., 1991].

As the short term memory is a limited capacity system [Miller, 1956; Todd and Marois, 2004], we also varied the set size of items. We hypothesized that the facilitation offered by spatial cues would be more significant for higher set sizes. We further expected a dissociation between color and shape features in the way their recall is affected by spatial cues.

5.2 Materials and Method

In this chapter we have tried to explore the hypothesis that spatial cues interact with serial order visual short term memory through 3 interconnected experiments. In the first experiment we only spatial cue for only one of the items presented in serial recall task and check the recall of cued items vs. uncued items. In the second experiment we tried to isolate the effect of cue itself from the spatial factor. Thus we contrasted a condition where all items were cued vs. the condition where no items were cued, thus making the cues irrelevant for spatial information. In the third experiment we checked for the effect of valid and invalid cues by presenting stable spatial cues throughout the trials and making the cues spatially uninformative regarding the stimulus position about 20% of the time.

Experiment 1

In the first experiment in a within subjects design (N = 32), we presented randomly chosen (in 50-50 ratio from a total of 128 trials) trials of serially shown colored squares or grayscale shapes in set sizes of 4, and 6 in different spatial location on both sides of a fixation cross in a horizontal line. Only one item in the sequence was presented with a spatial cue (in the form of underline). The task was a recency judgement where two of the items in sequence were presented on both sides of the fixation cross and subjects had to respond saying which came later in the sequence. The cued item was probed for recall in half the trials.

Experiment 2

In this experiment in a within subjects design (N = 61), we presented randomly chosen (in 50-50 ratio from a total of 128 trials) trials of serially shown colored

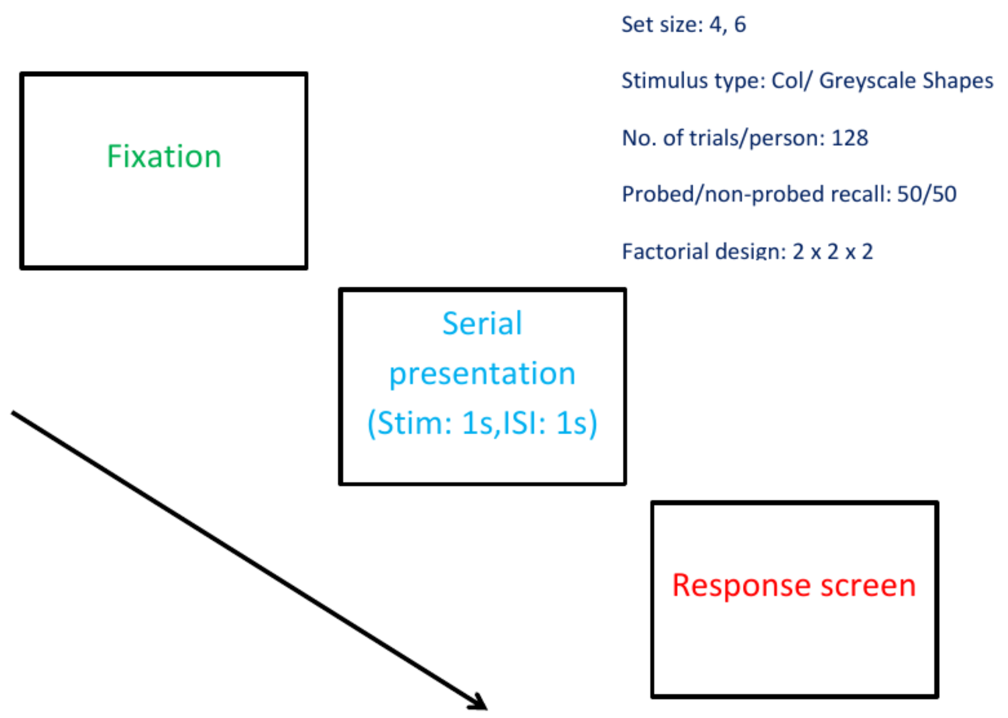


Figure 5.1: The schematics of Experiment 1

squares or grayscale shapes in set sizes of 2, 4, 6 and 8 in different spatial location on both sides of a fixation cross in a horizontal line. In half of the trials the sequence was presented with a stable spatial cue under each item as it appeared on screen (in the form of underline). The task was a recency judgement where two of the items in sequence were presented on both sides of the fixation cross and subjects had to respond saying which came later in the sequence.

Experiment 3

In the last experiment in a within subjects design (N = 20), we serially presented colored squares or grayscale shapes in set sizes of 2, 4, and 6 in different spatial location on both sides of a fixation cross in a horizontal line with persistent direct spatial cues (in the form of underlines). The spatial cues were helpful 80% of the time and misleading 20% of the time. Total number of trials was 100 per subject.

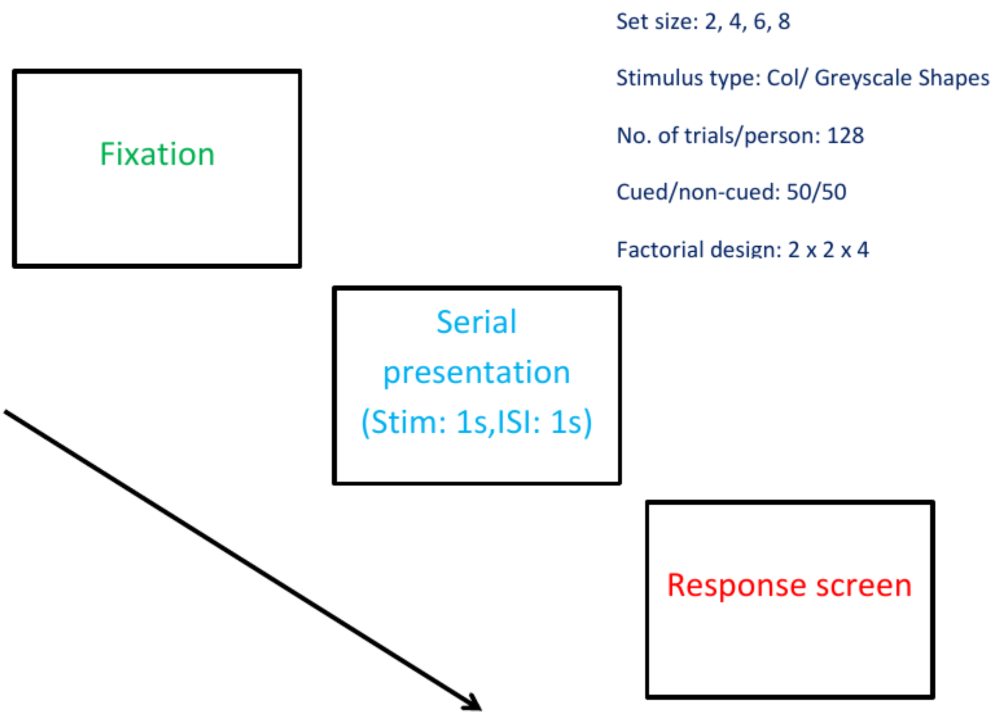


Figure 5.2: The schematics of Experiment 2

Participants:

All the participants in this study were 24-32 years of age. Informed consent was obtained for all participants in accordance with the local ethical committee guidelines.

Stimuli:

There were 2 kinds of stimuli- (i) 2/4/6/8 different colored squares (Red, Green, Cyan, Blue, Yellow, Pink, Magenta, Grey) of 80 px X 80 px each. (ii) 2/4/6/8 different grayscale shapes of 80 px X 80 px each. All the stimuli were presented on a white background. Locations were randomized in all the trials. A black Fixation Cross of 20 px X 20 px was presented at the centre of the screen. The order of occurrence of color and/or shapes was randomized in each block.

The location cues were 5 px X 80 px lines drawn appearing vertically 20 px

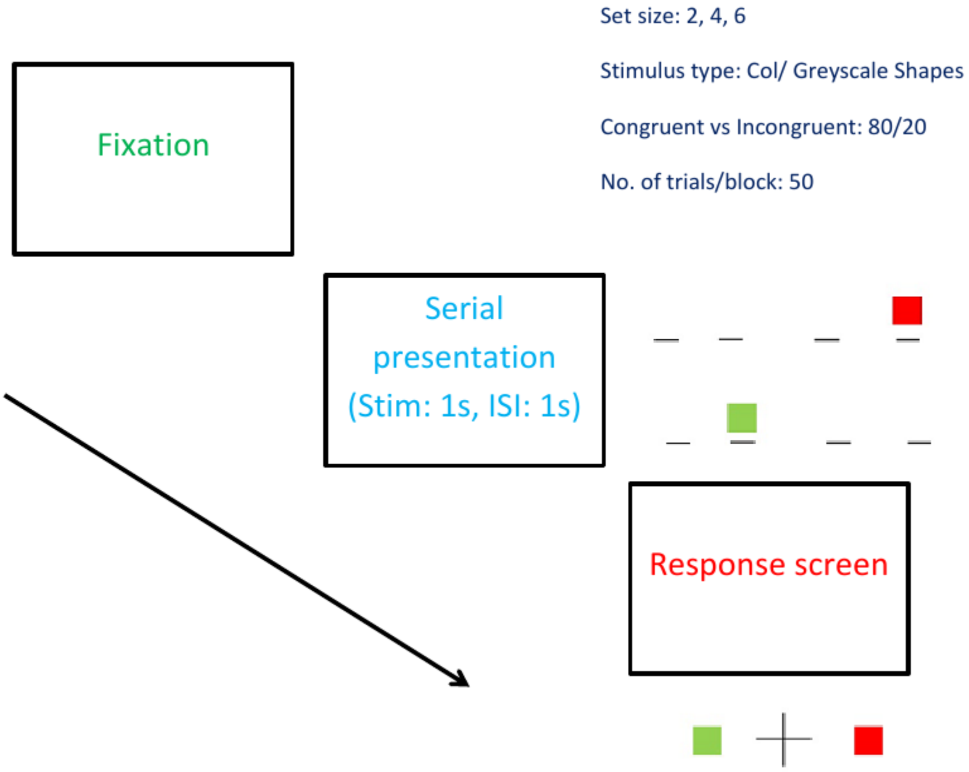


Figure 5.3: The schematics of Experiment 3

below the stimulus position.

Response Collection

A response screen was shown after a gap of 1 sec from the end of stimulus presentation. Response screen featured two items from the presented stream of stimuli on two sides of the fixation cross. The participant had to press the *LeftArrow* or *RightArrow* key on the keyboard to indicate which item had come later in the sequence. The choice and display location of the response items were randomized across trials within the blocks. There was no inter-trial interval, the next trial started as soon as the response was made.

5.3 Results

Experiment 1

The results show that the spatial cue helps recall in terms of reaction times for grayscale shapes ($F(1, 31) = 5.6$, $p < 0.05$), but not facilitatory for colored squares ($F(1, 31) = 0.07$, $p < 1$). The reaction times are shown in Fig. 5.4 and the accuracy results are shown in Fig. 5.5.

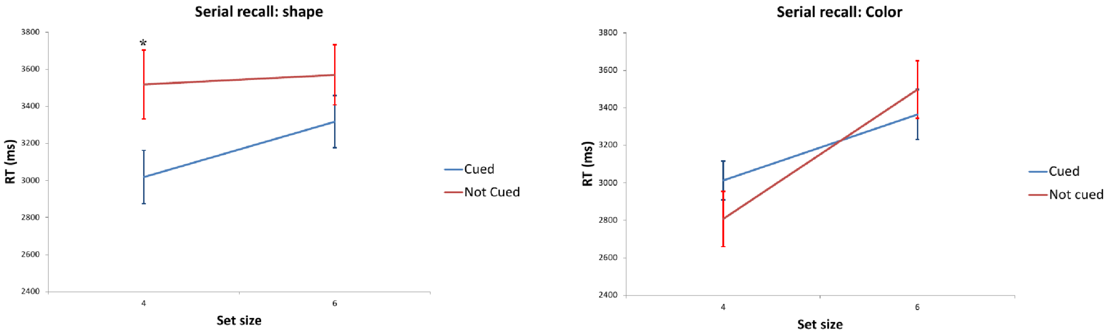
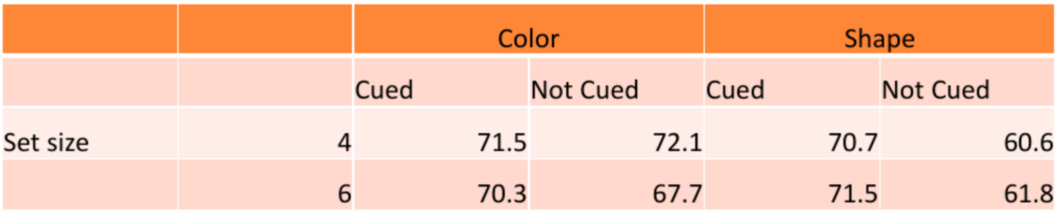


Figure 5.4: Reaction time results for Experiment 1. The results show clearly that in terms of reaction times, for shapes the recall of spatially cued item is facilitated, but not for colors.



		Color		Shape	
		Cued	Not Cued	Cued	Not Cued
Set size	4	71.5	72.1	70.7	60.6
	6	70.3	67.7	71.5	61.8

Figure 5.5: Accuracy results for Experiment 1. The results do not seem to show a speed accuracy trade-off for shape recall. Showing that they are better and faster at shape recall if the tested item's location was made spatially salient.

Experiment 2

In this task as the spatial cues were not task relevant, it did not have any effect on the reaction times for shape recall, $F(1, 60) = 0.05$, $p < 1$, or for color recall, $F(1, 60) = 0.68$, $p < 1$ (see Fig. 5.6). The accuracy values also show similar patterns (see Fig. 5.7).

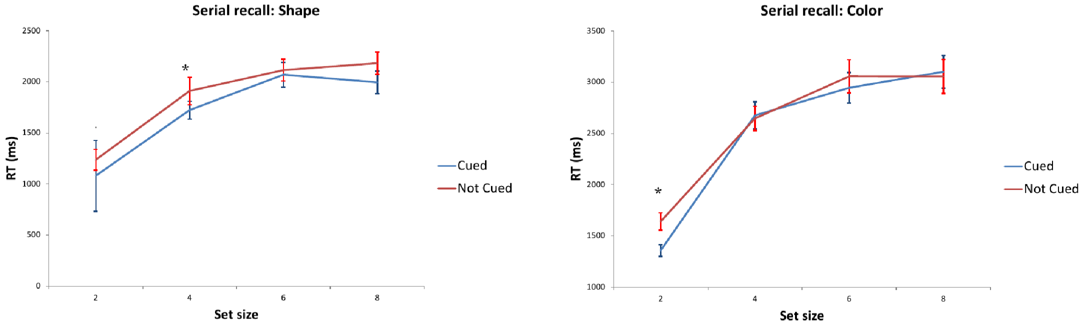
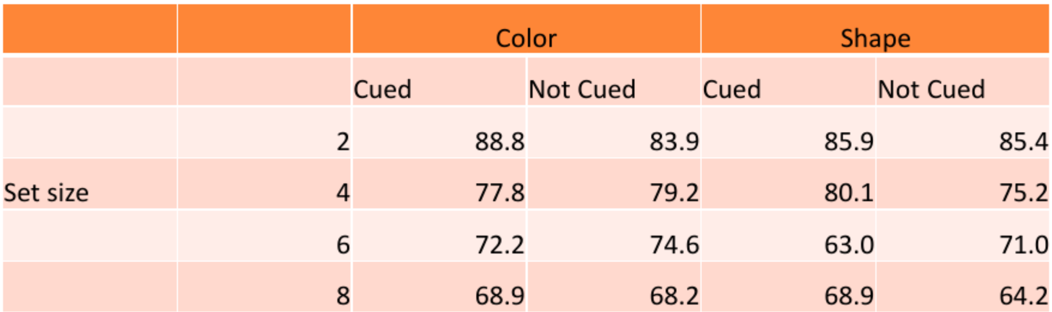


Figure 5.6: Reaction time results for Experiment 2. In this experiment we see that spatial cues do not have any noticeable effect on either shape or color recall for any set sizes.



		Color		Shape	
		Cued	Not Cued	Cued	Not Cued
	2	88.8	83.9	85.9	85.4
Set size	4	77.8	79.2	80.1	75.2
	6	72.2	74.6	63.0	71.0
	8	68.9	68.2	68.9	64.2

Figure 5.7: Accuracy results for Experiment 2. The accuracy results seem to be consistent with the trend in RT's

Experiment 3

The results show that although spatial cues do facilitate serial recall for grayscale shapes ($F(1, 19) = 6.16$, $p < 0.05$, see Fig. 5.8), they are disadvantageous for color recall ($F(1,19) = 3.85$, $p < 0.05$, see Fig. 5.8)). Similar effects are also seen for accuracy (see Fig. 5.9).

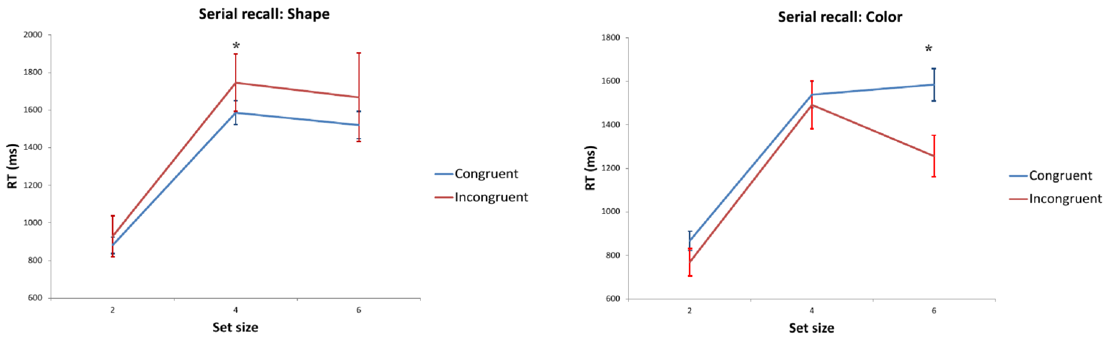
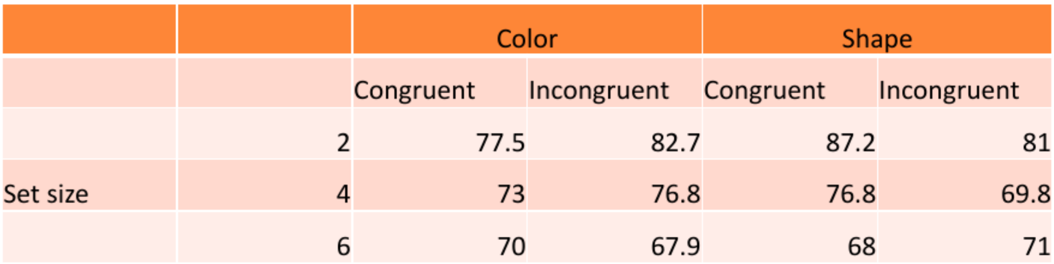


Figure 5.8: Reaction time results for Experiment 1. The results show clearly that in terms of reaction times, the spatial cue probe are facilitatory for serial recall of shapes, but not for color recall.



		Color		Shape	
		Congruent	Incongruent	Congruent	Incongruent
Set size	2	77.5	82.7	87.2	81
	4	73	76.8	76.8	69.8
	6	70	67.9	68	71

Figure 5.9: Accuracy results for Experiment 3

The interaction between color and shape recall is better illustrated in Fig. 5.10.

5.4 Discussion

The effect of spatial cue in Experiment 1 is task specific and thus seems to facilitate serial recall for shapes, when cued for single item. However, spatial cues interact with color and shape differentially. So in Experiment 2 we made the spatial cues task irrelevant by having them appear with all the items or none.

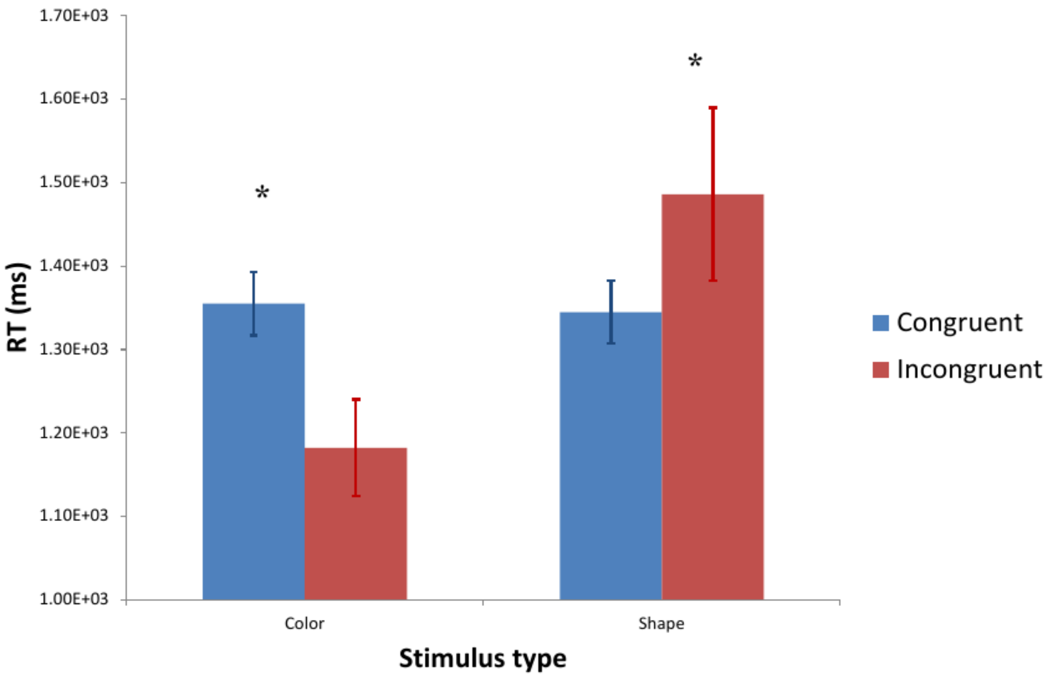


Figure 5.10: Interaction between color and shape recall. In terms of reaction times, the results show a clear interaction between color and shape recall performances.

The spatial cues were not task relevant and thus did not produce any remarkable effect as seen in the previous section. In Experiment 3 we again brought back task relevant cues and this time they were stable (and only 4 in number irrespective of set size). This time we see the interaction between shape and color recall again, with shape recall facilitated by spatial cues whereas color recall being impeded by spatial cues.

There can be at least two possible explanations:

- The inclusion of cues themselves has caused this anomalous effect and not anything to do with working memory per se.
- Spatial cues interfere with serial recall.

However we have seen from Experiment 2 that the effect disappears when the cues are task irrelevant. The computational saliency map model seems to account for shape recall but not color recall. Probably a more detailed computational account might give better insight into the mechanism of such interactions. Further neuroimaging studies are required to explore the neural substrates.

Overall, the experiments corroborate the variable precision dynamic models for working memory supported by van den Berg et al. [2012], as opposed to slot models for working memory elucidated by Zhang and Luck [2008].

Chapter 6

Empirical investigation of time's subjective expansion

> There is really nothing more to say except why. But since why is difficult to handle, one must take refuge in how.
>
> *The bluest eye*
> Toni Morrison

6.1 Introduction

Subjective expansion of time (TSE) is the phenomenon where human beings perceive an oddball stimulus in a stream/ series of identical standard stimuli to have lasted longer in terms of percept duration [Tse et al., 2004]. There are several competing theories that have tried to explain this phenomenon. However, they can be roughly classified into - a) dedicated internal clock b) information based decision mechanism [Hellström, 1985; Thomas and Weaver, 1975; Treisman, 1963]. Here we have used a computational model to explore the hypothesis that on-center off-surround saliency map [Sengupta et al., 2014b] (increased saliency of oddball) combined with winner-take-all decision can account for the oddball effect.

6.2 Model

The model and the parameters are described in detail in Chapter 3. Here we present the results relevant for the temporal oddball paradigm. We saw previously that the model predicts that habituation should have an effect on time's subjective expansion factor in temporal oddball paradigm. However, the model also predicts that this effect should be uneven across different standard duration levels. The effect should be most around 0.3-0.4 s region and should be very less around 1 s standard duration. Thus we chose 2 standard durations 0.38 s and 1.05 s, in order to test the qualitative predictions from the model.

6.3 Method

Participants

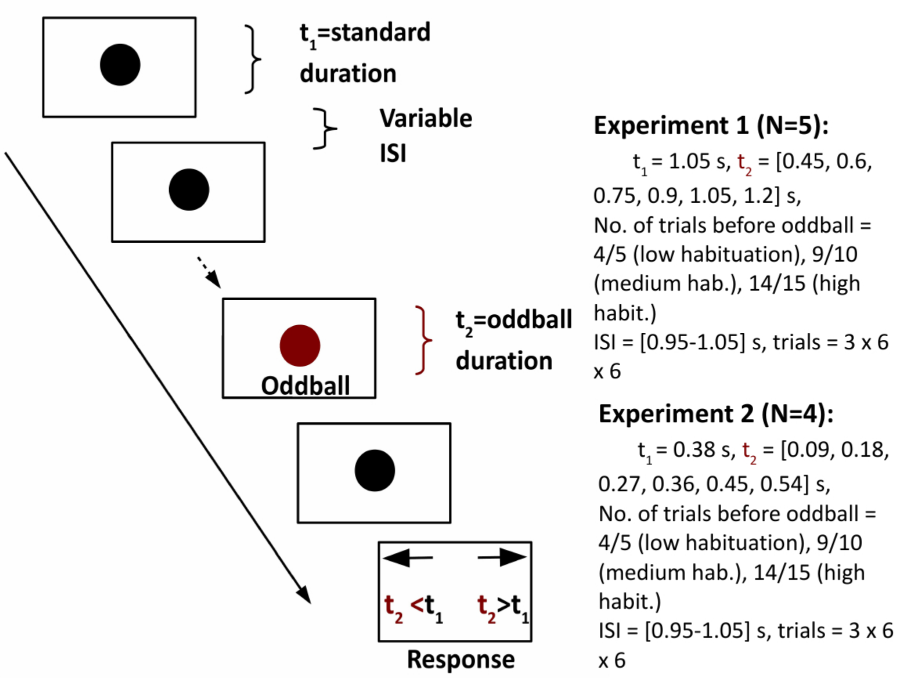
A total of 9 adult volunteers participated in this study (age range: 24-32). All the participants were tested at the Center for Cognitive Sciences, B. V. Raju Institute of Technology - Hyderabad, India. The participants included 5 male and 4 female participants. Informed consent was obtained from all participants in accordance with the local ethical committee guidelines.

Stimulus

We used a temporal oddball stimuli where a red circle at the center of the screen was used as an oddball among black circles presented at the center of the screen.

Design

Two standard durations were used 0.384 s and 1.05 s on same subjects on different days. The corresponding oddball durations used were [0.09, 0.18, 0.27, 0.36, 0.45, 0.54] s and [0.45, 0.6, 0.75, 0.9, 1.05, 1.2] s.There were three habituation conditions - the oddball could appear in a stream of 6, 11, or 16 displays. The exact position of the oddball was randomly manipulated between 5/6, 10/11, and 15/16. Habituation levels were presented in randomised trials. The inter-stimulus



Experiment 1 (N=5):
t_1 = 1.05 s, t_2 = [0.45, 0.6, 0.75, 0.9, 1.05, 1.2] s,
No. of trials before oddball = 4/5 (low habituation), 9/10 (medium hab.), 14/15 (high habit.)
ISI = [0.95-1.05] s, trials = 3 x 6 x 6

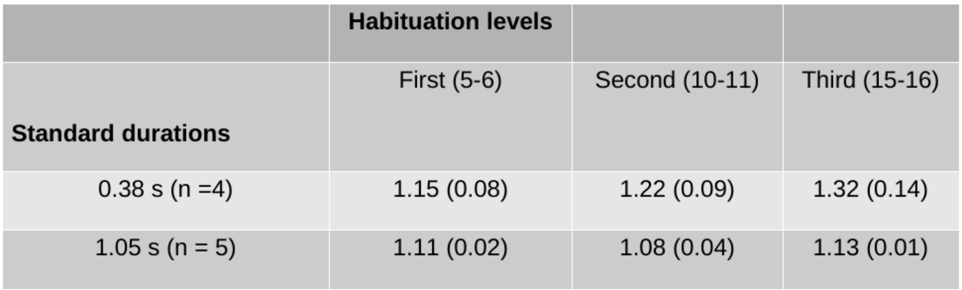
Experiment 2 (N=4):
t_1 = 0.38 s, t_2 = [0.09, 0.18, 0.27, 0.36, 0.45, 0.54] s,
No. of trials before oddball = 4/5 (low habituation), 9/10 (medium hab.), 14/15 (high habit.)
ISI = [0.95-1.05] s, trials = 3 x 6 x 6

Figure 6.1: The design of the temporal oddball experiment

interval (ISI) was jittered between 0.95-1.05 s. In each habituation block each oddball duration was presented 6 times. The total number of oddball trials for each standard duration was 3 x 6 x 6 (no. of habituation conditions x no. of oddball durations x no. of repeats per oddball duration) or 108. The design is shown in Fig. 6.1. 5 subjects were used for the experiment (of whom 4 underwent the experiment for both the standard durations).

Response Collection

The response screen was presented after each trial sequence (of 6, 11 or 16 displays). Where they were asked to indicate with a key press wheather they found the oddball to be longer in duration than the standard stimulus or not. The participants pressed right arrow key for ‘longer’ response and left arrow key otherwise.



	Habituation levels		
Standard durations	First (5-6)	Second (10-11)	Third (15-16)
0.38 s (n =4)	1.15 (0.08)	1.22 (0.09)	1.32 (0.14)
1.05 s (n = 5)	1.11 (0.02)	1.08 (0.04)	1.13 (0.01)

Figure 6.2: The average TSE values for all the subjects for all the conditions with standard error given in parenthesis.

6.4 Results and analysis

We analysed each habituation block for each standard duration separately. For each habituation block for one standard duration, there was 36 trial sequences. Among these 36 trial sequences six oddball durations appeared for six repetitions each. We used the proportion of trials in which the oddball duration was perceives as longer as the response variable (the proportions range 0, 1/6, 1/3, 1/2, 5/6, 1). For each subject we plotted this response variable against the oddball durations (for each habituation block for each standard duration) and used a Weibull fit [1] to determine the point of subjective equality (PSE), the oddball duration where the performance or response variable was 50%. The calculated TSE values along with the standard error (in parenntthesis) are given in Fig. 6.2.

We calculated the PSE from each standard duration condition at each habituation level for all the subjects. The calculated TSE values ($TSE =$ standard duration$/PSE$) show the trend predicted by the model. Generally the increasing habituation levels lead to more subjective expansion of time. However the effect of habituation is most pronounced for 0.38 s standard stimulus than for 1.05 s stimulus. The results are shown in Fig. 6.3.

---

[1] Weibull fit was calculated using the cumulative Weibull function $y = 1 - e^{-(\frac{x}{a})^b}$.

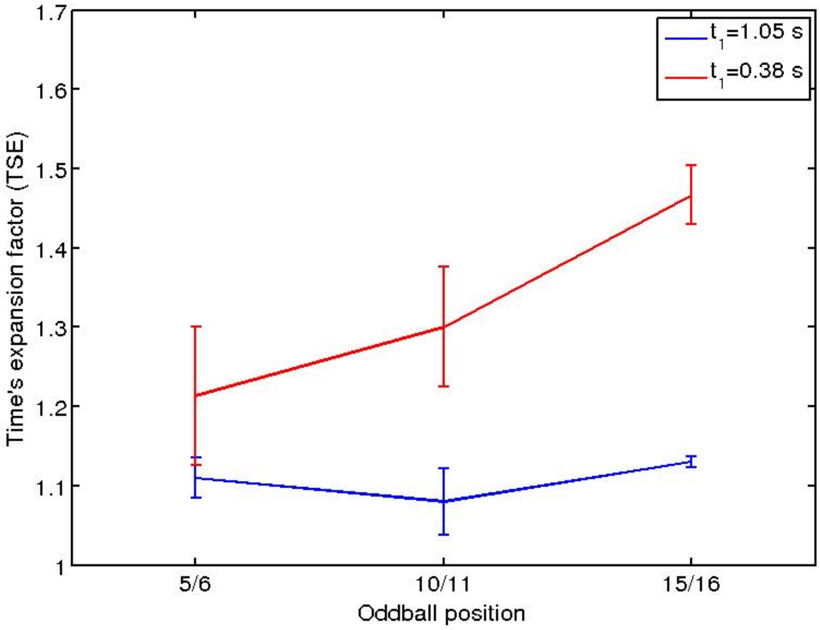


Figure 6.3: TSE values calculated for each habituation condition for both the standard durations. The results confirm the model predictions

6.5 Discussion

The model derives predictions for temporal judgments based on the assumption that oddball stimuli are perceptually more salient than the standards. The model output captures the pattern of temporal expansion (TSE) for different standard durations given in Tse et al. [2004], including the 120 ms threshold criterion. We also surmised that greater habituation to standard stimuli should lead to greater salience for the oddball stimulus, and thus influences TSE via a winner-take all decision mechanism. The model's prediction for differential effect of habituation on TSE for different standard intervals was confirmed empirically.

Chapter 7

Discussion and Conclusions

> There is more wisdom in your body than in your deepest philosophy.
>
> ---
>
> *Thus spoke Zarathustra*
>
> FRIEDRICH NIETZSCHE

7.1 Numbers, memory, and time

Visual sense of numbers

In our present work we have shown how a single, flexible model of object individuation and number perception can account for several important phenomena in the domain of human numerical cognition (e.g., distance effect, Fechner's law, etc) as well as adult human numerosity comparison data. The current model is different from previous attempts at modeling visual number [Dehaene and Changeux, 1993; Grossberg and Repin, 2003; Stoianov and Zorzi, 2012] because it did not assume the distinction between large and small numerosity a priori, but rather allowed the diverse number ranges to emerge through self-organization of the same network. An interesting and novel aspect of our approach is the modeling of reaction times based on an energy function. This formulation led to the prediction of specific asymmetric switch costs between small and large numerosities in terms of reaction times (RT) (see Fig. 4.1). The model output also led to the prediction of severe underestimation following a small-large enumeration switch. A hallmark of any successful computational model lies in its ability to offer testable predictions as well as explanation of known phenomena. In the current work we set out to test the prediction of the model in a behavioral enumeration task. We did find that for small numerosities there is a very significant ~ 150 ms switch cost for the sudden switch from large numerosity estimation to small numerosity enumeration, whereas no significant switch cost exists for small to large switch (see Fig. 4.6). The reaction time prediction from the model was highly correlated with the empirical data for both the non-switch and switch conditions in a linear fit (Fig. 4.8 and 4.9). Critically, the same parameters yielded good results for both numerical ranges and switch conditions. This formulation led to the prediction of severe underestimation following a small-large enumeration switch. Model's prediction for large underestimation following a small-large switch was also confirmed in error measurements (see Fig. 4.3).

It is fairly common to encounter such asymmetric switch costs in task-switching literature as Gilbert and Shallice [2002] have shown. Their PDP model accounts for such asymmetries in switch costs through postulations of interacting networks of colors and words in Stroop-related tasks. However, in the present case appears to be somewhat more complicated than the standard cases. Firstly, even though a very large switch cost exists for large to small switch in terms of RT, there is no such cost for small to large switch. One might attempt to explain that through speed accuracy trade-offs in the case of small to large switch, i.e., one might say that there is not much increase in switch cost but the precision suffers in the switch trials as a result. However, this does not explain the systematic underestimation that is seen the small to large switch, whereas no such systematic under- or overestimation effects are present for large to small switch.

It might be possible that small and large numbers are processed by related yet different networks. However, such a suggestion leads to some difficulties. The reaction time switch cost from large to small switch can be explained in the account with two networks, but then we have to assume different interaction weights for small to large switch which gives almost no switch cost in terms of RT. On the other hand the characteristic underestimation in case of small to large switch does not translate into an overestimation for large to small switch (although the accuracy suffers). It is also difficult to explain the duration of switch cost from trial to trial as seen in Fig. 4.4 and 4.7 in two-network account without resorting so some very complicated and questionable assumptions (like different weights for network A to network B transition compared to B to A transition). Our account is more parsimonious and able to account for all the observed effects within the same network.

Overall, the ability of a single, flexible model to account for human performance in both small and larger number ranges has implications towards the debate regarding mechanisms of subitizing and estimation [Dehaene and Changeux, 1993; Melcher and Piazza, 2011; Revkin et al., 2008; Whalen et al., 1999]. One possibility, suggested by recent work looking at the role of attention in numerical judgments (Burr et al, 2010), is that the use of focused attention in order to individuate a small number of items acts to increase the inhibition in the network. This idea is consistent with attention models, such as the biased competition model, which account for the ability of neurons to respond to an attended stimulus in the presence of distractors. Based on previous neuroimaging and neurophysiological studies, neurons representing an attentional priority (saliency) map in area LIP [Roggeman et al., 2010] might act in a way similar to the recurrent on-center off-surround network model used here. In other words, the remarkable overlap among studies of attention, object individuation and working memory, number and eye movements in the LIP area may not be simply a coincidence but rather implicate a shared, flexible attention priority map in representing the presence of important objects in the visual environment.

Working memory

We have seen that the model also captures the salient aspects working memory, especially capacity limits and primacy effects. The use of similar network for working memory and spatial salience allowed us to ask the question regarding the effect of spatial cuing on serial recall. The on-center off-surround architecture predicted interestingly that spatial cuing would facilitate serial recall of recent items. This shows that the modeling framework is fairly versatile in exploring problems in the visual spatial domain. The prediction was only partially confirmed (in case of serial recall of grayscale shapes and not for colors). However, it gave us an interesting insight into possible competing mechanisms within working memory for shapes and colors.

Time and expansion

We have also seen that the model through implementation of winner-take-all paradigm correctly replicates the experimental results on time's subjective expansion in the temporal oddball paradigm by Tse et al. [2004]. Moreover, it also predicts some novel paradigms regarding habituation and temporal oddball. It predicts that the effect of habituation on time's subjective expansion factor will be more if the standard stimuli are 200-400 ms in duration, but not so much in case of 1.05 s stimulus duration. The model prediction was confirmed [Sengupta et al., 2014c] in a series of experiments involving the temporal oddball paradigm.

7.2 Embodiment, representation, and saliency map

In essence embodiment stance in cognitive science asserts a fairly simple but deep departure from the standard approaches that are commonly labeled as Cognitivism (i.e., both functionalism and connectionism along with other derivatives of physical symbol systems stance) - human cognition is located within a dynamic goal-directed action-oriented process involving both the body and the environment [Wilson, 2002]. Even if specialized cognitive functions are served by different parts of the brain, the involvement of body and environment amounts to much more than input-output circuitry. We ascribe meaning to our world based on the constraints posed by the body [Wilson and Clark, 2009] and our roles as being-in-the-world is regulated by the body [Dreyfus, 1990]. The embodied stance is far from uniform. Early proponents like Lakoff and Johnson [1980] propose a view where a metaphorical process deriving from the body and its structure is central to human cognition. Dreyfus [1990] derives a Heideggerian account from his consideration of AI and robotics, to propose an anti-representational process using the world as its model Dreyfus [2008]. Freeman [1995] proposes an intentional arc (following Merleau-Ponty, 2012) through which an organism has its cognitive dispositions (determined by corresponding attractor states) altered through its action in the environment. These three major viewpoints have their advantages and also some inherent distinctions. In the next few paragraphs I would like to point out conceptual distinctions between these positions and will attempt to arrive at the necessary conceptual apparatus needed to address the question of the temporal experience of human cognition.

Lakoff and Johnson [1980]'s account in some form does adhere to at least a moderate stance of the Sapir-Whorf hypothesis [Sapir, 1983; Whorf, 1956] for linguistic relativism (seen more explicitly in their later work, Lakoff and Nuňez, 2000). Metaphoric process within the unconscious can impinge upon the meaning formation because the linguistic categories determine our world view. To extend their argument, one has to assume at least a linguistic representation of the world, but one that is grounded in the body rather than being arbitrary like in the case of physical symbol systems hypothesis (PSSH, see Newell and Simon, 1976). Figure 7.1 elucidates the major import of their position on embodiment.

Dreyfus [1990] has rejected Cartesian ideas inherent in traditional AI and Cognitive science (i.e., the position that context-free, disembodied, representations governed by rule-like algorithms separated from and independent of the cognitive being and the world inhabited by it, are sufficient for building AI or understanding cognitive psychology). From his analysis of Frame problem in AI (writing a closed set of axioms for all given and novel situations), Dreyfus [2008] points out that the fundamental difficulty of a Cartesian ontology where one does not just see something, but sees it as something - there is always an extra step of deciding how to apply meaning to a given situation. On the other hand in the Heideggerian approach meaning is always ready-at-hand,

> To say a hammer has the function of being for hammering leaves out the defining relation of hammers to nails and other equipment, to the point of building things, and to our skills all of which Heidegger called readiness-to-hand and so attributing functions to brute facts couldn't

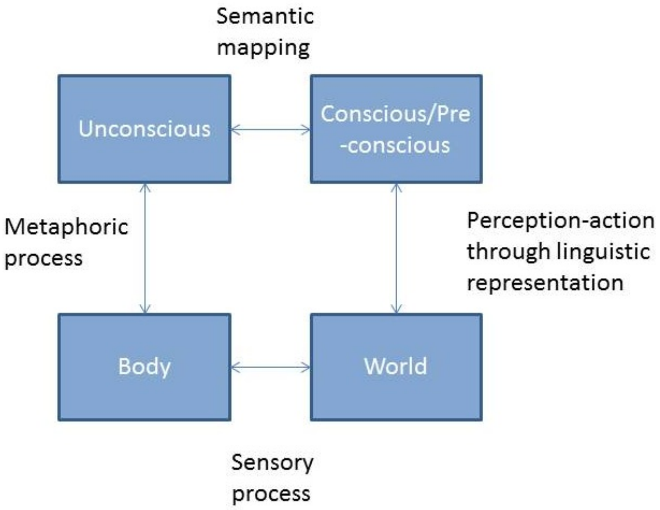


Figure 7.1: Embodied stance of Lakoff and Johnson [1980].

> capture the meaningful organization of the everyday world. -Dreyfus [2008]

In his account, embodiment exists in "being-in-the-world" relation between cognitive agent and the world, what Dreyfus [2008] describes as 'background coping'. He uses Gibson [1979]'s idea of affordances to explain how the world affords the actions available for the cognitive agent in a continuous and non-representational manner. In fact, he criticizes the idea of Cartesian self-enclosed entity itself, rather thinking of experience as a flow during 'everyday coping' [Dreyfus, 2000].

Although Freeman [1995]'s account is very similar to that of Dreyfus [2008], there are some key differences. His account derives from the works of Aquinas [2000] and Merleau-Ponty [2012]. The key concepts here are unity (by the mind's assertion of a boundary against the outside) and an intentional relation (between intellect and material objects) which shapes itself by changing its scales to accommodate the object through the process of knowing it. Intentionality as defined here is not "about-ness" [Searle, 1992]. Meaning is achieved through action in the world, and the self is altered by that action. Although [Aquinas, 2000] describes intentionality as directing action towards some future defined and chosen

	Representation	Idea of self	Basis of cognition
Lakoff & Johnson	Linguistic	Through metaphoric processes connecting the body and the world	Unconscious metaphoric process
Dreyfus	Non-representational account	Not self-enclosed, thus even unaware of self during ‘everyday coping’	“Being-in-the-world” relation between cognitive agent and the world
Freeman	Ordered dispositions act as surrogates to representation	Through unity and intent self distinguishes from and is modified by non-self	Intentionality rooted in the unconscious mind

Figure 7.2: The relation between different strands of embodied stance

by the actor, intentionality is fundamentally an unconscious process, i.e., the self is not necessarily a conscious agent. Merleau-Ponty [2012]’s formulation of “the intentional arc” that completes an organic loop from action through the world and back into the brain, resembles Aquinas’ active intellect. The intentional arc names the tight connection between the agent and the world, perception and action. The skills acquired by the agent are not stored as representations but as dispositions to respond to directing situations in the world. Thus for Freeman [1995], the term intentionality means the “process of a brain in action having the properties of unity, wholeness, and intent.” Unity refers to a state of integration by which a self distinguishes itself from non-self. Wholeness refers to a bounded process by which through stages a self actualizes its mature form, ultimately to die. Intent refers to a relation in which a self modifies itself in conformance with aspects of non-self. A living brain thus capable of actualizing these properties by purposive behavior is an intentional structure.

From the above discussion it emerges that we can distinguish the three aforementioned strands within the embodied cognition literature along three dimensions: representation, idea of self, and basis of cognition. The differences are summarized in Figure 7.2.

The above table makes it explicit that there are some major irreconcilable differences within the three strands of embodiment. Lakoff and Johnson’s idea of linguistic representation is incompatible with the others, but their idea of an unconscious process as a basis of cognition is fairly in line with Freeman's account. However, it is unclear how much the metaphoric process can be reconciled with the intentional arc in the perceptual domain. On the other hand, Dreyfus's idea of "being-in-the-world" relation is not compatible with Freeman's notion of self that distinguishes itself from non-self through unity.

In both Dreyfus and Freeman's account representation is disposed off almost completely. However, we know that there is some validity to the idea of representation from researches in retinotopic and topographic maps in the brain. It is not confined just to the visual domain either [Chen et al., 2013]. The main problem is that thinking of representation in a linguistic manner guarantees a Cartesian ontology. However, any idea of a representation that is completely non-linguistic is doomed to fall flat in the case of human cognition. As shown by Dreyfus [1990], on-line cognition can be traced to a state of flow (being-in-the-world) that does not require any mediating representation. However, we do employ quite a lot of off-line cognition that do definitely rely on representation. These representations may be derived from the body as some views of embodiment literature hold [Wilson, 2002]. Even so, it does not discount the possibility of a representational matrix (as seen in saliency map literature by Melcher and Piazza, 2011; Roggeman et al., 2010 or in literature on topographic maps, like Seelke et al., 2012) which is essentially a spatial one.

The idea of a spatial saliency map representation underlying perception and action is not limited to just vision, but involves auditory [Kayser et al., 2005] and tactile [Diederich et al., 2003] modalities as well. These saliency maps are part of pre-attentive visual stream and may not be cognitively penetrable [Pylyshyn, 1999]. A subsequent decoding of a saliency map requires linguistic features of cognition. Moreover, Bonato et al. [2012] have recently summarized an array of literature that points to various similarities between spatial and temporal processing and suggest that space, time, and number processing might be grounded in neuronal structures for spatial attention and sensori-motor processing. Spatial numerical abilities in other animals [Dehaene and Changeux, 1993] suggests that spatial processing might operate at least partly in an independent manner from linguistic ability. Given the above we can make a case for spatial representations that are grounded in sensori-motor process, i.e., embodied, and also can maintain relative autonomy from linguistic representations. This perspective is explored further in Appendix B.

7.3 Future directions

In the present thesis we saw a recurrent dynamical neural network that able to capture salient features of individuation, an spatial attentional phenomena, in a detailed fashion. However, individuation is also limited temporally [Wutz and Melcher, 2014]. We saw in previous chapters that some of the temporal effects can indeed arise from spatial representation of a self-organized recurrent network. In the future I would like explore the spatio-temporal integration windows for individuation and attention further using similar architechtures.

Another important aspect worth persuing in future seems to be locating the neural correlates of the network inhibition parameter in terms of both location and time, i.e., which parts of the brain do influence the network inhibition and how far in the pre-stimulus domain they need to be to affect the prection-action cycle.

Appendix A: Mathematical Details

Stability Analysis

The differential equation governing the time-evolution of the network of N nodes is given by

$$\frac{dx_i}{dt} = -\lambda x_i + \alpha F(x_i) - \beta \sum_{j=1, j\neq i}^{N} F(x_j) + I_i + noise \tag{1}$$

$x_i(t)$ is the activation of node i at time t. I_i represents the intensity of external input ($\forall i, 0 \leq I_i \leq 1$), it is zero if the stimulus is absent for a particular node at that time point. Input is only presented for a finite amount of time, typically much less than total time of simulation. $F(x)$ is the activation function given by the formula,

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \frac{x}{1+x} & \text{for } x > 0 \end{cases} \tag{2}$$

Network should reach steady state activity when the external input is taken away. If we disregard noise, at steady state, i.e. when, $\frac{dx_i}{dt} = 0$,

$$\lambda x_i = \alpha F(x_i) - \beta \sum_{j=1, j\neq i}^{N} F(x_j) \tag{3}$$

As the equation is symmetric under permutation of units, the system should have symmetric solutions characterized by number of active units n, and their activation $x(n)$, all other units having 0 activation.

$$x(n) = \left(\frac{\alpha - (n-1)\beta}{\lambda}\right) F(x(n)) \tag{4}$$

Using Eq. 2 and 4, we get

$$x(n) = \left(\frac{\alpha - (n-1)\beta}{\lambda}\right) - 1 \tag{5}$$

Now, noise can bring in additional fluctuation that can destabilize the solution for a pair of active modes (with equal activation according to Eq. 5), unless the the difference of activations between the said nodes $\Delta x = x_i - x_j$ decays. Using Eq. 1 & 5 we get

$$\frac{d\triangle x}{dt} = \triangle x \left[-\lambda + \lambda^2 \left(\frac{\alpha + \beta}{(\alpha - (n-1)\beta)^2}\right)\right] \tag{6}$$

Thus the fluctuation decays only if $\frac{d\Delta x}{dt} < 0$, i.e.,

$$\frac{\alpha + \beta}{(\alpha - (n-1)\beta)^2} < \frac{1}{\lambda} \tag{7}$$

As we can see that the decay parameter, excitation parameter and inhibition parameter are not completely independent for stable solutions. For the present purposes we use $\lambda = 1$.

Formulation for the expression of reaction time

In order to explore the stability of the network in a more rigorous manner, we tried to derive an expression for a Hamiltonian (or Energy function) for such a network from first principles. A desirable property for such a Hamiltonian should be that it exhibits properties similar to Lyapunov functions, at least under suitable range of parameter choices.

According to classical mechanics for a set of generalized co-ordinates x, the time evolution according to the Hamiltonian principle is given by,

$$\frac{\partial x}{\partial t} \propto \frac{\partial H}{\partial \dot{x}} \tag{8}$$

$$\frac{\partial \dot{x}}{\partial t} \propto -\frac{\partial H}{\partial x} \tag{9}$$

where H is the Hamiltonian (or the energy) of the system. In the present system the activations of the nodes can be taken as generalized co-ordinates for the system. Given Eq. 8 we can roughly say that at any point in time evolution (considering $\lambda = 1$ and proportionality constant as 1) disregarding noise, the Hamiltonian equation for a particular node i is given by,

$$dH_i \propto \frac{\partial H}{\partial \dot{x}_i} d\dot{x}_i + \frac{\partial H}{\partial x_i} dx_i = k_1 \dot{x}_i d\dot{x}_i - k_2 \ddot{x}_i dx_i \tag{10}$$

where k_1 and k_2 are proportionality constants (it is reasonable to assume that $k_1 \neq k_2$). Now from Eq. 1 we have

$$d\dot{x}_i = -dx_i + \alpha \left(\frac{F(x_i)}{x_i} \right)^2 dx_i \tag{11}$$

$$\ddot{x}_i = -\dot{x}_i + \alpha \left(\frac{F(x_i)}{x_i} \right)^2 \dot{x}_i \tag{12}$$

Inserting Eq. 11 in Eq. 10 we get

$$dH_i = -(k_1 - k_2) \left(1 - \alpha \left(\frac{F(x_i)}{x_i} \right)^2 \right) \dot{x}_i{}^2 dt \tag{13}$$

and the total energy,

$$H = \sum_i H_i \propto -\sum_i \int \left(1 - \alpha \left(\frac{F(x_i)}{x_i} \right)^2 \right) \dot{x}_i{}^2 dt \tag{14}$$

From Eq. 5 we can substitute terms in steady state to get

$$H \propto -\sum_i \int \left(1 - \frac{\alpha}{(\alpha - (n-1)\beta)^2}\right) \dot{x}_i{}^2 dt \tag{15}$$

Now if $dH < 0$ and thus a monotonically decreasing Lyapunov type function in absence of external input, we have the stability condition as

$$\frac{\alpha}{(\alpha - (n-1)\beta)^2} < 1 \tag{16}$$

Comparing this to Eq. 7, we can see that the conditions derived from the energy value is slightly different and diverges greatly for higher β. This is due to the fact that Eq. 7 excludes winner-take-all mechanisms operating at higher inhibition, whereas the energy function does not (See Chapter 5 for details regarding the winner-take-all mechanism as simulated from network dynamics). And it is evident that for all β,

$$\frac{\alpha}{(\alpha - (n-1)\beta)^2} < \frac{\alpha + \beta}{(\alpha - (n-1)\beta)^2} < 1 \tag{17}$$

and thus the energy function is a very suitable candidate for the network as it is in line with the stability analysis derived from the dynamics of the network. However, there was an implicit assumption in the analysis presented here, mainly $k_1 > k_2$, which is justified as otherwise the system will not reach convergent solution. Interestingly as mean activation values are higher at lower β, it stands to reason that higher inhibition will also have higher energy than the lower ones.

In order to apply the equation to get to reaction time (RT) distribution, we assume that reaction times correlate with the allowed fluctuation in energy (i.e., $-\int dH$). More energy to dissipate, more the reaction time. And thus,

$$RT \sim -\int dH \propto \sum \int \left(1 - \alpha \left(\frac{F(x_i)}{x_i}\right)^2\right) \dot{x}_i{}^2 dt \tag{18}$$

We know from Mohamed et al. [2004] and Yarkoni et al. [2009] that reaction times do correlate with highest values of cortical activation and also be used to predict trial-by-trial variability in reaction times. As we have seen from Fig. 3.4, that BOLD activity corresponds to the integral activity of the network, which in turn correlates with the RT given in Eq. 18, thus he choice of the formula for RT seems reasonable.

Convergence and parameter choice

Here we want to show how changes in α and *noise* affect the overall behavior of the network. In order to do so we have ran simulations for mean activation $x(n)$ and $log10(RT)$ values calculated for - 1) constant *noise* (sampled from normal distribution of mean 0 and standard deviation 0.01) and variable α (2.0-2.4) and 2) constant α (2.0) and variable *noise* (sampled from normal distribution with mean 0 and standard deviation 0.01-0.05). The rest of the simulation parameters remain same as Table 3.1. The results are given in Fig. A1 and A2. It shows that how different interesting behavior emerges out of manipulation of *noise* and α. However, there is a lot of similarity in the qualitative behavior as well. So we have chosen $\alpha = 2.3$ and *noise* of standard deviation 0.03 in order to probe a wide range of behaviors while varying least number of parameters.

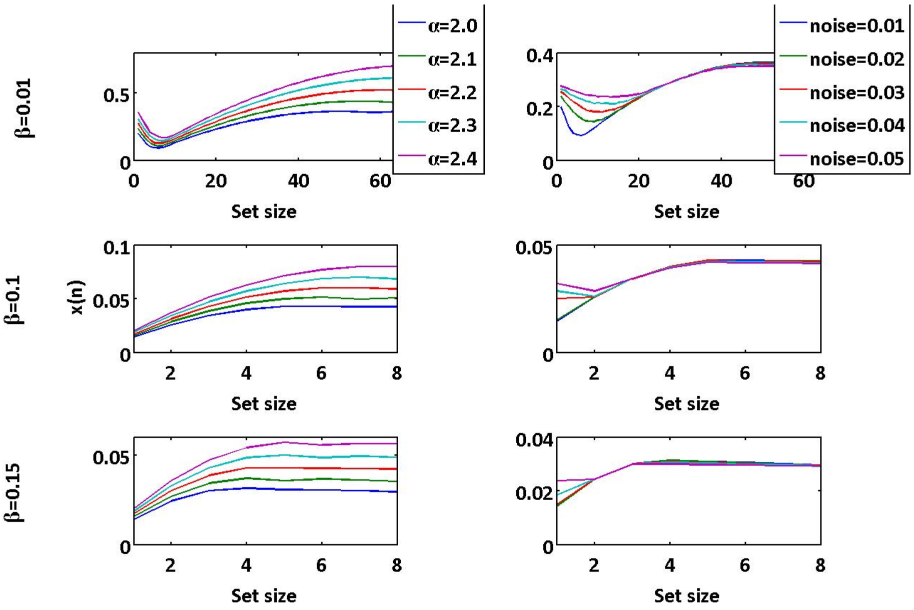


Figure A1: Here we show the mean activation ($x(n)$) pattern plotted for three β values (0.01, 0.1, and 0.15) and their variation with changing α and *noise*. The left column shows variation of $x(n)$ with α at constant level of *noise* (standard deviation 0.01) and the right column shows the variation of $x(n)$ at constant $\alpha = 2.0$ at different noise levels.

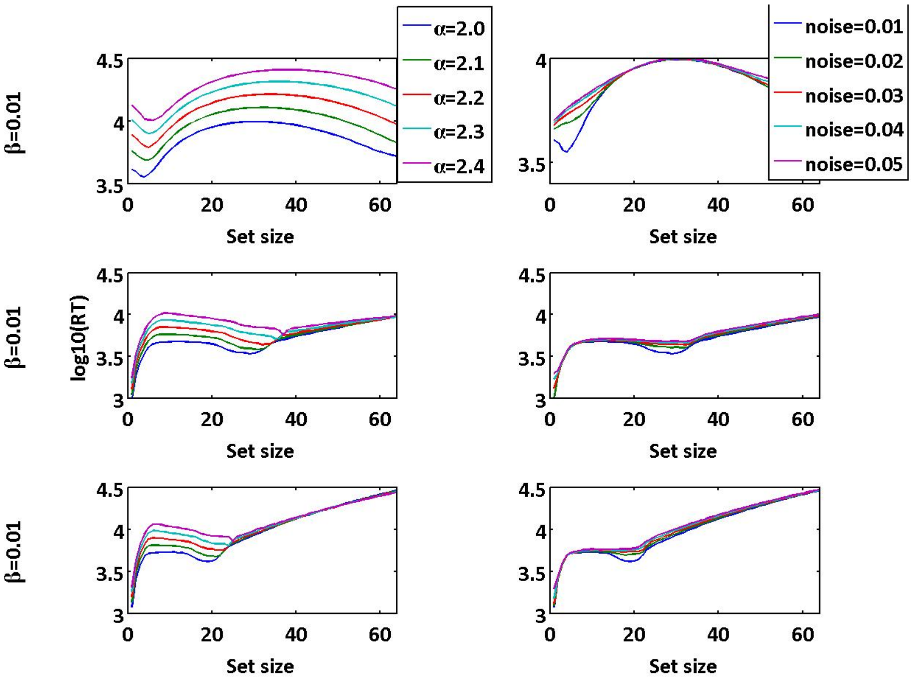


Figure A2: Here we show the $log10(RT)$ pattern plotted for three β values (0.01, 0.1, and 0.15) and their variation with changing α and $noise$. The left column shows variation of $log10(RT)$ with α at constant level of $noise$ (standard deviation 0.01) and the right column shows the variation of $x(n)$ at constant $\alpha = 2.0$ at different noise levels.

Example code

```
function params=numerosity_comparison

%clear all
%clc
params=setParams;

params.setsize=1:params.maxset;
ref1=16;
ref2=32;
ref3=1;
ref4=2;
ref5=3;
ref6=4;
ref7=5;
params.meanact=zeros(params.maxset,length(params.beta));
params.varact=zeros(params.maxset,length(params.beta));
params.prob_faith=zeros(params.maxset,length(params.beta));
params.var_faith=zeros(params.maxset,length(params.beta));
params.stableno=zeros(params.maxset,length(params.beta));
params.energy_mat=zeros(params.maxset,length(params.beta));



params.beta1=params.beta;
%params.beta1=[0.05 0.07 0.1 0.11 0.12];
%params.beta1=[0.01 0.015 0.02];
params.prob_ratio_ref1=zeros(length(params.beta1),length(params.setsize));
params.prob_ratio2_ref1=zeros(length(params.beta1),length(params.setsize));
params.prob_ratio_ref2=zeros(length(params.beta1),length(params.setsize));
params.prob_ratio2_ref2=zeros(length(params.beta1),length(params.setsize));
for k=1:length(params.beta1)
    params.mean_act=zeros(params.nrun,length(params.setsize));

    for j=1:length(params.setsize)
        testact=zeros(1,params.nrun);
```

```
        prob_faithfulness=zeros(1,params.nrun);
        for irun=1:params.nrun
            x=zeros(params.N,params.nt);
            dx=zeros(params.N,params.nt);
            en=zeros(params.N,params.nt);
            %gain=zeros(params.N,params.nt);
            %gainsum=zeros(N,1);
            alph=params.alpha*eye(params.N)+params.beta1(k)*(eye(params.N)-ones(params.N));
            a=randperm(params.N);
            b=a(:,1:params.setsize(j));
            I=zeros(params.N,params.nt);
            for kk=1:size(b,2)
                I(b(kk),:)=0.33*[ones(1,params.nstim) zeros(1,params.nt-params.nstim)];
            end
            noise=params.noise*randn(params.N,params.nt);
            for it=1:params.nt-1
            x((x(:,it)<=0),it)=0;
            %x(:,it+1)=0.99*x(:,it)+0.01*(alph*sumfx(x(:,it))+I(:,it))+noise(:,it);
            %dx(:,it+1)=0.01*(-x(:,it)+alph*sumfx(x(:,it))+I(:,it))+noise(:,it);
            x(:,it+1)=(alph*sumfx(x(:,it))+I(:,it))+noise(:,it);
            dx(:,it+1)=(-x(:,it)+alph*sumfx(x(:,it))+I(:,it))+noise(:,it);
            en(:,it+1)=en(:,it)-(1-params.alpha*(sumfx(x(:,it+1))./x(:,it+1)).^2).*(x(:,it+1).^2);
%            x(:,it+1)=alph*sumfx(x(:,it))+I(:,it)+noise(:,it);
            end
            x((x(:,params.nt)<=0),params.nt)=0;
            params.mean_act(irun,j)=mean(x(:,params.nt));
            testact(irun)=mean(x(:,params.nt));

            input_pattern=I(:,1);
            output_pattern=x(:,params.nt);
%             assignin('base','a',input_pattern);
            output_pattern(find(output_pattern>0))=1;
%             assignin('base','b',output_pattern);
            prob_faithfulness(irun)=pdist2(input_pattern',output_pattern','hamming');
```

```
            stableno(irun)=sum(output_pattern);
            energy(irun)=sum(sum(en));

        end
        params.meanact(j,k)=mean(testact);
        params.varact(j,k)=std(testact)/sqrt(params.nrun);
        params.prob_faith(j,k)=mean(1-prob_faithfulness);
        params.var_faith(j,k)=std(prob_faithfulness)/sqrt(
           params.nrun);
        params.stableno(j,k)=mean(stableno);
        params.energy_mat(j,k)=mean(energy);
    end

    params.all_meanact{1,k}=params.mean_act;

    for i=1:length(params.setsize)
        %if(i~=ref)

        b=params.mean_act(:,i)-mean(params.mean_act(:,ref1)
           );
        c=params.mean_act(:,i)-mean(params.mean_act(:,ref2)
           );
        d=params.mean_act(:,i)-mean(params.mean_act(:,ref3)
           );
        e=params.mean_act(:,i)-mean(params.mean_act(:,ref4)
           );
        f=params.mean_act(:,i)-mean(params.mean_act(:,ref5)
           );
        g=params.mean_act(:,i)-mean(params.mean_act(:,ref6)
           );
        h=params.mean_act(:,i)-mean(params.mean_act(:,ref7)
           );

        params.prob_ref1(i)=mean(b>0.01);
        params.prob2_ref1(i)=mean(abs(b)>0.01);
        params.prob_ref2(i)=mean(c>0.01);
        params.prob2_ref2(i)=mean(abs(c)>0.01);
        params.prob_ref3(i)=mean(d>0.01);
        params.prob2_ref3(i)=mean(abs(d)>0.01);
        params.prob_ref4(i)=mean(e>0.01);
        params.prob2_ref4(i)=mean(abs(e)>0.01);
        params.prob_ref5(i)=mean(f>0.01);
        params.prob2_ref5(i)=mean(abs(f)>0.01);
```

```
            params.prob_ref6(i)=mean(g>0.01);
            params.prob2_ref6(i)=mean(abs(g)>0.01);
            params.prob_ref7(i)=mean(h>0.01);
            params.prob2_ref7(i)=mean(abs(h)>0.01);

            %end
        end
        %params.prob(ref)=[];
        params.prob_ratio_ref1(k,:)=params.prob_ref1;
        params.prob_ratio2_ref1(k,:)=params.prob2_ref1;
        params.prob_ratio_ref2(k,:)=params.prob_ref2;
        params.prob_ratio2_ref2(k,:)=params.prob2_ref2;
        params.prob_ratio_ref3(k,:)=params.prob_ref3;
        params.prob_ratio2_ref3(k,:)=params.prob2_ref3;
        params.prob_ratio_ref4(k,:)=params.prob_ref4;
        params.prob_ratio2_ref4(k,:)=params.prob2_ref4;
        params.prob_ratio_ref5(k,:)=params.prob_ref5;
        params.prob_ratio2_ref5(k,:)=params.prob2_ref5;
        params.prob_ratio_ref6(k,:)=params.prob_ref6;
        params.prob_ratio2_ref6(k,:)=params.prob2_ref6;
        params.prob_ratio_ref7(k,:)=params.prob_ref7;
        params.prob_ratio2_ref7(k,:)=params.prob2_ref7;
    k

    end

    params.rt=-params.energy_mat;
```

Appendix B: How embodied is time?

Introduction

> We live in time – it holds us and moulds us – but I've never felt I understood it very well. And I'm not referring to theories about how it bends and doubles back, or may exist elsewhere in parallel versions. No, I mean ordinary, everyday time, which clocks and watches assure us passes regularly: tick-tock, click-clock. Is there anything more plausible than a second hand? And yet it takes only the smallest pleasure or pain to teach us time's malleability. Some emotions speed it up, others slow it down; occasionally, it seems to go missing – until the eventual point when it really does go missing, never to return.
>
> *The Sense of an Ending*, Julian Barnes

It is a standard understanding that we live in time. In fact, the whole physical world as described in sciences is based on the idea of objective (not absolute) time. For centuries we have defined time ever so minutely, basing them on finer and finer event measurements (uncoiling springs to atomic clocks) that we do not even notice that we have made an inductive leap when it comes to time - we can measure time, so we experience time. In the current work I wish to critique this inductive leap and examine what it means to experience time. We are embodied and embedded cognitive agents, constrained by our body as well as in continuous interaction with our environment (mostly in an unconscious manner, e.g., Are you standing or sitting? Are you paying complete attention to each part of your body and posture? etc). So another way to ask the question of temporal experience would be - how embodied is time? I posit that experience of time spoken of in general literature is a linguistic construct, in that, the idea of experience of time overshadows the actual phenomenal contents of time perception. Moreover, time perception (either as perception of duration or perception of an instant of time) itself comes from a post-facto judgment of events. It has also been observed that the order of events in time can be altered to create an illusion of violation of causality itself. This points to the possibility that events are arranged in a temporal map that can be read off by higher cognitive substrates. In the current work we go on to explore the nature of such a map as it emerges from an embodied mind.

Our experience is malleable. We retain traces of emotions generated from a beautiful melody even long after the actual music stops. Sometimes, we experience the same event alongside another in the form of a déjà vu. We are also in a constant state of growth and maturation, with malleable beliefs and imports, so that we have a sense of a unified ego (cogito encased in flesh) that moves through time[1]. In all standard theories of cognition, time is taken as an a priori category of consciousness. It is even hard to imagine how one can talk about the conscious experience and cognition in general if one did not have an ontological commitment to the idea of a subjective experience of time not only as embodied (our cognition is constrained, regulated and dependent in the body) and embedded/situated beings, but also as biological entities (our development, growth, and maturity form a linear progression).

However, such an ontological commitment to an a priori category of time leads to some very acute problems, both philosophically as well as theoretically. In the following section, I would try to show the difficulties with the ontological commitment to the subjective experience of time. However, merely questioning subjective experience of time leaves us without a sound account of temporal

---

[1] Even though the present ego might not identify with the beliefs and import of the ego-construct

phenomena in cognition. The challenge lies in giving an account of temporal existence on the basis of some other a priori category without giving in to the trap of Cartesian ontology.

In Section Three, I re-examine conceptual frameworks within the embodied stance of cognition to locate spatial representations as a possible solution. Although spatial representation has been shown to serve as an explanation of temporal phenomena in cognition, we must also locate the possibility of relative autonomy of spatial representation from a linguistic one, as a completely linguistic representation leads inevitably to Cartesian ontology.

However, spatial representation only forms one part of the story. It still leaves the question of the substrate of such a representation. In Section Four, I posit ontological priority of events and show how a possibility of a phenomenological account of event might operate. If events coupled with spatial representations replace the idea of subjective experience of time, it leads to a very interesting consequence for the embodied stance and intentionality defined therein. In the last section, I discuss the future directions that follow from this new ontological commitment.

Anatomy of time

In this section we are concerned with the time that is subjective - the time we are nostalgic about, the time that we are afraid of losing, the time we spend, the idea of time on the basis of which we plan for future, organize our day, the time through which we claim we move through and mature, and so on - as opposed to objective time. In the current work we take a realist position on objective time, but anti-realist position on subjective time. Subjective time is mostly in the domain of language. We talk about our subjective time through the categories of language and thus the subjective time is subject to linguistic relativism [Nùñez, 2008]. We also make a distinction between subjective time and subjective experience of time[2].

---

[2] We are not dealing with the question of biological time - in terms of circadian rhythms, heart beats, pulses, and so on. Organisms have a way of biological time keeping, in that, there is a periodicity that the organism can rely on for its survival. However, here we are more

In the perceptual domain time enters as instant of time (occurrence of an event) and duration (gap between two events in question), i.e., as a relation between events. Subjective experience of time thus refers to the experience of the relation itself. Subjective time on the other hand is a construct of language that helps us build a narrative around events in our memory. In the following we will try to show the relation of events can appear as a map (similar to the maps that are common in discussions on visuo-spatial attention). We will now explore two examples within the literature of time perception to explore the two aspects of instant of time and duration of time.

Motor-sensory recalibration There have been several demonstrations of temporal asynchrony in the visual domain [Moutoussis and Zeki, 1997a,b]. More recently, Stetson et al. [2006] have shown that by adapting to a fixed delay between a keypress by a human participant followed by a sensation, they could introduce illusory reversals by negating the delay after adaptation. This interesting case of reversal points us to a direction where instants of time-events are not integrated immediately, but rather an illusion-specific activation in anterior cingulate/medial frontal cortex point to a malleable representation of time events being compared to a more rigid representations [Stetson et al., 2006].

Johnston and Nishida [2001] have shown that attempts to explain the perceptual anomaly with neural processing delays raises very important philosophical issues. They have shown that, if we assume what they refer to it as 'brain time' hypothesis (equating perceptual time to processing time), we will run into two kinds of problems. In the explicit form, the relative time of events get encoded in higher level 'metaneurons'. However, it appears that it simply substitutes higher level processes for the events themselves. In the implicit form of brain time hypothesis, then Johnston and Nishida [2001] explains, "If our perception of the time of an event is coded implicitly, as the time at which a perceptual state of the brain is established - tantamount to the time at which we become aware of the contents of the event - then the medium of temporal sensory experience is no longer physical, as it is for the other five senses, it is conscious experience itself". Rather we should try to find answers within common spatio-temporal processing

---

interested in the time that is taken as content of subjective experience.

of percepts themselves.

Temporal oddball Events have a subjective duration. It remains an open question whether duration is perceived directly, like a visual feature, or f it depends mainly on a comparison process. Numerous studies have shown that subjective time experience depends on low-level visual properties and also the attentional focus. In the oddball paradigm, Tse et al. [2004] reported that duration judgments for stimuli longer than 120 ms showed temporal expansion. In another work we used a computational model to determine whether a decision-based account of temporal judgments could account for temporal expansion. We used a single layer recurrent dynamic on-center off-surround network of fully connected nodes with self-excitation and lateral inhibition (based on Usher and Cohen [1999]) optimized for winner-take-all dynamics for duration judgments [Sengupta et al., 2014b]. The 'winner' node, out of the two that receive the inputs of different duration values, determines the duration judgment. One node received the habituated input (standard) and the other received novel input (oddball). We ran a simulation over a range of durations (from 30 ms to 1200 ms) in order to calculate the subjective expansion factor for these durations if they were used as standard duration for oddball trials. The simulation results closely match the pattern of experimental results collected by Tse et al. [2004], including the 120 ms cutoff for TSE. These findings suggest that the TSE effect might arise out of comparison process rather than perceived difference in time itself [Sengupta et al., 2014c].

It seems extremely plausible from above discussion that temporal processing of events follows a map of events with spatial representation[3]. Given the above, there are two ways that subjective experience of time is possible. Firstly, the map's representation is completely dependent upon linguistic representation and thus cannot be separated from the content of experience. Secondly, the relation of events as given by map is itself a content of experience. In the following we will try to show that on both counts subjective experience of time is impossible. In the next section I challenge the first point by locating a possibility of non-linguistic

---

[3] For instance the illusory reversal of events could also be accounted for within a temporal map following spatial on-center off-surround representation.

representation within the embodied paradigm.

Autonomy of space

In essence embodiment stance in cognitive science asserts a fairly simple but deep departure from the standard approaches that are commonly labeled as Cognitivism (i.e., both functionalism and connectionism along with other derivatives of physical symbol systems stance) - human cognition is located within a dynamic goal-directed action-oriented process involving both the body and the environment [Wilson, 2002]. Even if specialized cognitive functions are served by different parts of the brain, the involvement of body and environment amounts to much more than input-output circuitry. We ascribe meaning to our world based on the constraints posed by the body [Wilson and Clark, 2009] and our roles as being-in-the-world is regulated by the body [Dreyfus, 1990]. The embodied stance is far from uniform. Early proponents like Lakoff and Johnson [1980] propose a view where a metaphorical process deriving from the body and its structure is central to human cognition. Dreyfus [1990] derives a Heideggerian account from his consideration of AI and robotics, to propose an anti-representational process using the world as its model Dreyfus [2008]. Freeman [1995] proposes an intentional arc (following Merleau-Ponty, 2012) through which an organism has its cognitive dispositions (determined by corresponding attractor states) altered through its action in the environment. These three major viewpoints have their advantages and also some inherent distinctions. In the next few paragraphs I would like to point out conceptual distinctions between these positions and will attempt to arrive at the necessary conceptual apparatus needed to address the question of the temporal experience of human cognition.

Lakoff and Johnson [1980]'s account in some form does adhere to at least a moderate stance of the Sapir-Whorf hypothesis [Sapir, 1983; Whorf, 1956] for linguistic relativism (seen more explicitly in their later work, Lakoff and Nuñez, 2000). Metaphoric process within the unconscious can impinge upon the meaning formation because the linguistic categories determine our world view. To extend their argument, one has to assume at least a linguistic representation of the world,

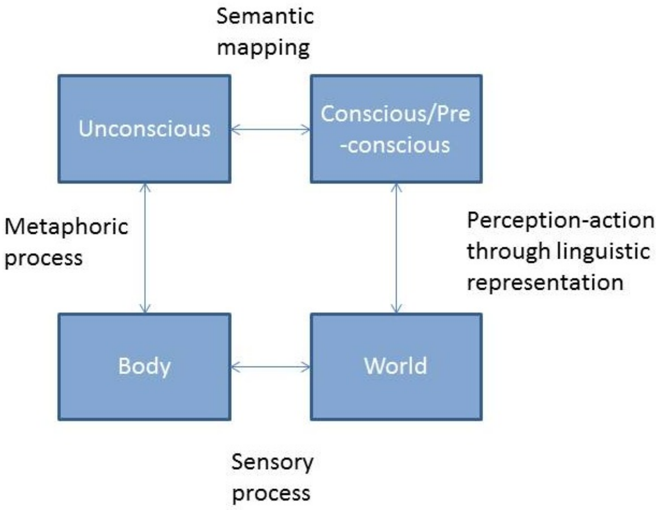


Figure A3: Embodied stance of Lakoff and Johnson [1980].

but one that is grounded in the body rather than being arbitrary like in the case of physical symbol systems hypothesis (PSSH, see Newell and Simon, 1976). Figure A3 elucidates the major import of their position on embodiment.

Dreyfus [1990] has rejected Cartesian ideas inherent in traditional AI and Cognitive science (i.e., the position that context-free, disembodied, representations governed by rule-like algorithms separated from and independent of the cognitive being and the world inhabited by it, are sufficient for building AI or understanding cognitive psychology). From his analysis of Frame problem in AI (writing a closed set of axioms for all given and novel situations), Dreyfus [2008] points out that the fundamental difficulty of a Cartesian ontology where one does not just see something, but sees it as something - there is always an extra step of deciding how to apply meaning to a given situation. On the other hand in the Heideggerian approach meaning is always ready-at-hand,

> To say a hammer has the function of being for hammering leaves out the defining relation of hammers to nails and other equipment, to the point of building things, and to our skills all of which Heidegger called readiness-to-hand and so attributing functions to brute facts couldn't

> capture the meaningful organization of the everyday world. -Dreyfus [2008]

In his account, embodiment exists in "being-in-the-world" relation between cognitive agent and the world, what Dreyfus [2008] describes as 'background coping'. He uses Gibson [1979]'s idea of affordances to explain how the world affords the actions available for the cognitive agent in a continuous and non-representational manner. In fact, he criticizes the idea of Cartesian self-enclosed entity itself, rather thinking of experience as a flow during 'everyday coping' [Dreyfus, 2000].

Although Freeman [1995]'s account is very similar to that of Dreyfus [2008], there are some key differences. His account derives from the works of Aquinas [2000] and Merleau-Ponty [2012]. The key concepts here are unity (by the mind's assertion of a boundary against the outside) and an intentional relation (between intellect and material objects) which shapes itself by changing its scales to accommodate the object through the process of knowing it. Intentionality as defined here is not "about-ness" [Searle, 1992]. Meaning is achieved through action in the world, and the self is altered by that action. Although [Aquinas, 2000] describes intentionality as directing action towards some future defined and chosen by the actor, intentionality is fundamentally an unconscious process, i.e., the self is not necessarily a conscious agent. Merleau-Ponty [2012]'s formulation of "the intentional arc" that completes an organic loop from action through the world and back into the brain, resembles Aquinas' active intellect. The intentional arc names the tight connection between the agent and the world, perception and action. The skills acquired by the agent are not stored as representations but as dispositions to respond to directing situations in the world. Thus for Freeman [1995], the term intentionality means the "process of a brain in action having the properties of unity, wholeness, and intent." Unity refers to a state of integration by which a self distinguishes itself from non-self. Wholeness refers to a bounded process by which through stages a self actualizes its mature form, ultimately to die. Intent refers to a relation in which a self modifies itself in conformance with aspects of non-self. A living brain thus capable of actualizing these properties by purposive behavior is an intentional structure.

From the above discussion it emerges that we can distinguish the three aforementioned strands within the embodied cognition literature along three dimen-

	Representation	Idea of self	Basis of cognition
Lakoff & Johnson	Linguistic	Through metaphoric processes connecting the body and the world	Unconscious metaphoric process
Dreyfus	Non-representational account	Not self-enclosed, thus even unaware of self during ‘everyday coping’	“Being-in-the-world” relation between cognitive agent and the world
Freeman	Ordered dispositions act as surrogates to representation	Through unity and intent self distinguishes from and is modified by non-self	Intentionality rooted in the unconscious mind

Figure A4: The relation between different strands of embodied stance

sions: representation, idea of self, and basis of cognition. The differences are summarized in Figure A4.

The above table makes it explicit that there are some major irreconcilable differences within the three strands of embodiment. Lakoff and Johnson’s idea of linguistic representation is incompatible with the others, but their idea of an unconscious process as a basis of cognition is fairly in line with Freeman’s account. However, it is unclear how much the metaphoric process can be reconciled with the intentional arc in the perceptual domain. On the other hand, Dreyfus’s idea of “being-in-the-world” relation is not compatible with Freeman’s notion of self that distinguishes itself from non-self through unity.

In both Dreyfus and Freeman’s account representation is disposed off almost completely. However, we know that there is some validity to the idea of representation from researches in retinotopic and topographic maps in the brain. It is not confined just to the visual domain either [Chen et al., 2013]. The main problem is that thinking of representation in a linguistic manner guarantees a Cartesian ontology. However, any idea of a representation that is completely non-linguistic is doomed to fall flat in the case of human cognition. As shown by Dreyfus [1990], on-line cognition can be traced to a state of flow (being-in-the-world) that does not require any mediating representation. However, we do employ quite a lot of off-line cognition that do definitely rely on representation. These representations may be derived from the body as some views of embodiment literature hold [Wilson, 2002]. Even so, it does not discount the possibility of a representational matrix (as seen in saliency map literature by Melcher and Piazza, 2011; Roggeman et al., 2010 or in literature on topographic maps, like Seelke et al., 2012) which is essentially a spatial one.

The idea of a spatial saliency map representation underlying perception and action is not limited to just vision, but involves auditory [Kayser et al., 2005] and tactile [Diederich et al., 2003] modalities as well. These saliency maps are part of pre-attentive visual stream and may not be cognitively penetrable [Pylyshyn, 1999]. A subsequent decoding of a saliency map requires linguistic features of cognition. Moreover, Bonato et al. [2012] have recently summarized an array of literature that points to various similarities between spatial and temporal processing and suggest that space, time, and number processing might be grounded in neuronal structures for spatial attention and sensori-motor processing. Spatial numerical abilities in other animals [Dehaene and Changeux, 1993] suggests that spatial processing might operate at least partly in an independent manner from linguistic ability. Given the above we can make a case for spatial representations that are grounded in sensori-motor process, i.e., embodied, and also can maintain relative autonomy from linguistic representations.

Ontology of the event

Phenomenology, either in the line of Heidegger or Merleau-Ponty, forms the theoretical kernel for much of the philosophy of mind in the embodied tradition. Thus, it comes as no surprise that the idea of a discreet mental event seems contrary to the embodied stance. The phenomenological idea is to explore how reality appears to consciousness. Consciousness is continuous. Then where will be the scope of something as discreet as an onset of an event, and what is the meaning of the order of events given in time? As Dastur [2000] explains

> As Maurice Merleau-Ponty shows in his Phenomenology of Perception ..., philosophy can give neither a realist nor an idealist solution to the problem of time. It does not succeed in locating it either in

> things themselves or in consciousness. If, on the one hand, we consider time to be no more than a dimension of reality, we can no longer explain the relationship between what comes first and what follows. The succession of events can only be established by consciousness, a consciousness which requires, in order to have a general view of the succession of events, not to be completely immersed in time. But what if, on the other hand, we consider time to be a mere construction of consciousness? Temporality itself becomes incomprehensible, insofar as it is the essence of time to be incompletely present to consciousness, to remain incompletely constituted, as Husserl would say. For time, precisely, is not identical to being, it is a process which is always in becoming. It is always of the order of the process, the passage, and that which comes. Therefore realism (which immerses the subject in time to the point of destroying all possibility of a time-consciousness) and idealism (which places consciousness in a position of overviewing a time which no longer proceeds), are both unable to clarify what they pretend to explain, that is, the relation of consciousness to time. For in both cases, what remains out of range for a philosophical inquiry which wants to see in time either a reality or an idea is precisely its transitional character, its non-being or non-essence, which is not, but proceeds.
>
> Dastur [2000]

The problem of temporality rather appears in giving time a priori ontological basis in consciousness. Rather, we should try to understand the phenomenal character of the event and proceed from there. We propose here an ontological priority of the event - in that it is not that event is something that breaks the continuous flow of consciousness through time as a surprise[4], rather we should

---

[4] "The event constitutes the "dehiscence" of time, its coming out of itself in different directions, which Heidegger calls "ekstasis," the fact that it never coincides with itself, and which Levinas names diachrony ... For the event, as such, is upsetting. It does not integrate itself as a specific moment in the flow of time. It changes drastically the whole style of an existence ... It does not happen in a world–it is, on the contrary, as if a new world opens up through its happening. The event constitutes the critical moment of temporality–a critical moment which nevertheless allows the continuity of time."-Dastur [2000]

understand how a limited consciousness would construct the flow of events post-priori due to its inability to cope with them all at once. Attending to many events is expensive for the organism in terms of energy. To deal with multitude of events we need coherence, repression of events, we need order - we construct time. This is inexorably linked to how the organism forms an idea of self.

In autism, subjects display "pattern of impairment in cognitive tasks that demand contextual processing, coupled with ... superiority at tasks that demand piecemeal processing of individual features" [Belmonte, 2008]. Thus for the autistic, the surrounding is 'threateningly intractable'. Thus they exhibit repetitive behavior akin to what Belmonte [2008] calls re-viewing the 'Cartesian cinema', to come to terms with their selves and surrounding. However, one can also look at this as multiple events as constituting a single instant of integrated time for the self.

The other side of the spectrum appears to be Schizophrenia where a decoherence of internal integration leads to hallucinations and delusions and a host of other symptoms. Schizophrenia presents an unique problem for embodiment - how can the same embodied being embedded in a single environment have different minds so to say. To put it better, we can ask how an embodied being can fail to integrate its own temporal existence. The failure of integration would not arise if time is more fundamental to embodied existence than an event is. Rather we propose here that events are prior to an embodied being. The flow of consciousness itself might arise due to a finite organism's strategy to keep track of multiple events.

A serious objection to the above position can be raised from the perspective of Dreyfus - how does the primacy of events allow us to explain on-line cognitive activity like hammering the nail example. The question hinges upon the assumption that, on-line cognition by virtue of its unconscious fluidity negates the ontological priority of events that enter consciousness. However, there is no reason to think so. The unconscious dynamical processes are driving forces of cognition, but conscious experience may get its structure from events. Both are not mutually incompatible.

Magnitude of intentionality

> As an alternative to this understanding of volition, I want to describe a neural basis for goal-directed actions that is common to both humans and other animals, because it reflects the evolution of human mechanisms from simpler animals, in which intent can operate without will. The concept - "intentionality" - was first described by Thomas Aquinas in 1272 to denote the process by which humans and other animals act in accordance with their own growth and maturation. An intent is the directing of an action toward some future goal that is defined and chosen by the actor. It differs from a motive, which is the reason and explanation of the action, and from a desire, which is the awareness and experience stemming from the intent. A man shoots another with the intent to kill, which is separate from why he does it and with what feeling.
>
> Freeman [2000]

In light of the above discussion it is important to address some very important issues. Our biological growth and maturity is linear, so is the goal-directed intentional arc. In other words the intentional arc that doubles back to the individual, is forward seeking. If we have an event-based spatial ontology to replace time, we still leave out another possibility. We experience linearity in time because things happen linearly. So, in spite of all my previous arguments, a forward seeking, goal-directed intentional arc can still experience time relation because the world affords linearity in cognition.

I have a series of interconnected arguments against the possibility.

1. As far as linearity in thinking goes, the idea of linearity itself is subservient to the recognition of simultaneity and order (event A and event B occurred together, but prior to event C, and so on), which themselves possibly have spatial representation and thus are subject to the limits of capacity.

2. The limits of capacity appear only in the case of conscious recognition of this simultaneity and order (i.e., operating on them with working memory).

However, it does not preclude the very large storage of patterns in long term memory.

3. With regard to ideas of biological linear progression, I think evolution is as parallel as it is linear (one can even say probably more parallel than linear). Evolution operates in jumps and parallel mutations that accumulate for an environment, which itself undergoes drastic change to halt the progression and lead through another parallel branch (consider the reptilian to mammalian evolution).

4. Regarding the idea of change from life to death and of physical time itself, we know that Newtonian and Quantum physics are both time-symmetric (i.e., they do not distinguish between t and -t). Only when we come to macroscopic results (in the realm of statistical physics and thermodynamics) do we get time as the direction of increase of entropy in a closed system. However, there are some recent works that lean towards the possibility that a recorder/measurement itself is involved in giving the arrow of time.

Keeping in mind the above, there is a need to probably let go of the idea of non-reducible intentionality. We must reduce intentionality at least at a theoretical level if not at the entity level. We need to operationalize intentionality in terms of other *a priori* ontological categories in order to get the full blooded idea of intentionality promised by Freeman.

The future of embodiment stance will lie in the way we use the tools available to us to get a much better idea of human cognition. For instance, categorical off-line thinking has been a challenge to embodiment paradigm. However categorical thinking or conceptual thought depends upon the idea of sets, sets depend upon the notion of numbers. We have made progress in establishing numbers on the foundation of spatial representation [Grossberg and Repin, 2003; Sengupta et al., 2014b]. The other steps are sure to follow to give us a full fledged "dynamical systems" ontology of mind.

List of publications

Publications

Journal Articles

Rakesh Sengupta, Bapi Raju Surampudi, and David Melcher (2014), “A visual sense of number emerges from the dynamics of a recurrent on-center off-surround neural network”, *Brain Research*, 1582: 114-124

Andr Knops, Manuela Piazza, Rakesh Sengupta, Evelyn Eger, and David Melcher (2014) “A shared, flexible neural map architecture reflects capacity limits in both visual short term memory and enumeration”, *Journal of Neuroscience* 34(30): 9857-9866.

Rakesh Sengupta, Bapi Raju Surampudi, and David Melcher (2014), “Big and small numbers: empirical support for a single, flexible mechanism for human numerosity perception”. *Attention, Perception, & Psychophysics* In revision.

Rakesh Sengupta, Bapi Raju Surampudi, Prajit Basu and David Melcher (2014), “The influence of spatial cueing on serial order visual memory”, *Manuscript in Preparation*

Conference presentations: oral

Rakesh Sengupta, Bapi Raju Surampudi, Prajit Basu and David Melcher (March, 2014), “Spatial cueing and serial recall”, *Annual conference on Cognitive Science,*

Delhi, India
Rakesh Sengupta (February, 2014), “How embodied is time?”, *National Seminar on Embodied Cognition, Hyderabad, India*

Conference presentations: poster

Rakesh Sengupta, Bapi Raju Surampudi, Prajit Basu and David Melcher (May, 2014), “Accounting for subjective time expansion based on a decision, rather than perceptual, mechanism”, *Vision Science Society Annual Conference, Florida, USA*
Rakesh Sengupta, Bapi Raju Surampudi, and David Melcher, (May, 2013), “Subitizing and estimation emerge from a computational saliency map model”, *Vision Science Society Annual Conference, Florida, USA*
Rakesh Sengupta and Bapi Raju Surampudi (December, 2011), “Saliency map and spatial attention”, *Federation of Asian and Oceanian Neuroscience Societies, Lucknow, India*

References

Abramov, I., Gordon, J., Feldman, O., and Chavarga, A. (2012). Sex and vision ii: color appearance of monochromatic lights. *Biology of sex differences*, 3(1):1–15. 9

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15:122–31. 24

Amiez, C. and Petrides, M. (2007). Selective involvement of the mid-dorsolateral prefrontal cortex in the coding of the serial order of visual stimuli in working memory. *Proceedings of the National Academy of Sciences*, 104:13786–13791. 31

Aquinas, S. T. (2000). *The "Summa Theologica"*. Hayes Barton Press. 19, 111, 133

Ariely, D. (2001). Seeing sets: representation by statistical properties. *Psychological Science*, 12:157–162. 78

Aristotle (2007). *The Nicomachean Ethics*. Filiquarian Publishing. 2

Austin, J. L. (1970). *Philosophical papers*. Oxford University Press. 17

Awh, E. and Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, 5:119–126. 24, 28, 31, 79

Awh, E., Vogel, E. K., and Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience*, 139:201–208. 24, 30, 31

B B Murdock, J. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64(5):482. 31

Belmonte, M. (2008). Human, but more so: What the autistic brain tells us about the process of narrative. In *Autism and Representation*. New York: Routledge. 137

Bickhard, M. H. (2008). *Handbook of Cognitive Science: an Embodied approach*, chapter Is Embodiment Necessary. Elsevier. 14

Bisley, J. W. and Goldberg, M. E. (2006). Neural correlates of attention and distractibility in the lateral intraparietal area. *Journal of Neurophysiology*, 95:1696–1717. 28, 31

Bogacz, R., Usher, M., Zhang, J., and McClelland, J. L. (2007). Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362:1655–1670. 36

Bonato, M., Zorzi, M., and Umiltà, C. (2012). When time is space: evidence for a mental time line. *Neuroscience and Behavioral Reviews*, 36:2257–2273. 113, 135

Brooks, R. A. (1991). New approaches to robotics. *Science*, 253:1227–1232. 20

Brown, R. W. and Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology*, 49:454–462. 5

Buechner, J. (2008). *Gödel, Putnam, and Functionalism: A New Reading of Representation and Reality*. The MIT Press. 8

Buller, D. J. (2005). Get over: massive modularity. *Biology and Philosophy*, 20:881–891. 8

Burr, D. and Ross, J. (2008a). Response: Visual number. *Current Biology*, 18:R855–R856. 32

Burr, D. and Ross, J. (2008b). Visual sense of numbers. *Current Biology*, 18:425–428. 32, 78

Burr, D., Turi, M., and Anobile, G. (2010). Subitizing but not estimation of numerosity requires attentional resources. *Journal of Vision*, 10:1–10. 33, 43, 54, 78

Calvo, P. (2005). Rules, similarity, and the information-processing blind alley. *Behavioral and Brain Sciences*, 28:17–18. 11, 12

Chen, K., Zhou, B., Chen, S., He, S., and Zhou, W. (2013). Olfaction spontaneously highlights visual saliency map. *Proceedings of the Royal Society B: Biological Sciences*, 280(1768):20131729–20131729. 113, 134

Chomsky, N. (1975). *The Logical Structure of Linguistic Theory*. Springer. 17

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Jornal of Philosophy*, 78:67–90. 9

Churchland, P. M. (1989). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. The MIT Press. 9

Clark, A. (2008). *Handbook of cognition: an embodied approach*, chapter Embodiment and explanation, pages 41–58. Elsevier. 17

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., Linenweber, M. R., Peterson, S. E., Raichle, M. E., Van Essen, D. C., and Shulman, G. L. (1998). A common network of functional areas for attention and eye movements. *Neuron*, 21:761–773. 24, 31

Corbetta, M., Kicade, J. M., and Shulman, G. L. (2002). Neural systems for visual orienting and their relationships to spatial working memory. *Journal of Cognitive Neuroscience*, 14:508–523. 30, 31

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., and Peterson, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. *The Journal of Neuroscience*, 11:2383–2402. 91

Daston, L. J. and Galison, P. L. (2010). *Objectivity*. Zone Books. 4

Dastur, F. (2000). Phenomenology of the event: Waiting and surprise. *Hypatia*, 15:178–189. 135, 136

Davidson, D. (1970). *Experience and Theory*, chapter Mental Events. Humanities Press. 5

Dehaene, S. and Changeux, J. P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, 5:390–407. 33, 34, 45, 54, 55, 79, 106, 108, 113, 135

Dennett, D. (1988). Quining qualia. In *Consciousness in modern science*. Oxford University Press. 9

Dennett, D. (1996). *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, chapter Cognitive Wheels: The Frame Problem in Artificial Intelligence. Greenwood Publishing Group. 12

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences*, 93:13494–13499. 24

Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1):193–222. 26, 30

Deubel, H., Schneider, W., and Paprotta, I. (1998). Selective dorsal and ventral processing: Evidence for a common attentional mechanism in reaching and perception. *Visual Cognition*, 5(1-2):81–107. 26

Diederich, A., Colonius, H., Bockhorst, D., and Tabeling, S. (2003). Visual-tactile spatial interaction in saccade generation. *Experimental Brain Research*, 148(3):328–337. 113, 135

Dreyfus, H. (1990). *Being in the World: A Commentary on Heidegger's Being and Time*. The MIT Press. 109, 110, 113, 131, 132, 134

Dreyfus, H. (2000). Telepistemology: Descartes's last stand. In *The robot in the garden*. MIT Press. 111, 133

Dreyfus, H. (2008). Why heideggerian ai failed and how fixing it would require making it more heideggerian. *Artificial Intelligence*, 171:1137–1160. 13, 14, 110, 111, 131, 132, 133

Eccles, J. C. (1993). *Rethinking Neural Networks: Quantum Fields and Biological Data*, chapter Evolution of complexity of the brain with the emergence of consciousness. Lawrence Erlbaum. 6

Elman, J. L. (1998). Generalization, simple recurrent networks, and the emergence of structure. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*. 11

Farah, M. (1990). *Visual agnosia*. MIT press Cambridge, MA. 26

Feigenson, L., Dehane, S., and Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8:307–314. 78

Fellaman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate visual cortex. *Cerebral Cortex*, 1:1–47. 28

Fernandez-Duque, D. and Thornton, I. M. (2000). Change detection without awareness: Do explicit reports underestimate the representation of change in the visual system? *Visual Cognition*, 7:323–344. 31

Fodor, J. A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. The MIT Press. 8

Fodor, J. A. (1987). *Psychosemantics*. The MIT Press. 17

Fodor, J. A. (2000). *The mind doesn't work that way*. The MIT Press. 13

Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28:3–71. 11

Franconeri, S., Alvarez, G., and Cavanagh, P. (2013). Flexible cognitive resources: competitive content maps for attention and memory. *Trends in Cognitive Sciences*, 17:134–141. 78

Freeman, W. J. (1995). *Societies of brains*. Lawrence Erlbaum. 19, 110, 111, 112, 131, 133

Freeman, W. J. (2000). *How brains make up their minds*. Columbia University Press. 14, 138

Freud, S. (1895). *Project for a scientific psychology.* The Hogarth Press. 4, 9

Gallistel, C. R. and Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44(1):43–74. 33, 78

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception.* Lawrence Erlbaum. 19, 111, 133

Gilbert, S. J. and Shallice, T. (2002). Task switching: a PDP model. *Cognitive Psychology*, 44(3):297–337. 107

Gödel, K. (2002). On formally undecidable propositions of principia mathematica and related systems:i. In *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, pages 592–617. Harvard University Press. 8

Gordon, P. (2004). Numerical cognition without words: Evidence from amazonia. *Science*, 306:496–99. 33

Gottlieb, J. and Goldberg, M. E. (1999). Activity of neurons in the lateral intraparietal area of the monkey during an antisaccade task. *Nature Neuroscience*, 2:906–912. 55, 78

Gregory, R. L. (1998). *Eye and Brain.* Oxford University Press. 28

Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, LII:213–257. 37, 53

Grossberg, S. and Repin, D. V. (2003). A neural model of how the brain represents and compares multi-digit numbers: spatial and categorical processes. *Neural Networks*, 16(8):1107–1140. 34, 54, 55, 79, 106, 139

Hadamard, J. (1945). *The psychology of invention in the mathematical field.* Dover Publication. 16

Halberda, J. and Feigenson, L. (2008). Set representations required for the acquisition of the “natural number” concept. *Behavioral and Brain Sciiences*, 31:655–656. 77

Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42:335–346. 12

Harnad, S. (2005). To cognize is to categorize: Cognition is categorization. In Cohen, H. and Lefebvre, C., editors, *Handbook of Categorization in Cognitive Science*, pages 20–44. Elsevier. 5, 24

Hauser, M. D., Tsao, F., Garcia, P., and Spelke, E. S. (2003). Evolutionary foundations of number: Spontaneous representation of numerical magnitudes by cotton-top tamarins. *Proceedings of Biological Science*, 270:1441–1446. 33

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological approach*. John Wiley & Sons. 9

Hegel, G. W. F. (1988). *Phenomenology Of Spirit*. Motilal Banarsidass Publ. 3

Heisenberg, M. and Wolf, R. (1988). Reafferent control of optomotor yaw torque in drosophilia melanogaster. *Journal of Comparative Physiology*, 163:373–388. 21

Hellström, A. (1985). The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin*, 97:35–61. 35, 101

Holway, A. H. and Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *American Journal of Psychology*, 54:21–37. 24

Hommel, B., Müsseler, J., Aschersleben, G., and Prinz, W. (2001). The theory of event coding (tec): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24:849–937. 12

Hubel, D. H. and Wiesel, T. N. (1970). Stereopsis vision in the macaque monkey. *Nature*, 225:41–42. 28

Hurvich, L. M. and Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review*, 64:384–404. 5

Hyvärinen, J. (2012). *The Parietal Cortex of Monkey and Man*. Springer London. 24

Itti, L. and Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2:194–203. viii, 26, 27

James, W. (1950). *Principles of psychology.* Dover Publication. 4

Johnston, A. and Nishida, S. (2001). Time perception: Brain time or event time? *Current Biology*, 11(11):R427–R430. 129

Kafka, F. (2008). *Metamorphosis.* Aventura Press. 3

Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology*, 24:174–219. 24

Katz, J. T. and Fodor, J. A. (1963). The structure of a semantic theory. *Language*, 39:170–210. 17

Kaufman, E. L., Lord, M. W., Reese, T., and Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology*, 62:498–525. 33, 43, 50, 54, 77, 85

Kawasaki, M., Watanabe, M., Okuda, J., Sakagami, M., and Aihara, K. (2008). Human posterior parietal cortex maintains color, shape and motion in visual short-term memory. *Brain Research*, 1213:91–97. 48, 79

Kayser, C., Petkov, C. I., Lippert, M., and Logothetis, N. K. (2005). Mechanisms for allocating auditory attention: An auditory saliency map. *Current Biology*, 15(21):1943–1947. 113, 135

Knops, A., Piazza, M., Sengupta, R., Eger, E., and Melcher, D. (2014). A shared, flexible neural map architecture reflects capacity limits in both visual short term memory and enumeration. *Journal of Neuroscience*, 34(30):9857–9866. 78, 79

Koestler, A. (1964). *The act of creation.* Macmillan. 16

Koffka, K. (1935). *Principles of gestalt psychology.* Lund Humphries. 4

Köhler, W. (1920/1939). *Psycsical gestalten.* Routledge & Kegan. 20

Kripke, S. A. (1980). *Naming and necessity*. Harvard University Press. 6

Lakoff, G. and Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press. xiv, 9, 16, 109, 110, 111, 131, 132

Lakoff, G. and Nuňez, R. (2000). *Where mathematics comes from: how the embodied mind brings mathematics into being*. Basic Books. 16, 110, 131

Lewontin, R. C. (1991). *Biology as ideology*. Harper Perennial. 18

Logan, G. D. (2004). Cumulative progress in formal theories of attention. *Annual Review of Psychology*, 55:207–234. 28

Marcus, G. F., Vijayan, S., Rao, S. B., and Vishton, P. M. (1999). Rule learning in seven-month-old infants. *Science*, 283:77–80. 11

Marr, D. (1976). Early processing of visual processing. *Phil. Tr. R. Soc. Lond. B*, 275:485–526. 24

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. The MIT Press. 24

Marx, K. (1970). *The German Ideology, Part One*. International Publishers. 3

Marx, K. (2007). *Capital: A Critique of Political Economy - Vol. I*. Cosimo, Inc. 2

McCandliss, B. D., Cohen, L., and Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, 7:293–299. 24

McCulloch, W. S. and Pitts, W. H. (1943). A logical calculus of ideas immanent in nervous activity. bulletin of mathematical biophysics. *Bulletin of Mathematical Biophysics*, 5:115–133. 9

Melcher, D. and Piazza, M. (2011). The role of attentional priority and saliency in determining capacity limits in enumeration and visual working memory. *PLoS one*, 6:1–11. 31, 32, 33, 78, 108, 113, 135

Merleau-Ponty, M. (2012). *Phenomenology of perception*. Routledge. 19, 110, 111, 112, 131, 133

Miller, A. G. (1956). The magical number seven, plus or minus two: Some limits on ourcapacity for processing information. *Psychological Review*, 63:81–97. 31, 57, 92

Milner, D. and Goodale, M. (1995). *The visual brain in action*. Oxford University Press. 26

Milner, D. and Goodale, M. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46(3):774–785. 26

Minsky, M. L. and Papert, S. (1988). *Perceptrons: An introduction to computational geometry*. The MIT Press. 9

Modell, A. H. (2003). *Imagination and the meaningful brain*. The MIT Press. 17

Mohamed, M., Yousem, D. M., Tekes, A., Browner, N., and Calhoun, V. D. (2004). Correlation between the amplitude of cortical activation and reaction time: A functional mri study. *American Journal of Roentgenology*, 183(3):759–765. 118

Moore, A. and Cambell, E. (2006). *From Hell*. Top Shelf Productions. 3

Moutoussis, K. and Zeki, S. (1997a). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society B: Biological Sciences*, 264(1380):393–399. 129

Moutoussis, K. and Zeki, S. (1997b). Functional segregation and temporal hierarchy of the visual perceptive systems. *Proceedings of the Royal Society B: Biological Sciences*, 264(1387):1407–1414. 129

Müller, H. J. and Rabbitt, P. M. A. (1989). Spatial cueing and the relation between the accuracy of “where” and “what” decisions in visual search. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 41:747–773. 31

Nagel, T. (1974). What is it like to be a bat? *Psychological Review*, 83:435–450. 6

Neisser, U. (1967). *Cognitive Psychology*. Appleton Century Crofts. 28

Newell, A. and Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 19(3):113–126. 110, 132

Nieder, A., Freedman, D. J., and Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science*, 297:1708–1711. 34

Nöe, A. (2004). *Action in perception*. The MIT Press. 24

Nùñez, R. E. (2008). Mathematics, the ultimate challenge to embodiment: Truth and the grounding of axiomatic systems. In *Cambridge Handbook of Metaphor and Thought . Cambridge*. MA: Cambridge University Press. 128

Ono, H. (1966). Distal and proximal size under reduced and non-reduced viewing conditions. *American Journal of Psychology*, 79:234–241. 24

Pavlov, I. P. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. Oxford University Press. 4

Penrose, R. (1994). *Shadows of the Mind*. Oxford University Press. 8

Piaget, J. (1954). *The Construction of Reality in the Child*. Basic Books. 14

Piazza, M., Fumarola, A., Chinello, A., and Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition*, 121:147–53. 32, 33, 77, 78

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, 44(3):547–555. 45, 53, 79

Piazza, M., Pinel, P., Le Bihan, D., and Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, 53:293–305. 77, 79

Piccolino, M. and Wade, N. J. (2008). Galileo galilei's vision of the senses. *Trends in Neurosciences*, 31:585–590. 24

Pinker, S. and Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6:456–463. 11

Poggio, G. and Poggio, T. (1984). The analysis of stereopsis. *Annual Review of Neuroscience*, 7:379–412. 28

Poggio, T. and Reichhardt, W. (1976). Visual control of orientation behaviour in the fly. *Quarterly Journal of Biophysics*, 9:377–438. 20

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32:3–25. 28, 31

Putnam, H. (1991). *Representation and Reality*. The MIT Press. 8, 12

Pylyshyn, Z. (1999). Is vision continuous with cognition? the case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22:341–423. 24, 113, 135

Pylyshyn, Z. W. (1987). *The robot's dilemma: the frame problem in artificial intelligence*. Ablex. 13

Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology*, 45:45–94. 11

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., and Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science*, 19:607–14. 33, 54, 78, 108

Robertson, L. (2003). Binding, spatial attention and perceptual awareness. *Nature Reviews Neuroscience*, 4(2):93–102. 26, 28, 30

Robertson, L., Treisman, A., Friedman-Hill, S., and Grabowecky, M. (1997). The interaction of spatial and object pathways: Evidence from balint's syndrome. *Journal of Cognitive Neuroscience*, 9(3):295–317. 30

Roggeman, C., Fias, W., and Verguts, T. (2010). Salience maps in parietal cortex: Imaging and computational modeling. *NeuroImage*, 52:1005–1014. 31, 32, 36, 38, 41, 48, 53, 55, 79, 108, 113, 135

Roitman, J. D., Brannon, E. M., and Platt, M. L. (2007). Monotonic coding of numerosity in macaque lateral intraparietal area. *PLos Biology*, 5:1673–1682. 34, 50, 79

Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms.* Spartan Books. 9

Ross, W. D. (1931). *The works of Aristotle, Vol. 3.* Clarenden, Oxford. 23

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). *Parallel Distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations*, chapter Learning internal representations by error propagation, pages 318–362. The MIT Press. 9

Salay, N. (2009). Why dreyfus frame problem argument cannot justify anti-representational ai. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society.* 14

Sapir, E. (1983). *Selected Writings of Edward Sapir in Language, Culture, and Personality.* University of California Press. 5, 110, 131

Schöner, G. (2008). *Cambridge handbook of computational psychology*, chapter Dynamical systems approaches to cognition, pages 101–126. Cambridge University Press. 20, 21

Schöner, G. and Kelso, J. A. S. (1988). Dynamical pattern generation in behavioral and neural systems. *Science*, 239:1513–1520. 21

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, 3:417–457. 8

Searle, J. R. (1992). *The Rediscovery of the mind.* The MIT Press. 5, 6, 7, 19, 111, 133

Seelke, A. M. H., Padberg, J. J., Disbrow, E., Purnell, S. M., Recanzone, G., and Krubitzer, L. (2012). Topographic maps within brodmann's area 5 of macaque monkeys. *Cerebral Cortex*, 22(8):1834–1850. 113, 135

Seidenberg, M. and Elman, J. L. (1999). Networks are not 'hidden rules'. *Trends in Cognitive Sciences*, 3:288–289. 11

Sengupta, R. (2013). Excommunication to cogito: Exploring hegelian dialectic. *Journal of Eurasian Studies*, V:35–45. Submitted to History and Theory. 2

Sengupta, R. (2014). Subaltern's image and the real: an enquiry. *Postcolonial Studies*, page In review. 2, 3

Sengupta, R., Bapi Raju, S., and Melcher, D. (2014a). Big and small numbers: empirical support for a single, flexible mechanism for human numerosity perception. *Attention, Perception & Psychophysics*, page In revision. 79

Sengupta, R., Bapi Raju, S., and Melcher, D. (2014b). A visual sense of number emerges from the dynamics of a recurrent on-center off-surround neural network. *Brain Research*, 1582:114–124. 78, 79, 101, 130, 139

Sengupta, R., Basu, P., Melcher, D., and Bapi Raju, S. (2014c). Accounting for subjective time expansion based on a decision, rather than perceptual, mechanism. In *Vision Sciences Society Annual Conference*. 109, 130

Shapiro, L. (2004). *The mind incarnate*. The MIT Press. 8

Shea, J. J. and Sisk, M. L. (2010). Complex projectile technology and homo sapiens dispersal into western eurasia. *PaleoAnthropology*, 2010:100–122. 24

Skinner, B. F. (1938). *The Behavior of Organisms*. Copley Publishing Group. 4

Spencer, H. (1872). *Principles of Psychology*. Longman, Brown, Green and Longmans. 9

Sperling, G. and Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review*, 102:503–532. 28

Stetson, C., Cui, X., Montague, P., and Eagleman, D. (2006). Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron*, 51(5):651–659. 129

Stoianov, I. and Zorzi, M. (2012). Emergence of a 'visual number sense' in hierarchical generative models. *Nature Neuroscience*, 15:194–196. 34, 54, 55, 79, 106

Styles, E. (2005). *Attention, Perception and Memory: An integrated introduction*. Psychology Press. 4, 26, 28

Thomas, E. A. C. and Weaver, W. B. (1975). Cognitive processing and time perception. *Peception and Psychophysics*, 17:363–367. 35, 101

Thomas, M. S. S. and McClelland, J. L. (2008). *The Cambridge Handbook of Computational Psychology*, chapter Connectionist Models of Cognition, pages 23–58. Cambridge University Press. 10

Todd, J. J. and Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, 428:751–754. 31, 48, 79, 92

Todd, J. J. and Marois, R. (2005). Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. *Cognitive, Affective, & Behavioral Neuroscience*, 5:144–55. 48, 79

Treisman, A. (1988). Features and objects: the fourteenth bartlett memorial lectures. *Quarterly Journal of Experimental Psychology*, 40:201–237. 24

Treisman, A. and Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12:97–136. viii, 28, 29

Treisman, M. (1963). Temporal discrimination and the indifference interval. implications for a model of the "internal clock". *Psychological Monographs*, 77:1–31. 35, 101

Trick, L. and Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? a limited-capacity preattentive stage in vision. *Psychological Review*, 101(1):80–102. 33, 78

Tse, P. U., Intriligator, J., Rivest, J., and Cavanagh, P. (2004). Attention and the subjective expansion of time. *Perception & psychophysics*, 66(7):1171–1189. xii, 35, 73, 74, 101, 105, 109, 130

Usher, M. and Cohen, J. D. (1999). Short term memory and selection processes in a frontal-lobe model. In *Connectionist models in cognitive neuroscience*, pages 78–91. Springer. 36, 38, 69, 130

van den Berg, R., Shin, H., Chou, W., George, R., and Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, 109(22):8780–8785. 63, 100

Vico, G. (1984). *The New Science of Giambattista Vico*. Cornell University Press. 2, 17, 18

Vogel, E. K. and Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428:784–751. 31, 55

von der Malsburg, C. and Buhmann, J. (1992). Sensory segmentation with coupled neural oscillators. *Bilogical Cybernetics*, 67:233–242. 37

Warjri, J. A. and Sengupta, R. (2011). Subjectivation: the two-headed symptom. *Journal of Eurasian Studies*, IV:109–113. 3

Wässle, H. (1999). Colour vision: A patchwork of cones. *Nature*, 397:473–475. 5

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20:158–177. 4

Waugh, N. C. and Norman, D. A. (1965). Primary memory. *Psychological Review*, 72:89–104. x, 31, 60

Weber, E. H. (1978). De pulsu, resorptione, auditu et tactu. annotationes anatomicae et psychologicae. In Ross, H. E. and Murray, D. J., editors, *The sense of touch*. Academic Press. 4

Webster, M. J. and Ungerleider, L. G. (1999). Neuroanatomy of visual attention. In Parasuraman, R., editor, *The attentive Brain*, pages 11–34. The MIT Press. 24

Whalen, J., Gallistel, C. R., and Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10:130–137. 33, 78, 108

Whorf, B. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. The MIT Press. 5, 110, 131

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625–636. 109, 113, 131, 135

Wilson, R. A. and Clark, A. (2009). How to situate cognition: Letting nature take its course. In *The Cambridge Handbook of Situated Cognition*. Cambridge. 109, 131

Wong, K.-F. and Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4):1314–1328. 74

Wundt, W. (1912). *An introduction to psychology*. Allen & Unwin. 4

Wutz, A. and Melcher, D. (2014). The temporal window of individuation limits visual capacity. *Frontiers of Psychology*, 5:1–13. 114

Xu, Y. and Chun, M. M. (2005). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, 440:91–95. 31, 32

Yarkoni, T., Barch, D. M., Gray, J. R., Conturo, T. E., and Braver, T. S. (2009). Bold correlates of trial-by-trial reaction time variability in gray and white matter: a multi-study fmri analysis. *PLoS One*, 4(1):e4257. 118

Young, T. (1802). The bakerian lecture: On the theory of light and colours. *Philosophical Transactions of the Royal Society of London*, 92:12–48. 5

Zhang, W. and Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192):233–235. 63, 100

BRAIN RESEARCH 1582 (2014) 114–124

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

Research Report

A visual sense of number emerges from the dynamics of a recurrent on-center off-surround neural network

Rakesh Sengupta[a], *Bapi Raju Surampudi*[a,c,*], *David Melcher*[b]

[a]*Center for Neural and Cognitive Sciences (CNCS), University of Hyderabad, Hyderabad 500046, India*
[b]*Center for Mind/Brain Sciences (CiMEC), University of Trento, Trento 38068, Italy*
[c]*School of Computer and Information Sciences, University of Hyderabad, Gachibowli, Hyderabad 500046, India*

ARTICLE INFO

Article history:
Accepted 9 March 2014
Available online 6 August 2014

Keywords:
Visual sense of numbers
Computational model
On-center off-surround
Neural network
Enumeration
Spatial attention
Individuation
Numerical cognition

ABSTRACT

It has been proposed that the ability of humans to quickly perceive numerosity involves a visual sense of number. Different paradigms of enumeration and numerosity comparison have produced a gamut of behavioral and neuroimaging data, but there has been no unified conceptual framework that can explain results across the entire range of numerosity. The current work tries to address the ongoing debate concerning whether the same mechanism operates for enumeration of small and large numbers, through a computational approach. We describe the workings of a single-layered, fully connected network characterized by self-excitation and recurrent inhibition that operates at both subitizing and estimation ranges. We show that such a network can account for classic numerical cognition effects (the distance effect, Fechner's law, Weber fraction for numerosity comparison) through the network steady state activation response across different recurrent inhibition values. The model also accounts for fMRI data previously reported for different enumeration related tasks. The model also allows us to generate an estimate of the pattern of reaction times in enumeration tasks. Overall, these findings suggest that a single network architecture can account for both small and large number processing.

1. Introduction

According to the theory of a visual sense of number (Burr and Ross, 2008a, 2008b), the ability to rapidly estimate the numerosity of a set of items reflects a basic, perceptual process. Developmental studies have shown that infants show an ability to distinguish between different numerosities at a young age (Xu and Spelke, 2000). Studies of non-human animals, such as cotton-top tamarins (Hauser et al., 2003), point towards the possible evolutionary origins of the visual

---

[*] Correspondence to: Current address: Cognitive Science Lab, International Institute of Information Technology (IIIT), Gachibowli, Hyderabad 500032, India.
E-mail addresses: qg.rakesh@gmail.com (R. Sengupta), bapiks@yahoo.co.in (B.R. Surampudi), david.melcher@unitn.it (D. Melcher).

http://dx.doi.org/10.1016/j.brainres.2014.03.014
0006-8993/© 2014 Published by Elsevier B.V.

Behavioral/Cognitive

A Shared, Flexible Neural Map Architecture Reflects Capacity Limits in Both Visual Short-Term Memory and Enumeration

André Knops,[1,2] **Manuela Piazza,**[1,3,4*] **Rakesh Sengupta,**[5] **Evelyn Eger,**[3,4] **and David Melcher**[1*]
[1]Center for Mind/Brain Sciences, University of Trento, I-38068 Rovereto, Italy, [2]Department of Psychology, Humboldt University of Berlin, D-12489 Berlin, Germany, [3]National Institute of Health and Medical Research, Unit 992, Cognitive Neuroimaging Unit, F-91191 Gif/Yvette, France, [4]Atomic Energy and Alternative Energies Commission, Departement of Life Sciences (DSV)/Biomedical Imaging Institute (I2BM), NeuroSpin Center, F-91191 Gif/Yvette, France, and [5]Center for Neural and Cognitive Sciences, University of Hyderabad, 500046 Hyderabad, India

Human cognition is characterized by severe capacity limits: we can accurately track, enumerate, or hold in mind only a small number of items at a time. It remains debated whether capacity limitations across tasks are determined by a common system. Here we measure brain activation of adult subjects performing either a visual short-term memory (vSTM) task consisting of holding in mind precise information about the orientation and position of a variable number of items, or an enumeration task consisting of assessing the number of items in those sets. We show that task-specific capacity limits (three to four items in enumeration and two to three in vSTM) are neurally reflected in the activity of the posterior parietal cortex (PPC): an identical set of voxels in this region, commonly activated during the two tasks, changed its overall response profile reflecting task-specific capacity limitations. These results, replicated in a second experiment, were further supported by multivariate pattern analysis in which we could decode the number of items presented over a larger range during enumeration than during vSTM. Finally, we simulated our results with a computational model of PPC using a saliency map architecture in which the level of mutual inhibition between nodes gives rise to capacity limitations and reflects the task-dependent precision with which objects need to be encoded (high precision for vSTM, lower precision for enumeration). Together, our work supports the existence of a common, flexible system underlying capacity limits across tasks in PPC that may take the form of a saliency map.

Introduction

Visual cognition is characterized by high flexibility but also capacity limits. Although the visual system can adapt its representational accuracy, the number of items concurrently processed is limited: in tasks as different as rapid object enumeration or visual short-term memory (vSTM), subjects can only process three or four items at a time. These capacity limits may reflect a general mechanism of object individuation (Piazza et al., 2011; Wutz and Melcher, 2013), commonly accessed in many different attentional tasks and that we suggested may take the form of a saliency (or priority) map (Bisley and Goldberg, 2003). Saliency maps topographically represent the conspicuity (or "saliency") of items at every location. Map-like architectures for spatial attention have been observed previously in the monkey lateral intraparietal (LIP) area (Bisley and Goldberg, 2003) and the putative human homolog posterior parietal cortex (PPC; Connolly et al., 2002). Critically, PPC has been implicated in studies of capacity limits in both enumeration (Piazza et al., 2002) and vSTM (Todd and Marois, 2004), as well as in visuospatial attention tasks in general, suggesting a shared neural substrate for capacity limits across tasks (Colby and Goldberg, 1999). Proof for the hypothesis of shared neural systems across tasks remains scarce because of a lack of studies investigating more than one task at a time (but see Silk et al., 2010). Here we directly test the hypothesis that a map architecture in human PPC (Gottlieb, 2007; Bays et al., 2010; Melcher and Piazza, 2011; Franconeri et al., 2013) represents individual items with a flexible degree of precision (e.g., modulable by context and task requirements) and reflects capacity limits across different tasks. Recent empirical and computational evidence link lateral inhibition strength between items to the precision of represented items within a map (Roggeman et al., 2010; Dempere-Marco et al., 2012; Sengupta and Melcher, 2014). High inhibition reduces the noise within a map, allowing for precise representations of items, but restricts capacity to few items. Conversely, low inhibition allows for a larger number of items to be represented yet less precisely. The representational precision of a given item varies with the observer's current goals. Whereas in a vSTM task participants encode multiple features, such as location and orientation of items, in enumeration tasks, no precise encoding of object features is necessary. The mere individuation of items is sufficient to encode them as units (Melcher and Piazza, 2011; Wutz and Melcher, 2013).

Here, we manipulated the required representational precision of objects by engaging participants in two tasks: (1) a vSTM task,

Received June 28, 2013; revised May 19, 2014; accepted May 22, 2014.

Author contributions: A.K., M.P., and D.M. designed research; A.K. and R.S. performed research; A.K., R.S., and E.E. analyzed data; A.K., M.P., E.E., and D.M. wrote the paper.

This work was supported by the Italian Ministry of Universities and Research (Research Programs of Relevant National Interest 2009), the Caritro Foundation (Fondazione Cassa di Risparmio di Trento e Rovereto), German Research Foundation Grant KN 959/2 (A.K.), India-Trento Programme for Advanced Research (R.S.), and European Research Council Starting Grant 313658 (D.M.).

*M.P. and D.M. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. André Knops, Department of Psychology, Humboldt University of Berlin, Rudower Chaussee 18, D-12489 Berlin, Germany. E-mail: knops.andre@gmail.com.

DOI:10.1523/JNEUROSCI.2758-13.2014

0270-6474/14/349857-10$15.00/0