Title
Differential Equations Modeling of Integration of Ordered Memory, Perception, and Attention Activities in Episode Processing

Author
Cadez, Eva

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, MERCED

Differential Equations modeling of Integration of Ordered Memory, Perception, and Attention Activities in Episode Processing

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

in

Cognitive & Information Sciences

by

Eva Cadez

Committee in charge:

Professor Michael J. Spivey, Chair
Professor Christopher T. Kello
Professor David C. Noelle
Professor Jeffrey Yoshimi

2013

My sincere thanks to: Zora and Vladimir Cadez, my parents, Eva Csonka, Igor V. Cadez, Jeff Yoshimi, Michael Spivey, Akesha Horton, my thesis committee, Evan Heit, and Joseph J. Campos.

The Dissertation of Eva Cadez is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, Merced

2013

Contents

1 Introduction **2**

2 Episode processing modeled and simulated by differential equations **4**

2.1 Preliminary notes on general theoretical and modeling aspects of this dissertation . . . 5

2.1.1 The CDIE model and some other models use similar mathematical equations . . . 7

2.2 Mathematical models and dynamical systems . . . 11

2.3 The Complex Dynamical Integrated Episode Processing hypothesis . . 15

2.3.1 The logic of the CDIE equation . . . 15

2.3.2 The dynamical systems and CDIE model . . . 20

2.3.3 CDIE details and solutions . . . 21

2.3.4 "Order in Time As Order in Space" idea from cognitive linguistics 24

2.3.5 Arguments for a distinct order information . . . 25

2.3.6 CDIE order information . . . 27

2.4 Differential equations in modeling changes in systems elsewhere in science 28

3 Time-Based (spontaneous) Decay in Episodic Processing **32**

3.1 Spontaneous decay and the CDIE model . . . 32

3.2 Background on decay research . . . 35

3.2.1 Literature on decay, rehearsal, and interference . . . 37

3.2.2 Summary of the studies and conclusions . . . 39

3.2.3 Decay in other mathematical models of cognition: present but may be hidden. . . . 41

3.2.4 More reasons to consider decay . . . 45

3.2.5 Conclusions about time based decay in research . . . 47

4 CDIE application on spontaneous decay issues: Decay is important for Distinctiveness of stimuli **48**

4.1 The Word Length Effect and Distinctiveness analysis. . . . 48

4.1.1 Intermediate discussion of distinctiveness . . . 67

4.2 Amount of decay and the WLE . . . 72

4.2.1 Very fast spontaneous decay in CDIE . . . 73

4.2.2 Perfect memory, Decay amount in CDIE and the Learning Rate and Penalty Terms in Connectionist Networks . . . 75

4.3 Serial Position Curves (SPC) in CDIE and Distinctiveness analysis. . 76

4.3.1 The Primacy Effect . . . 76

4.3.2 The Recency Effect . . . 77

4.4 Distinctiveness Revisited . . . 81

5 Chunking and temporal gaps effects **87**

5.1 Conclusions from this section . . . 90

6 The role of non-constant time-based decay in attention and false memory **91**
6.1 False Memory experimental results foreshadowed 101

7 Forgetting Curves - Emergence **104**
7.1 The shape of forgetting 106

8 Concluding remarks **110**

9 References **115**

List of Figures

1 A single Gaussian input. Every concept is modeled by a Gaussian input with its spread along the parameter dimension, "x", to acknowledge that activating one concept usually means activation of several more psychologically related concepts and that this activation weakens as concepts become more different (distant). . . 22

2 A 3D snapshot one time point of the "2D" field development. On the vertical axis, the intensity of a SUM of two fields is plotted (TAS + CS) and this intensity at each point changes at each time step; the figure shows only one of many distributions of activations during the evolution. Each activation has two coordinates: the position in some ordered sequence of events and the position in the conceptual space. Only the highest activations will be the inputs into another field with a different time scale that will be used during recall, for example. 23

3 "2D" field. This is, again, a 3D snapshot one time-instance of the "2D" field development, as in Fig. 2, except that the activation is color coded instead of the vertical axis being visible. Only the highest activations will be inputs into another field with a different time scale that will be used during recall, for example. . . . 23

4 Marine Angelfish 34

5 Fish stripes as they appear on the fish Pomacanthus (picture "d") and as obtained by process simulation using the system of equations 5.12 and 5.13 (picture "g"). (From Kondo & Arai, 1995, Nature. Author permission obtained. Publisher permission in progress.) 35

6 To capture abrupt change in brightness, a periodic function slightly different than the sine function can be used to fit the change of intensity of a dark color along one direction on the skin of the fish Pomacanthus. Variables such as decay rates and concentration do not play a role in this fitting. 36

7 Minimal decay of for stimuli presented one by one to a field with a flat initial distribution. The stimuli are the same in every respect but their timing. This is not essential here but it will be in figures 28 and 30. There, the sole difference from this simulation is in the amount of decay in the system's dynamics. 39

8 Nine stimuli of duration one are presented in time (Red, Green, Blue, Cyan, Magenta, Black, Gray, Orange, Purple) and their activations (intensities) at each time step. The primacy effect is pronounced. Several lines and peaks of the same color- belonging to the same stimulus- are visible because the stimulus intensity was plotted at each time step for the entire evolution time. This way it is clearly visible how long the stimulus lasts at which strength of intensity AND multiple peaks (of various colors) that are visible at the same time location, indicate that these other-than-the-highest peaks exist in the vicinity of the highest one along the CS dimension. 54

9 For the same stimuli, the serial position is on the x-axis, and the activation of each stimulus on y-axis. For this figure, the three highest peaks from the Fig. 8 for each stimulus are averaged. Note that since the intensity units and thresholds are arbitrary, this is not problematic. Very similar general shape of the curve is obvious from Fig. 8 if just the highest peaks are connected. 55
10 Duration of stimuli is 2 time units, the break between them is always 0. 56
11 Duration of stimuli is 2 time units with zero duration break. 57
12 Duration of stimuli is 3 time units and all breaks last 0 units. 58
13 Duration of stimuli is 3 time units and zero duration of breaks. 59
14 Long words add more total activation to the field but at the same time the activation of all of them seems to be very similar compared to short words. Everything else being equal, this adds distinctiveness to words in the short-words lists. Note, however, in experiments rarely everything else is equal. 60
15 Order and duration of two-unit of stimuli do not change from Fig. 15 to Fig. 18. All the pauses between stimuli are the same within each of the four subplots. The pauses here are zero units. The depth of gaps between stimuli increases from Fig. 15 to Fig. 18 as a consequence of the slower rate of presentation of stimuli. . . . 61
16 Order and duration of two-unit of stimuli do not change from Fig. 15 to Fig. 18. All the pauses between stimuli are the same within each of the four subplots. The pauses here are one unit. The depth of gaps between stimuli increases from Fig. 15 to Fig. 18 as a consequence of the slower rate of presentation of stimuli. . . . 62
17 Order and duration of two-unit of stimuli do not change from Fig. 15 to Fig. 18. All the pauses between stimuli are the same within each of the four subplots. The pauses here are two units. The depth of gaps between stimuli increases from Fig. 15 to Fig. 18 as a consequence of the slower rate of presentation of stimuli. . . . 63
18 Order and duration of two-unit of stimuli do not change from Fig. 15 to Fig. 18. All the pauses between stimuli are the same within each of the four subplots. The pauses here are three units. The depth of gaps between stimuli increases from Fig. 15 to Fig. 18 as a consequence of the slower rate of presentation of stimuli. . . . 64
19 Order and duration (3 units) of stimuli does not change from Fig. 19 to Fig. 21. All the pauses between stimuli are the same within each of the figures. The pauses here are of duration zero units. The depth of gaps between stimuli increases from Fig. 19 to Fig. 21 as a consequence of the slower rate of presentation of stimuli. . 65
20 Order and duration (3 units) of stimuli does not change from Fig. 19 to Fig. 21. All the pauses between stimuli are the same within each of the figures. The pauses here are of duration one unit. The depth of gaps between stimuli increases from Fig. 19 to Fig. 21 as a consequence of the slower rate of presentation of stimuli. . 66
21 Order and duration (3 units) of stimuli does not change from Fig. 19 to Fig. 21. All the pauses between stimuli are the same within each of the figures. The pauses here are of duration two units. The depth of gaps between stimuli increases from Fig. 19 to Fig. 21 as a consequence of the slower rate of presentation of stimuli. . 67
22 Five short stimuli (red, blue, magenta, gray, and purple) and four long ones are presented with no breaks between them. More explanation is in the text. 68

23 Five short stimuli (red, blue, magenta, gray, and purple) and four long ones are presented with 3-unit breaks between them. More explanation is in the text. . . 69
24 A simple calculation of distinctiveness: the intensity difference between the peak and the following gap. 70
25 All the inputs are the same width along the CS dimension. Compare with Fig. 26 73
26 Longer words have wider Gaussians along the CS dimension than short words, and compared to long words in Fig.25. 74
27 WLE in the CDIE fits well with experimental data for lists containing both long and short words. The long words' probability of recall is reduced due to their larger neighborhoods but their large peaks offset this, as illustrated in Fig.24. 75
28 Very fast decay: The four stimuli of same duration are presented on at a time. The stimuli are identical to those in figures 29 and 30, but the decay speed of the system is different. 78
29 Very slow decay: The four stimuli of same duration are presented on at a time. The stimuli are identical to those in figures 28 and 30, but the decay speed of the system is different. 78
30 Moderate Decay: The four stimuli, exactly the same as those in figures 28 and 30 (Red, Green, Blue, Cyan,) and their activations (intensities) at each time step are presented. The primacy effect is obvious. Prolonged distinct activation of the last stimulus might contribute to the recency effect. The stimuli are identical to those in figures 28 and 29, but the decay speed of the system is different. 79
31 (same as Fig. 19.) Compare the last and the second to the last stimulus' evolution after they reach their peaks. The break between stimuli is 0. 82
32 (same as Fig. 20.) Compare the last and second to the last stimulus' evolution after they reach their peaks. The break between stimuli is 1unit. 83
33 (same as Fig. 21.) Compare the last and second to the last stimulus' evolution after they reach their peaks. The break between stimuli is 2 units. 84
34 Serial presentation for short words (duration=1) 87
35 Serial presentation for long words (duration=3) 88
36 Serial Position Curve for short words (duration=1). Item presentation position is on the x-axis. 89
37 Serial Position Curve for short words (duration=3). Item presentation position is on the x-axis. 90
38 As the stimulus duration – break ratio gets smaller, stimuli get temporally very distinct but in terms of other features of their shape, they become almost identical. The recency effect is not present because of this. 91
39 Forward and backward latencies from the experiments in Anderson et. al., 1998 . 92
40 CDIE simulation nicely captures the results from Fig. 40. 94

41 Two Gaussians representing two concepts, in this case, which are activated in the interaction field. Notice the area containing x-positions with low activations between the two peaks of activation. This indicates that the concepts are easily distinguishable and that no other concepts are significantly activated. This distribution of activations is not a necessary one, as discussed later in the text and depicted in Fig. 45 97

42 The shape of the attenuation function Alpha, $\alpha(x,t)$, along the x - axis. At each point, "x", the intensity of α is different indicating that some concepts are more potentiated or suppressed by it. (For more discussion about the consequences of this, see discussion below.) In addition, as explained in the text, the position of the minimum of the function α changes in time (not depicted in this figure). . . . 97

43 This image represents the normalized solution of the development of the CDIE sum of the two Gaussians. The distribution of activation along the x-axis changes as time progresses (from the top towards the bottom of the graph). Notice the emerging large activations between the original peaks of activation that appear after the attention has been applied for some time. 99

44 This image depicts the non-normalized solution of the development of the CDIE sum of the two Gaussians. In general, the process of normalization of intensities, "I", will change the contributions of points' intensities to each other in many ways due to the complexity of the system, as described by the CDIE equations, as compared to non-normalized solutions. This results in different evolutions of the same initial distribution. 99

45 In BLUE: Two close Gaussians. Unlike the situation in Fig. 10, the two concepts activated here are very similar, as depicted by their positions. After they are introduced to the interaction field, they are integrated with each other and other activations and will evolve according to the CDIE equations. In PURPLE: CDIE sum of the two close Gaussians. An emerging highly activated peak forms between the original activation peaks. Unlike in the case depicted in Fig. 43 and 44, this is not the result of attention. However, as explained in the text, the attention may now be used to suppress the emergent activation peak. 100

46 This image illustrates the normalized solution of the development of the CDIE sum of the two close Gaussians. This figure is analogous to the Fig. 43 in every way. The only difference is that the two initial Gaussians are closer together. More discussion is given in the text. 101

47 This is a non-normalized solution of the development of the CDIE sum of the two close Gaussians. This figure is analogous to the Fig. 44 in every way. The only difference is that the two initial Gaussians are closer together. More discussion is given in the text. 102

48 The sum of two one-dimensional Gaussians. 104

49 The two-dimensional Gaussian. 105

50 CDIE simulation of attention suppressing false memory in the Forewarning condition from Heit et al., (2004); simulation captures the experimental results very well. 106

51 CDIE simulation of the Inclusion condition from Heit et al., (2004); simulation captures the experimental results very well. 107
52 A possible set of learned material is being slowly forgotten over time. 109
53 Two exponential functions with the same half-life. The red one begins its decay earlier than the blue one. Importantly, they never intersect. 111
54 Two exponential functions with different half-life. The red one begins its decay after the blue one. Importantly, they intersect. 112

Permissions have been granted for used copyrighted material from John R. Anderson, Richard King Mellon Professor of Psychology and Computer Science, Carnegie Mellon University; Journal of Memory and Language; Shigeru Kondo,Professor, Laboratory of Pattern Formation, Graduate School of Frontier Biosciences, Osaka University; journal Nature.

Education:

- Ph.D. in Cognitive and Information Sciences, University of California, Merced, Fall 2013.
- B. A. Psychology, University of California Berkeley, Fall 2005.
- Psychology, Orange Coast College, Costa Mesa, CA, Spring 2003.

Research:

Dynamical systems models of episode processing and conceptual change; Theoretical (Mathematical) Cognitive Science.

Adviser: Professor Michael J. Spivey, UC Merced.

Differential Equations modeling of Integration of Ordered Memory, Perception, and Attention Activities in Episode Processing

Eva Cadez

Doctor of Philosophy in Cognitive & Information Sciences

University of California, Merced

2013

Chair of the Committee: Professor Michael J. Spivey

Abstract

By using differential equations to simulate cognition at the level of abstract mental constructs, the Complex Dynamical Integrated Episode Processing Hypothesis (CDIE) sheds new light on how ordered events are encoded in episode processing. It assumes that during the episode processing current distribution of activated information is influenced by external inputs of information in a complex way due to additional influences of spatiotemporal context and spontaneous decay. In particular, it integrates perception and existing memory, attention related to time based decay modification and as possible additional simple source of information, forgetting (including both constant and non-constant time-based decay and interference), relationship of this decay and item distinctiveness, false memories resulting from three possible mechanisms and their suppression, etc. Cognition is treated as a complex dynamical system and this model for the first time in the literature integrates all of the mentioned classical concepts and sheds light on their temporal dynamics. The following are some of the unique contributions of this project.

Episode processing is currently not a particularly well identified traditional research area. Nonetheless, there exists a continuity in cognitive processing explored in many other areas and episode processing or the episode processor are defined as a complex system changing in time.

This dynamical process model is an innovative application of an infinite-dimensional dynamical system to the ordered cognitive events in episode processing. On the one hand, while dynamical systems (DS) theory has been recognized as an important tool in cognitive science, this kind of infinite-dimensional system, where a continuum of values evolves in time, has not been heavily studied by modelers or philosophers discussing dynamical approaches to non-neural level cognitive science.

The model assumes that order in time is cognitively treated as order in space. A complex systems differential equations approach to time-as-space ideas of cognitive linguistics has not been quantitatively implemented in cognitive science and it shows promise to open an entirely new area of research in cognitive linguistics.

1 Introduction

Episode processing is not a unified research area, such as memory research. There exists a continuity in cognitive processing but the classical research areas are relatively arbitrary categories that are not always helpful. While all of the models in these areas "are aware" of inputs, encoding, retrieval, and retention, concepts like "short term serial recall" are operational definitions, based on experimental paradigms. In principle, however, they are fully amenable to explicit references to "episode processing" and the "episode processor," (which I later define as a dynamical systems construct). Connectionist network models use probably the most technical description of their work such as "working with data varying over time", "solving temporal problems," "temporal models" and this is adopted here, as well as the most research-area-neutral name – episode processing.

This dynamical process model is an innovative application of an infinite – dimensional dynamical system to the ordered cognitive events in episode processing. On the one hand, while dynamical systems (DS) theory has been recognized as an important tool in cognitive science, this kind of infinite – dimensional system, where a continuum of values evolves in time, has not been heavily studied by modelers or philosophers discussing dynamical approaches to non-neural level cognitive science. On the other hand, the specific application of the DS to numerous aspects of ordered events of episode processing, which are parts of many areas of cognitive science, are here studied together as integrated complex processes. Therefore it seems more appropriate to describe the hypothesis presented in this work as being about episode processing – it unites in a new way perception, attention, memory, and more. It assumes that during episode processing the current distribution of activated information is influenced by external inputs of information in a complex way due to additional influences of spatiotemporal context and spontaneous decay.

In the current literature there is work that is related to the work presented here.

Namely, in 1969, Stephen Grossberg published the paper "On the serial learning of lists." There he used a very similar differential equation to model one aspect of the processing of episodes - learning of a list of items. The work presented here has a broader scope, beyond learning, but it also has at least four other important novelties. The first is the non-constant time based decay. The decay term is in general (if present) in related models always assumed to be a constant. The second, arguably more important difference is that in the present model the conceptual space - the space of items that might be learned in order - is a continuous dimension while it is a relatively small set of discrete items in Grossberg's work. Third, in the present model the order in time is treated as order in space (an idea that comes from research in cognitive linguistics) while in Grossberg's work only the real continuous time exists in the model and the order in time is dealt with via the Primacy Gradient, as we shall see later. Finally, the list learning is studied via connectionist networks in Grossberg's work but not here, which produces fundamentally different, but equally important insight in cognitive processes involved.

More recently, a similar class of DS models, the Dynamic Field Theory (DFT), has been used by Kopecz, Schöner, Erlhagen and their colleagues (Kopecz, Engels, & Schöner, 1993). Since the Complex Dynamical Integrated Episode processing hypothesis (CDIE) and the Dynamical Field Theory (DFT) simulations are visually similar, and along with Grossberg, they all share the general radiative transfer equation, it is useful to specifically note several fundamental distinctions between these models.

First, in comparison with the CDIE, the level of cognition modeled buy the DFT is different. "The DFT is in a class of bi-stable neural networks first developed by Amari" (Johnson, Spencer, & Schöner, 2008). While the DFT, following the above mentioned work of Amari, uses two layers a layer of "excitatory neurons" and another one, the layer of "inhibitory interneurons," (and often actually several layers like these), the CDIE does not deal with these concepts of neural or connectionist networks at all; it contains one "layer," in DFT terms, which is, fundamentally, a mathematical concept of an interaction place in CDIE. Therefore, just as is the case with the DFT term "field", the "layer" in the CDIE is not a neural tissue or its analog.

Second, a dimension that in the CDIE represents an abstract conceptual space, in DFT application represents the real space in front of a subject in an experiment, specifically "the retinal position of a point of light along a horizontal axis" or "the direction of a goal-directed movement" in some other applications. In their later work, the authors added other dimensions to simulate two dimensional real space as seen by subjects. The CDIE does not deal with this real perceived space at all. On the other hand, just like with Grossberg's work, the CDIEs time-as-space dimension and its dynamics are not present in any of DFT applications or theory.

Third, the details about the time-based decay term, which is a function of time and position along two dimensions in CDIE, but is a constant in DFT , and the details of how the context influences current information in CDIE differ from the DFT model and therefore in its applications. This substantially changes the dynamics of these two systems. As we shall see, this will have profound implications for modeling and theorizing about several cognitive phenomena attention and false memories (which are not explored in DFT modeling).

Fourth, the way the separate dimensions are combined into a two-dimensional system and into other postulated systems (or "layers" in the DFT terminology) differ in CDIE and DFT. In CDIE, the two dimensions are simply added together: space of concepts and order position do not otherwise interact with each other. In DFT these combinations are different. This difference is a direct consequence of differences in natures of phenomena modeled. Thus, currently, the CDIE uses a simple sum of Gaussians along two of its dimensions to simulate input information, while the DFT uses a two-dimensional Gaussian in cases where their "layers" are two-dimensional (e.g., Spencer, Barich, Goldberg, & Perone, 2012).

Finally, the "resting level" term from the DFT equation is not contributing activation in each time step as is the case in the DFT but it is taken into account as a part of an initial state. This changes the dynamics of the model.

The model presented here includes a time based decay, which is currently a largely disputed issue in episodic memory research (e.g., Brown & Lewandowsky, 2010). This model mathematically defines time-based decay, and using this definition in a completely novel way it examines the role of time-based decay in Serial Position Effects, Forgetting Curves, Item Distinctiveness and Confusability (in Serial Recall, Attention, False Memory, Word Length Effect, and Chunking in memory.) These analyses provide novel insights in complexity of the decay's influences on cognitive phenomena as a part of their integrated dynamics but also via its role in attention and distinctiveness of items. A more complex view on time-based decay and distinctiveness emerges from these novel analyses of complex interactions happening during the episode processing.

Lastly, a complex systems differential equations approach to time-as-space notion of cognitive linguistics presented here have not been done before in cognitive science and it might, perhaps, open an entirely new area of research in cognitive linguistics.

This dissertation additionally suggests that differential equations in general might be useful in modeling cognition at more abstract levels than thus far. Rich DE modeling is relatively well represented at the level of neuronal activities, mostly related to bio-physical processes but their use diminishes as the level of modeling increases to more abstract cognitive concepts. Comparing processes at different biological, including cognitive, levels might shed more light on each of those, in addition to offering insights into processes evolving in time at each level alone.

2 Episode processing modeled and simulated by differential equations

Every experience seems to be some kind of a sum of ordered events intricately combined in episodes we are dealing with right now and have dealt with in the past. An episodic memory test, the serial recall paradigm for example, might involve presenting subjects with a list of items, perhaps words, to be remembered in their presentation order and then recalled. In this process, new events interact with remains of what we have previously encountered. Many variations of these experiments – where a series of perceptual episodes must be processed, individuated, and ordered – are possible and have been used to manipulate different aspects of cognitive processes involved. Episode-processing tasks play a role in various research areas related to cognition. If one thinks about the procedures involved in these tasks, it seems obvious that a subject in the experiment has to remember at least two things: the items themselves and the order of the items. Detailed mechanisms of episode processing including interactions of attention, memory, learning, and forgetting are still a mystery.

Many different kinds of models of these processes have been proposed during the decades of scientific research (e.g., Grossberg, 1964, Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974; Brown, Preece, & Hulme, 2000; Grossberg, 2013). The models typically capture some aspects of episode processing well but leave other aspects to different models. A variety of models is itself very useful because they may inform each other and may be gradually integrated into progressively more realistic understanding of cognition. A dynamical systems (DS) approach to modeling in cognitive science, which may include DE simulations like those presented here, is a relatively recent endeavor (e.g., Grossberg, 1964, 1969; Rumelhart, Smolensky, McClelland, & Hinton, 1986; Smolensky, 1986; Kopecz, Engels, & Schöner, 1993; Van Geert, 1991; van Gelder, 1995; for review of recent works see Riley & Holden, 2012).

However, modeling various other cognitive processes at an abstract level (not neural or connectionist networks) using differential equations other than the ones similar to the radiative transfer equation used here is not widely present in cognitive science. Stephen Grossberg (Grossberg 1964, 1967, 1969, 1971, 1976) has started modeling both neuronal activities and list learning (a part of episode processing, at the abstract level) using DEs of the same kind as the radiative transfer equation nowadays used here as well as elsewhere in literature (e.g., Kopecz, Engels, & Schöner, 1993). For the time being, it seems that it is a coincidence that the same type of mathematics may be used to model a radiative transfer in physics, individual or group neuronal activities in neuroscience of, say, perception, and list learning, for example, in cognitive science. It is possible, of course that in the future it will be shown that some of these similarities are at a deeper level, after all.

2.1 Preliminary notes on general theoretical and modeling aspects of this dissertation

This section discusses a highly abstract, not connectionist networks, level of modeling used by the CDIE, it offers additional clarifications about the most general theoretical terms it employs, and clarifies differences between the CDIE and some other models that use the same general mathematical equation either at a different level and therefore with different applications or at the same level but with different simulations that shed light on similar issues but from a different perspective.

In this dissertation, episode processing is treated as a complex dynamical system of existing memories and traces of individual inputs, among other sources of activations explained in detail below. The inputs are ordered in time, they are parts of episodes. For systems that change in time and whose current state (a distribution of information referred to as "activations" with "intensities" in order to make an analogy with energy states of a physical or biological system or subsystems) depends on their own past state(s) and possible external influences, differential (and related) equations are the most frequently used tool in science and economics. In this model, I implemented an equation whose use is elsewhere in science widespread. In physics of radiative transfer, this form of the equation is used to model spatial distributions of photons that move along a specific direction. In this equation, the distribution changes along a spatial dimension, while in CDIE the change is in time. The terms of this equation describe a local generation or absorption of photons, and an addition of photons initially traveling in different directions which scatter in to a given path, those that due to scattering out leave the given path in all directions. In addition, it is possible to have a material through which photons travel with such properties that they induce some kind of residual photon generation (such as scintillation) at a point after the photons have traveled through that specific point (for an example of use of a radiative transfer equation and similar equations in propagation of light in biological tissue see Kim & Keller, 2003). Using the radiative transfer equation, I simulate how a single information input's trace interacts with existing information (activations in a system) (e.g., a person's activated concepts in memory) and how this system evolves in time. This way, for example, it was possible to simulate the emergence of well-known forgetting curves from a cognitive complex system (Čadež & Heit, 2011).

Fundamentally, this work *does not* make more specific assumptions about memory traces beyond the commitment to the general idea that they are the neural substrates of remembering (Brown, 1958). Presented here is *a formal mathematical model* of an evolution of an input trace information arriving to and combining with other information. This model is *completely* abstracted from biological networks in the brain and is even more abstract than connectionist network models as described by Smolensky (1988).

Smolensky's distinction between three levels of cognitive explanation: neural, sub-symbolic, and symbolic, might be is helpful to clarify how this abstract model should be interpreted in relation to other models. At each of the three levels dynamical models can be developed. As explained later, this classification is not a perfect one for this model but it nonetheless may be helpful. A dynamical system at the neural level describes a group of interconnected neurons in a region of a brain while a dynamical system at the sub-symbolic level is something like a connectionist network – a more abstract approximation of the brain both in terms of structure and dynamics. It is still based on *general* neural principles but abstracts from the details (Smolensky, 1988). A dynamical system at the conceptual level describes the processing of symbols and relations between them, for example the words and concepts in a natural language. Here is how Smolensky (1988, p. 6) describes the sub-symbolic level *in relation* to the symbolic level:

> The interactions between individual units are simple, but these units do not have conceptual semantics: they are subconceptual. The interactions between entities with conceptual semantics – interactions between complex patterns of activity – are not at all simple. Interactions at the level of activity patterns are not directly described by the formal definition of a subsymbolic model; they must be computed by the analyst.

This computation is, in a sense, what the simulation of interacting activations using differential equations presented in this context might do, although this model is in no way committed to this. This model might not even be implementable in connectionist networks at their current state of development. We can consider very formal, high-level, mathematical, dynamical systems that just describe abstract elements and relations that change in time: for example a system of objects related by abstract forces. I treat an *episodic processor* (defined later in the text) as one such system. Importantly, the CDIE model, as well as other models, possibly at various levels of explanation, should shed light on how the episode processor is implemented in the brain from an additional perspectives, and may be informed by all the other models and their perspectives, and this "communication" might significantly improve our understanding of the brain and cognition.

As mentioned above, Smolensky's classification is not a perfect tool to classify this model because the term "state" that is used here has the below defined *mathematical* sense (a set of variable values of a system at a single moment in time) and because every state changes in continuous time (which is not directly discussed in Smolensky's classification), but such conceptions have consequences for understanding what is meant by "symbols," "mental representation," "concepts," and so on. Most of fundamental cognitive concepts that this model tries to explore are best understood in terms of the "Continuity of Mind" thesis (Spivey, 2007). They are understood as a kind of a running sum of complex, interacting, integrated, and dynamic processes happening continuously in physical time. Examples of these processes are external input information, such as perceptual stimuli, electromagnetic influences from the environment or the rest of the body, activated memories, attentional processes, and so on. They all contribute to the constantly changing distribution of total activity of the mind.

2.1.1 The CDIE model and some other models use similar mathematical equations

This subsection attempts to clarify the use of the general mathematical equation of the CDIE an its use in some closely related research areas: first, in modeling biological processes related to neuronal activities and, second, in models of more abstract cognition. Particularly relevant for the work presented here is the following point there are two prominent lines of research using this same general equation. One is Stephen Grossberg's work (Grossberg, 1964, 1969) on list learning implemented in connectionist networks and the other is the Dynamic Field Theory by Kopecz, Engels, & Schöner (1993), which is at the level of neural tissues but it is simulated the same way as the CDIE.

On the one hand in 1964, Stephen Grossberg published a monograph "The Theory of Embedding Fields with applications to psychology and neurophysiology" in which he notes the need for deeper understanding of dynamics of processes underlying cognition and attempts to give mathematical formulations of *both* neural and more abstract levels of brain functioning in order to begin connecting them. To do this, the author uses mathematical geometry (and the notion of an "embedding field", not a biological term, just like in the CDIE), relates its concepts to cognition and adds time evolutions of basic concepts. The author makes an important point here about the beginning of a modeling process in general noticing an underlying structure of data, which, as mentioned above, might come from researcher's experiences with dynamical systems and mathematical structures in completely unrelated areas of science (in Grossberg's case, geometry and physics):

> In order to initiate our theoretical study, it shall be necessary to perform an inductive leap to a mathematical system which is, in a natural sense, the simplest system that provides nontrivial insight into the data at large. Such a leap is never particularly easy to motivate at first since it originally arises as a direct perception of the underlying structure of the data. It shall therefore be initially justified by showing that the mathematical object which arises actually gives a qualitative description of a large quantity of data from a small number of principles. (p. 3)

In this work, the author uses the same general equation as the CDIE in order to model spiking neurons (e.g., post-synaptic habituation) as well as more abstract list learning of nonsense syllables, among other things. In other words, he uses the same mathematical model and applies it at both the biological level of neuronal spikes but also at the abstract level of verbal learning. In his later works, the author pursues each track of modeling and analyses and tunes the equations in each context. Importantly for the work presented here, the verbal learning modeling by Grossberg is mathematically almost the same as the CDIE. The significant differences are that the conceptual dimension in Grossbergs work is discrete and that the model is embedded in a connectionist networks framework. The main focus of the model is list *learning*. Therefore, the simulations of basically the same mathematical equation are producing different insights. These insights inform each other which should be of great value for cognitive science, as argued throughout the present work.

On the other hand, the Dynamic Field Theory (DFT) approach promoted by Kopecz, Schöner, Erlhagen and their colleagues (Kopecz, Engels, and Schöner, 1993), has similar simulation framework as the CDIE, but applies the general equation at mostly biological level in sensory motor related cognition as well as in the area of robotics. In a sense, the DFT and similar models are continuation of the ideas mentioned in Grossberg's work above, on the lower abstraction level of application. The idea for these kinds of models comes from the convenience of continuous approximations for a very large number of discrete units, namely neurons or synapses. Constructing connectionist models with this huge amount of units is extremely hard even using today's computers. The differential equations models (which are those approximations) may sometimes be solved mathematically (analytically or numerically) or may be simulated, as is the case in this work. When this is not possible, another kind of analyses including direction fields, etc, may be used, as mentioned.

Keeping in mind the enormous complexity of the brain, it seems that any insight from any, necessarily limited, perspective is welcome. Beurle (1956), Griffith (1963), Grossberg (1964), Amari (1975), etc. did this DE type of neuronal tissue modeling.

In his 2005 paper "Waves and bumps in neural field theories" (p. 2), Coombes states: In many continuum models for the propagation of electrical activity in neural tissue it is assumed that the synaptic input current is a function of the pre – synaptic firing rate function [5]. These infinite – dimensional dynamical systems are typically variations on the form [32]:

$$\frac{1}{\alpha}\frac{\partial u(x,t,)}{\partial t} = -u + \int_{-\infty}^{\infty} dy\, W(y) f \circ u(x-y,t) \tag{1}$$

This equation is of a similar type as others mentioned here. As mentioned earlier, this kind of infinite – dimensional system, where a continuum of values evolves in time, has not been emphasized by philosophers discussing dynamical approaches to abstract level concepts of cognitive science such as those presented in this work.

Amari (1977) has shown that the activity of two layers (the CDIE has one) of fully interconnected neurons may be modeled using approximations on a real number line. Amari's work, in addition to others', was identified by some authors of the DFT as the one with the most similar mathematical details to the equations for the two layers they were using in robotics (Erlhagen, Bastian, Jancke, Riehle, & Schöner, 1999). These authors of the DFT state: "Our thinking about the role of population activity was guided by ideas from dynamical systems theory (Wilson, Cowan, 1973; Amari, 1977; Grossberg, 1980; & Schöner, Kopecz, Erlhagen, 1997)." The idea of application of the equation was, in authors' words to "describe how distributions of population activation (DPAs) constructed from populations of cortical neurons can be used as a tool for analyzing representations of simple task or stimulus parameters."

In between these extremes of spiking neurons being statistically approximated by differential equations and cognitive processes being simulated by differential equations a neuroelectrodynamic account of neural information transmission allows for the actual physical electrical fields themselves to perform analog computation (Poznanski & Cacha, 2011). I share with the other authors the awareness of the importance of the fact that many different cognitive phenomena, as well as physical ones, may be modeled by the same general mathematical equation. While this fact may be a pure coincidence it may also prove to be a fruitful food for scientific thought about nature of phenomena at deeper levels of analysis.

Several applications of the DFT are described next:

The idea of application of the equation was, in authors' words to "describe how distributions of population activation (DPAs) constructed from populations of cortical neurons can be used as a tool for analyzing representations of simple task or stimulus parameters." In comparison with the CDIE, this means that the level of cognition modeled is slightly different -the biophysical level- but more specifically, the dimension that in the CDIE represents the abstract conceptual space, in DFT application represents the real space in front of a subject in an experiment, specifically "the retinal position of a point of light along a horizontal axis" or "the direction of a goal-directed movement in the other case". In the early work, they used only that one dimension as is appropriate for some tasks, and later they have added other dimensions to simulate two dimensional space as seen by subjects (which were robots in some cases). Let me now illustrate some of these simulations.

Likely the most cited study involving the DFT is the simulation a motor control and development of the A-not-B error in infants 7-12 months old. When they have two boxes in front of them, one to the left and the other to the right, but at the same distance from them, they are taught to find a toy hidden in, say, the left box. After this, even if they see the toy being placed in the other box, if there is a short delay in reaching, they often reach for the first location when asked to get the toy. This is called "perseverative reaching" or "A-not-B error" (Piaget, 1954). The model includes a dynamic field in which are registered the learned position of the environment and the events during the task. In addition, another field keeps track of the history of the first one. When these are combined, added together in simulations, the authors were able to replicate results from the experiments. From simulation it became obvious that some activations need to subside so that the others may become dominant, and in this case become actions. They concluded, that unlike other models that assume that some responses get inhibited (but not for infants) in this task, the seeming lack of inhibition is actually just an emergent property of evolution in time of mechanisms that do not have any "inhibition of response mechanisms". The emergence of properties, as opposed as preexisting separate mechanism for functions of those properties in these authors' work is also an important point that is mentioned elsewhere in this dissertation.

Spencer, Simmering, Schutte, and Schöner did another interesting application of the DFT in 2007. Here, five activation fields, all 2D corresponding to the 2D environment of the task, were combined to simulate spatial cognition, the recall of a spatial position in that 2D environment, to be specific. The five fields were: a perceptual field, a long-term memory field associated with this perceptual field, a shared layer of (inhibitory) interneurons, a spatial working memory field (SWM), and a long-term memory field associated with the spatial working memory field. All of these fields represent spatial positions of an object. Their simulations were intended to examine geometric biases in memory for spatial position and interaction of different timescales. All the fields have the same type of dynamics and only the specific time "tau" was varied to slow down the dynamics of memory of sensory perception fields. In these simulations, shared with the CDIE as well as some other models, multiple unique time scales where events may be processed exist. Unfortunately, the exact way of combining the fields is not evident from the text, those are details of implementation which are typically not reported in these kinds of papers. In principle, the differential equations can be coupled so that at any time moment value of the one may be a function of the first one, or their values may be just simply added together without true interaction. The choice of how to combine fields is dependent on ideas about processes involved. Anyhow, fields are sources of additional excitatory or inhibitory inputs for each other and finally to the working memory field, which high activation position is interpreted as the candidate spatial position for recall.

As a continuation of this work, Johnson, Spencer, and Schöner (2007), reported encouraging results in 7- field modeling, where the parameter dimension was continuous color space. The position in color space was treated as a position in a 2D real space, and the position of colored objects was encoded directly into the 2D field. Binding of color with position was achieved by combining 2D fields. In both of these works, the fields were combined into a field with slower decay than the other fields and because of that they were called memory fields for the specific task. In their paper presented next the authors clarify the idea of memory here :" Thus, SWM remembers the target location even in the absence of input. We refer to this as a "working" memory state because the activation peak is stable against perturbations (e.g., neural noise). Such stability is a pre-requisite to use the contents of working memory in the service of another task a central component of Baddeley's classic definition of the working memory construct (Baddeley, 1986)." (p. 344, Spencer, Barich, Goldberg, and Perone, 2012).

As the last illustration of the DFT, I mention the Spencer, Barich, Goldberg, and Perone (2012) study on neural activity in a multi-object tracking (MOT) task. Here, a 3-layer 2D fields model is used to represent 2D space. One field (perceptual field, PF) captures positions of distractors and the other one (spatial working memory SWM) the position of targets (Fig 3). A task display feeds into both PF and SWM. The third field is a field of inhibitory interneurons. The goal of the system is to correctly identify, or have sufficiently activated peaks at the position of targets after they have been moving for about 20 s. Because the changes in activations take time, based on the parameters of the fields, and because the speed of moving objects basically changes activated positions that evolve in addition to that speed, both speed of targets and time of tracking influence the outcome of tracking. The model was found to nicely fit behavioral findings and qualitatively captures findings for the ERP studies. ERP studies, in general, with a high temporal resolution are naturally particularly close to the investigation of or by dynamical systems approaches, as evident and pointed out by these authors.

2.2 Mathematical models and dynamical systems

Mathematical and scientific uses of the notion of dynamical systems are not identical. It seems important for the present argument about the use of DEs in the CDIE but more importantly – in cognitive science, to clearly understand consequences of this distinction. In particular, one and the same mathematical dynamical system equation may be used in relation to many different natural complex systems (e.g., solar systems or neural tissues). As mentioned above, the CDIE uses an equation, more or less similar to many others, that is nonetheless used for novel detailed analyses, at a different level of modeling, of cognitive concepts that might be centuries old.

Hence, in the following text, some basic ideas about mathematical functions and, importantly, differential equations (involving derivatives of variables, their rates of change) are reviewed. The difference between a formal mathematical equation, its use for data fitting, and possible interpretations of that equation used to model specific processes is mentioned here in order to highlight the reason why the DE are a useful *tool* in cognitive science. The purpose of this is to draw attention to differential equations as an arguably underused mathematical tool to model cognition *at levels of abstraction higher than and abstracted from connectionist networks.* While DE modeling in neuroscience is widespread and some of this modeling is used in cognitive neuroscience, this kind of modeling is not popular beyond these levels. The main advantage of this and related classes of equations is that they connect various variable as well as the changes in variables over time. This means that they may model many interacting processes as they evolve, which is a marked characteristic of complex systems of high-level cognition.

Mathematical models of various kinds in cognition research may have different primary uses based on their goals. The experimental data can be described using curves of the best fit to summarize them or make predictions. The resulting equations constitute a mathematical model of the phenomenon. Most often this process is followed by theoretical proposals about the nature of processes that use insights from the fitting procedures. In physics, the experimental physics deals with measuring phenomena and representing and summarizing data found, whereas the theoretical physics is a complementary field that looks for the unifying theories to explain what processes produce the experimental data. Arguably, in most parts of cognitive science, mathematical models are not widely used to describe the processes themselves where entities modeled are concepts of cognitive science (as opposed to the biological processes of neurons, for example). Theoretical cognitive science might use differential equations to obtain the experimental results and therefore is different from the philosophy of cognitive science which deals with concepts such as the metaphysical and epistemological issues related to constructs of cognitive science.

The nature of the physical process, in principle, is not the main consideration of a mathematician. For example, Sir Isaac Newton in *The Mathematical Principles of Natural Philosophy* (1729/1968 p. 5, 6) writes about forces:

> [...] I refer the motive force to the body as an endeavor and propensity of the whole towards a centre, arising from the propensities of the several parts taken together;
>
> [...] For I here design only to give a mathematical notion of those forces, without considering their physical causes and seats.
>
> [...] (I am) considering those forces not physically, but mathematically: wherefore the reader is not to imagine that by those words I anywhere take upon me to define the kind, or the manner of any action, the causes or the physical reason thereof, or that I attribute forces, in a true and physical sense, to certain centres (which are only mathematical points); when at any time I happen to speak of centres as attracting, or as endued with attractive powers.
>
> [...] Hitherto I have laid down the definitions of such words as are less known, and explained the sense in which I would have them to be under-

> stood in the following discourse. I do not define time, space, place, and motion, as being well known to all. Only I must observe, that the common people conceive those quantities under no other notions but from the relation they bear to sensible objects. And thence arise certain prejudices, for the removing of which it will be convenient to distinguish them into absolute and relative, true and apparent, mathematical and common.

This does not mean that the physical nature of processes is not the main concern of other researchers, nor does it mean mathematical formulas do not shed any light on the nature of physical processes. Further, any model is just a human approximation of reality and as such necessarily suffers deficits that, importantly, can be partially overcome by combining different models, experimental data, statistical inferences, etc. Some simplifications do not pose a problem for our purposes so we often do not deal with them at all. With this perspective in mind, it seems obvious that mathematical models both by summarizing results of experiments and by simulating complex dynamical processes, shed light on real processes that interest various researchers. We need both experimental and theoretical cognitive science.

A dynamical system in mathematics is defined in the following way:
(http://mathworld.wolfram.com/classroom/classes/DynamicalSystems.html)

"Technically, a dynamical system is a smooth action of the reals or the integers on another object (usually a manifold)." "A means of describing how one state develops into another state over the course of time."

The "system," "state," and the "complex system" are described/defined as:

Some kind of collection of elements (possibly described by collection of variables) is usually regarded as a system, with many possible detailed definitions related to this main idea, but which are not central to this work.

The state of a system is often (but not necessarily) a set of numbers that represent one possible configuration of variables' values of the system, the current height, weight, hair length, and shoe size of a person, for example. This state can change in the next instance of time to become a new state of the system.

"A complex system is one whose evolution is very sensitive to initial conditions or to small perturbations, one in which the number of independent interacting components is large, or one in which there are multiple pathways by which the system can evolve" (Whitesides & Ismagilov, 1999). In other words: a dynamical system is a description of the way a complex system changes over time.

As with the notion of a mathematical relation (a function, for example), a dynamical system is a mathematical term (or "object") that may be related to a physical process, such as an animal episode processing. It is one way of describing the "real-life" process of changing states of a complex system of cognitive activations. Mathematically speaking, episode processing is not a dynamical system or even a mathematical object. Episode processing, like the forces described in above in Newton's work are not physical objects. For a physicist or a cognitive scientist, however, the episode processing may be a hypothetical construct/object of their science. For a physicist or a cognitive scientist, however, the forces and the episode processing may be hypothetical constructs/objects of their science. While it may seem obvious that the "episode processor" may be regarded as a complex system changing in time, it may be described using the dynamical systems of mathematics. Outside of mathematics, the dynamical system has a related but obviously not identical meaning – it is a physical, "real life", system changing in time. In sum, outside of mathematics, the dynamical system is a physical system changing in time while in mathematics it is a *description* of the way a complex system changes over time.

Van Geert and Steenbeek, (2005), in their paper "Explaining after by before: Basic aspects of a dynamic systems approach to the study of development" point out a useful conceptual distinction between static and dynamic equations:

$$y_{(t+1)} = f\ (y_t) \tag{2}$$

is a dynamical systems equation, and

$$y_i = f\ (x_i) \tag{3}$$

is a static systems equation.

The important difference between these two kinds of equations is in natures of their domains and codomains. In static equation, for example, the domain containing various heights, when used as an input of an appropriate mathematical relation, produces a weight output. In the dynamical equation the codomain has the same nature as the domain. Both may be numbers of photons, but at different distances from the source. This describes "after by before", the changes from one state to another state of the same system.

Dynamical systems may be described by one or more differential equations. This system of equations only rarely has an exact analytical solution that allows for detailed mathematical analyses of the systems behavior. Most physical dynamical systems are also very hard to implement in connectionist network models due to their complexity. However, when these kinds of analyses are not possible, the dynamical systems approach offers other, qualitative methods which allow researchers to gain some insight into the systems behavior over time. Henri Poincare pioneered these methods in late XIX century (Farlow, Hall, McDill, & West, 2002), which include analyses of state spaces, direction fields, attractors, etc. At this time, I do not use these methods. I rely only on numerical solutions for simulations of the systems behavior based on the mathematical model I describe in the next section. I refer to all of these approaches in the investigation of systems as the dynamical systems (DS) approach.

2.3 The Complex Dynamical Integrated Episode Processing hypothesis

2.3.1 The logic of the CDIE equation

In this section, more intuitive verbal explanations are provided setting the foundation for more mathematically rigorous details of the model first and later a more mathematically rigorous details of the model are given.

First it is assumed that there is a concept in a concept space, at some position $x = x_j$ which can have some activation intensity $I(x)$. We are interested in how this activation changes in time (t).

To estimate this process we start from some existing activation $I_{old}(x)$, at time $t = t_{old}$ and assume a change $\Delta_I(x)$ in the activation I during the time interval Δt where $\Delta t = t_{new} - t_{old}$. In this case we say that the $I_{new}(x)$, at time $t = t_{new}$ follows from the previous activation $I_{old}(x)$ at time $t = t_{old}$ as follows:

$$I_{new}(x) = I_{old}(x) + \Delta_I(x) \tag{4}$$

This procedure is then iterated the required number of times, depending on how far of a future we wish to simulate. The first activation I_{old} at time $t = t_{old} = 0$ (the initial condition) remains mathematically arbitrary and has to be specified according to some data obtained from experiments.

Now we have to hypothesize about the processes that induce the change Δ_I during each time step of the evolution.

To do so, we first hypothesize that the intensity of a trace will decrease if there is nothing else to prevent this. The rate of this decrease is proportionate to the current intensity of the trace and it may be modified by some constant of proportionality. This is analogous to the spontaneous decay in radioactive materials, for example. Therefore, there is a possibility of a diffusion-like attenuation of intensity I within the time interval Δt, which has its typical form,

$$\Delta_1 = -\alpha I(x)\Delta t \tag{5}$$

where α is a damping coefficient which in general case may be both t and x dependent i.e. $\alpha = \alpha(x, t)$.

Second, the influences on the rate of change may come from one or more additional sources (or sinks) of activation. When we talk about the brain activations, this may be a trace elicited by presentation of a word in a list in a recall task, or chemical or electromagnetic influences from the environment or the rest of the body, emotional influences, memories, and so on. Therefore, there is a possibility of additional external source of activation on x_j, which generally may be time-dependent, giving the change $\Delta_2 \equiv$S(x,t) Δt.

Finally, as mentioned, we may argue that the rate of change of intensity is influenced by the temporal and spatial context of the specific activated point. As previously noted, specific class of dynamical systems, systems with memory, in which the evolution of the system explicitly depends on the past history of the system may be assumed in this modeling. Depending on the shape of a path leading the system to its current state, it will behave in different ways in the future. The equations used for modeling these kinds of systems are called the Functional Differential Equations (FDE) (Kim, 1999).

Therefore, we allow for additional activations present at time t related to x_j that come as an integrated "memory" effect (not in a cognitive sense used elsewhere in this text) from intensities that existed in the past at time t', i. e. $t' < t$. In addition, there may also be some integrated influences on x due to possible presence of other objects in the conceptual space at $x' \neq x$. Thus, we may write that $\Delta_3 \equiv C(x, x'; t, t')\Delta t$.

Thus, we have:

$$\Delta_I = \Delta_1 + \Delta_2 + \Delta_3 = (S(x,t) + C(x, x'; t, t'))\Delta t \tag{6}$$

Finally the evolution equation 1 , 2 and 3 can be written as:

$$\Delta I/\Delta t = S(x,t) + C(x, x'; t, t'),\ where,\ \Delta I \equiv I_{new} - I_{old} \tag{7}$$

By adequately reducing the time step $\Delta t \longrightarrow dt$, the time evolution takes a realistic form of a smooth function rather than a sequence of peacewise sequence of steps of the length Δt. Consequently, the Eq. 3 reduces to the first order DE in time t, with a parameter x, which is the essential equation of dynamics to be used in CDIE.

$$\frac{dI}{dt} = -\alpha\,(t,x)\,I\,(t,x) + S\,(t,x) + C\,(x, x'; t, t') \tag{8}$$

where $\alpha > 0$.

In sum, the change of the intensity of a an activation (percept, memory, etc.) depends on the intensity itself, on added external sources of activation such as other words presented in a list, as well as on the influence of internal "memories" that make a context for the evolution. This mathematical model is implemented in MATLAB to simulate various experiments in the episode processing research. External inputs can be presented at different rates in time and at different positions corresponding to different stimuli, memories, attentional mechanisms, etc. This allows for simulations of many tasks used in the episode processing research. Additional constraints on model parameters may be posited this way and the model itself may be a unifying account of many findings.

At this point, it is important to mention parameter fitting in CDIE and similar models. Mathematically, the model is very complex and implementation of the model in MATLAB necessarily introduces additional issues related to parameter fitting. However, fitting experimental data quantitatively is not the only goal of this research whose main theoretical purpose is supporting the assumptions of these models. Qualitative fits are more easily obtained in this modeling and may suffice for most purposes of DE modeling. This work follows the tradition from previous research- mentioned above- and uses qualitative fits. Here I refer readers to Erlhagen and

Schöner (2002), for details on parameter fitting issues in these kinds of simulations and to Brown, Preece, & Hulme (2000, p. 172) for a discussion on complex models' number of parameters. In addition, this model, unlike the DFT, includes a non-linear parameter in the decay term, which makes issues of parameter fitting even more complicated.

The simplest version of the model has three dimensions. It is important to note here that these dimensions are not used in manner equivalent to use of dimensions of neural networks. Here, the dimensions are not the number of nodes of a network. For modeling purposes, the three dimensions are on the real numbers scales, which is infinite and uncountable, and this would render them infinitely dimensional in the terminology of most neural networks researchers. One dimension represents a Concept Space (CS). Some concepts are similar to each other, while others are very different from each other. Their distance from each other on the x-axis or the y-axis in this model describes this. It should be noted here that the name Concept Space here is intentionally very broad to possibly include many dimensions along which variously defined notions of concepts may differ from each other (e.g., phonological, semantic, orthographic, etc. differences). Fig. 2 shows a single Gaussian function, an example of intensity of activation varying across one dimension. Several active concepts (activations) may evolve simultaneously, each point's intensity changing according to the model's general equation. The CDIE is a combination of two one-dimensional ("1D") fields, the second one being the order space, mathematically equivalent to CS. In figures, it is called Time-As-Space dimension (TAS), or theoretically more precisely, the "Order in Time As Order in Space" dimension. The third dimension corresponds to the intensity of each activation. This is represented as a height on the vertical axis and/or with different color in figures. Next, the evolution of each activation happens in time which is the third dimension and will be represented on another horizontal axis in some, but not all, figures (as noted in figures and text). Namely, the real-time dimension remains essential, the system evolves in continuous time, and it will be closely related to the TAS, but it will only implicitly be present in figures with the TAS dimension shown. The notion of the field here is strictly mathematical; the set of real numbers is a mathematical field; it satisfies a certain set of mathematical properties. This is the place of integration of activities. Theoretically, in further development of this model, from their simple sum at any point in time, only selected (activated above some threshold) traces get registered in another one or more "2D" fields. It is likely even more than one of these, on different time scales, will then have their own dynamics (at this stage of model development, those fields dynamics are at present not the focus, but see Grossberg, 2013). Finally, in interpretations of CDIE simulations, it is assumed that the intensity. "I", of an activated position is directly proportional to, but not solely responsible for the probability of this activation being recorded elsewhere, recalled later, and to the speed of its recall. Everything else being equal, the more intense activation is both more likely to be recalled and faster to recall than a low intensity activation.

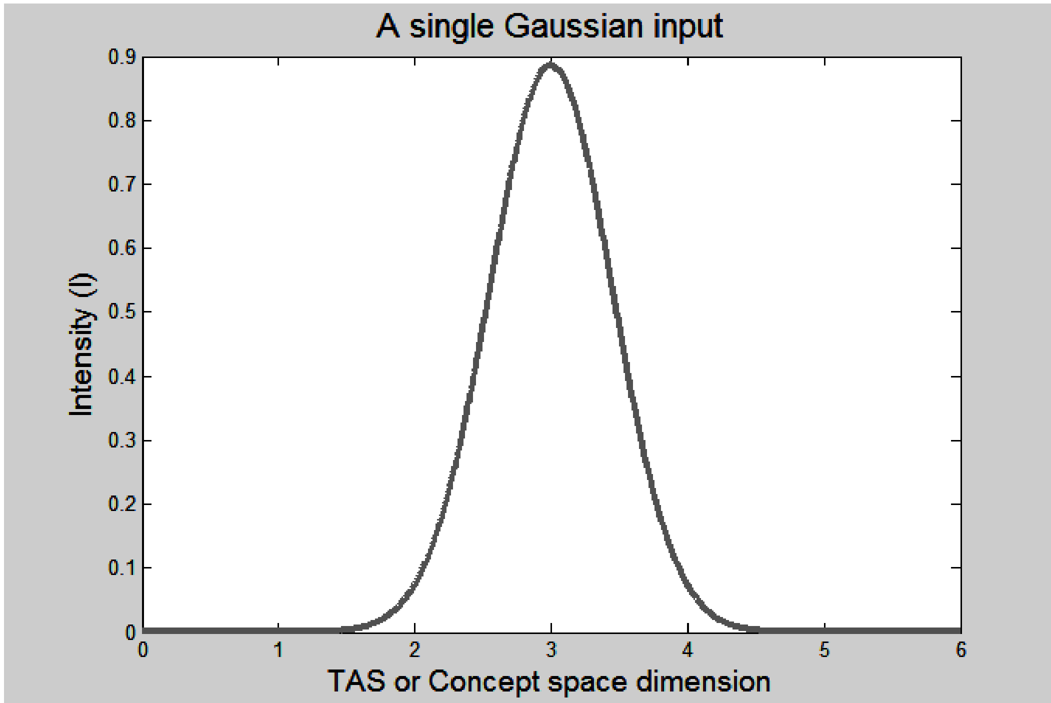

Figure 1: A single Gaussian input. Every concept is modeled by a Gaussian input with its spread along the parameter dimension, "x", to acknowledge that activating one concept usually means activation of several more psychologically related concepts and that this activation weakens as concepts become more different (distant).

The following figures illustrate the "2D" field. Note that "1D" and "2D" names refer only to the number of free parameter dimensions for a task (*and not to how multiple dimensions are combined*) - 1D includes only the CS or the TAS dimension evolving in time, while the "2D" includes both. *In the CDIE model these two dimensions are summed up, they are not a coupled system.* In both cases the mathematical dimensions of time and activations are also present. Furthermore, they *do not* show development of an activation from left to right along the TAS dimension. The entire plane TAS x CS with all its peaks changes in every moment in time. Figures 2 and 3 are snapshots of these changes at one arbitrary moment in time. The TAS dimension represents the order in time as order in space- it is essentially not a time dimension.

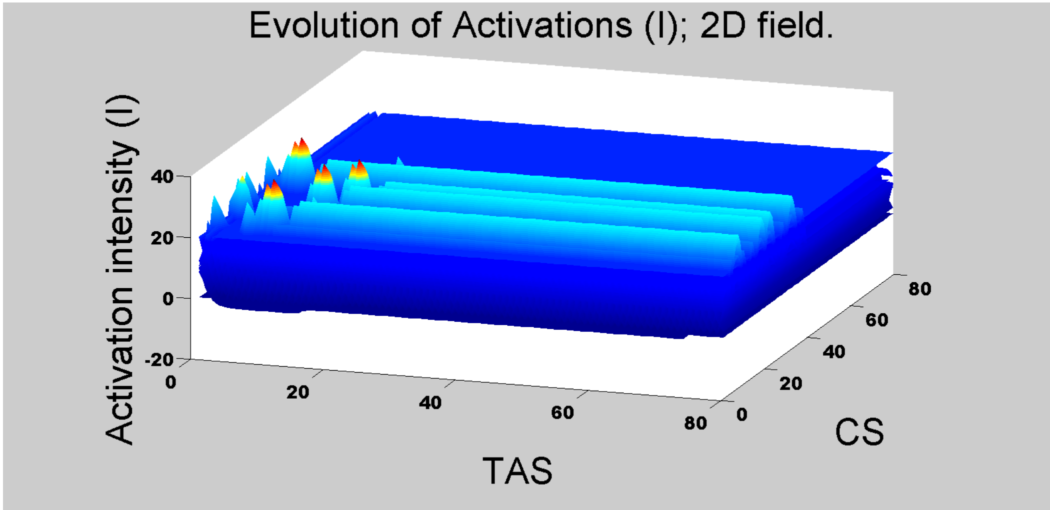

Figure 2: A 3D snapshot one time point of the "2D" field development. On the vertical axis, the intensity of a SUM of two fields is plotted (TAS + CS) and this intensity at each point changes at each time step; the figure shows only one of many distributions of activations during the evolution. Each activation has two coordinates: the position in some ordered sequence of events and the position in the conceptual space. Only the highest activations will be the inputs into another field with a different time scale that will be used during recall, for example.

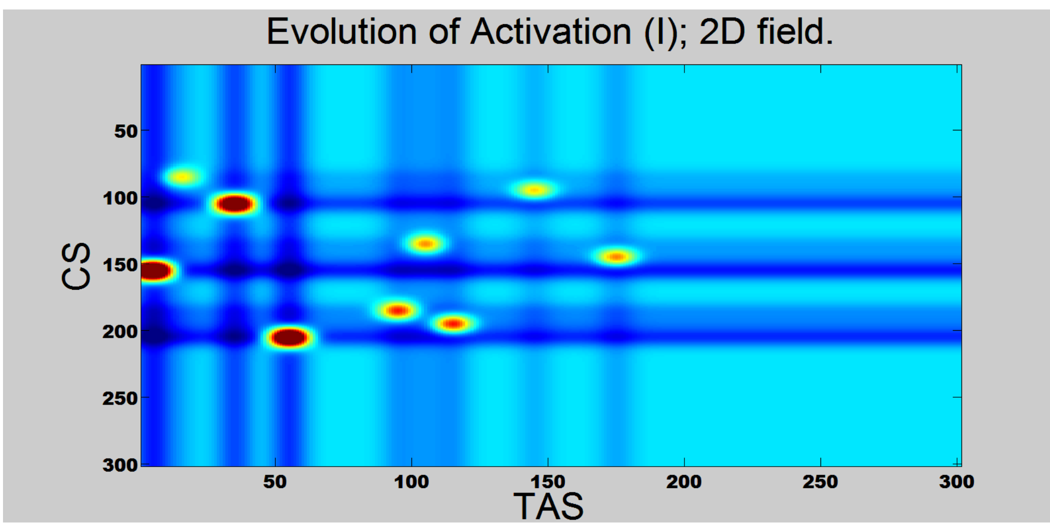

Figure 3: "2D" field. This is, again, a 3D snapshot one time-instance of the "2D" field development, as in Fig. 2, except that the activation is color coded instead of the vertical axis being visible. Only the highest activations will be inputs into another field with a different time scale that will be used during recall, for example.

In this dissertation, it is argued that the "episode processor" (integrated episode processing mechanisms) is a complex dynamical system. The "episode processor" should here be understood as a "dynamically organized and modulated system" as explained by Bechtel (2012, p. 14):

> [...T]he brain is organized as a small-world network. This still allows for local clusters to specialize in performing different operations and for researchers to refer to them, but the components are also regularly modulated by activity elsewhere in the brain. Such interactions do not entail a holism that defeats reference to localized components, but a perspective of a dynamically organized and modulated system in which the component operations are contextually modified by activity elsewhere. The initial assumption of a hierarchically organized nearly decomposable system must be modified, in the course of research, to take into account the sorts of modulation that occurs in a small-world network. These modulating effects are represented in equations whose solution reveals the temporal dynamics of the system. Within the context of such dynamic mechanistic explanations, localizing functions still plays an important role, but it is only an initial step and the resulting localization claims must be modified as researchers recognize how the whole mechanism functions in time by modifying the operation of its own constituents. The reference to parts and operations in the brain needs to be couched within a dynamic perspective in which both the parts and operations change through time in complex ways.

2.3.2 The dynamical systems and CDIE model

Complex dynamical systems are usually aggregates of a large number of elements. In this case, these elements are input information traces combined with memory, attention, etc., which produce activations at specific points in time for specific concepts. The numerous elements of a dynamical system also interact with each other in numerous ways. These interactions are very complex because the value of any activation's intensity at any point in time influences, in a more or less direct way, every other activation peak's intensity. It is argued here that this complexity has to be taken into account to further explore relevant cognitive phenomena. If the model proposed here was implemented in a connectionist network, it would be most closely related to the Adaptive Resonance Theory (Grossberg, 1976; Carpenter, Grossberg, Markuzon, Reynolds, & Rosen, 1992). In addition, the model also may be seen as sharing some hierarchical organization ideas with the Oscillator-based Memory for Serial Order model (Brown, Preece, & Hulme, 2000), by hypothesizing multiple systems in which activity traces may be recorded, with each system possibly having different parameters or even different dynamics (see also, Bromberg-Martin, Matsumoto, Nakahara, & Hikosaka, 2010; Ulanovsky, Las, Farkas, & Nelken, 2004.) Finally, the present model's dynamical equations inherently models ordering effects of events –

which accommodates ideas from recent developments in theoretical cognitive science of human judgment (e.g., Busemeyer, Pothos, Franco, & Trueblood, 2011).

2.3.3 CDIE details and solutions

This model considers the temporal dynamics of episode processing in a multidimensional parameter space. The first concept in the model is that of time, representing a parameter in the system: The system evolves in time. This can either be a real physical time "t" or it may be taken as an individual's psychological time (PT) in which case their relation has to be additionally specified, $PT = f(t)$. In the model, the parameter "x" represents the position of a concept in the Concept Space. For example, one can look at the "x" as a result of the following hypothetical experiment:

> A person is given a stimulus, say a color red, and is asked to make a list of somehow associated concepts. The person is asked to place the concepts in a sequence according to their sense of closeness or distance from the presented concept. In this way, it is possible to place some objects to the left and some to the right of the original concept, depending on the individuals sense of mutual relationships between the associates. For example, colors green and orange may be at the opposite sides of red. Thus, one can introduce the parameter x which defines the position of a concept relative to the initial associate located at some arbitrarily chosen position $x = x_0$. The quantity one can now call the distance between two concepts is therefore given by $x_0 - x$. Similarities between concepts are represented by the spatial proximities of their positions.

The concept of distance described bares a weak resemblance to the distance in physics, but some assumptions about relative positions and similarity can be reasonably put forward. The literature on distinctiveness and psychological distance, in several meanings of this phrase (e.g., Eysenck & Eysenck, 1980; Shepard, 1987; Chater & Vitanyi, 2003; Trope & Liberman, 2010) suggests that concepts indeed differ on various dimensions and that at least some measurements of distances are possible. In the Small World networks research, sometimes the spatial distance between concepts is ignored, which may seem to suggest that the distance parameter in this model is unrealistic. However, cognitive structures even when conceptualized as small world networks (Watts & Strogatz,1998) are still networks *with delay* (Yang, 2002) that allows for various measures of distance (not necessarily Euclidean) and reduction of dimensionality (e. g. Tenenbaum, de Silva, & Langford, 2000).

The use of a one dimensional continuous parameter space to represent a multidimensional concept space is a simplification and only a model of that space. Strictly speaking, some information is necessarily lost in doing this, but the model is mathematically more convenient and computationally simplified this way, while there are justifications for the claim that it is still functional enough to model data and give insights in memory processes.

First, one may conceive that even though there are multiple relations between objects in multidimensional space, in reality we do not use all the information and relations pertaining to one object when we do, say, serial recall of words in the task. Therefore, a one-parameter dimension that may be used for specific modeling task can be thought of as a rotation of a multidimensional space to align the parameter dimension of the model with a chosen single dimension along which the items differ in the most relevant way for the current task only. For any given task, the appropriate dimension(s) used may vary.

Second, for the purpose of this modeling, it is sufficient to determine whether the items in the "to be remembered" list, for example, are facilitating recall of each other. The exact distances are not very intuitive in this kind of research unlike in, say, physics, but this might not be making the modeling too simplified so that the results are completely non-informative of memory and related processes. Because in experiments similar results are obtained for a wide variety of lists, it seems that indeed the items are independent enough in this sense, so that aligning them along one dimension does not significantly distort results.

If the preceding discussion has taken into account the thought experiment presented above, it seems reasonable to conclude that at least the described ordering of inter-item distances along the parameter dimension in CDIE is acceptable for this modeling, and careful consideration of choices of elements of the model. As in any modeling work, one has to be aware of situations where assumptions and approximations may crucially influence the results and interpretations.

The next aim is to model the temporal evolution of an activation point's intensity. In the most general case, the quantity whose temporal evolution we are interested in will be called Initial Pattern Intensity (IPI) and let its mathematical designation be a positive function $I(t, x) > 0$ given by $I_0(x) = I(t, x)$. The activated objects in CS show up as peaks in the initial profile of function $I(t, x)$. In the simplest case, this function is some flat general activation level but it is not necessarily such in episode processing.

As described above in more intuitive terms, to find a suitable mathematical formalism that will model a typical process of time evolution of an initial intensity distribution, one starts from setting up a simple first order differential equation in time t for the IPI function $I(t, x)$. The rate of change of the IPI can be positive or negative depending on whether an object is "gaining" or "losing" activation in time which in turn depends on certain factors, each of which may be represented by a suitable analytical model term.

First, due to the decay processes, the memory intensity $I(t, x)$ spontaneously decays in time if no additional influences are present. This decay/diffusion part in the model, although using mathematics of diffusion common in science, is not used for the same purposes as in the diffusion models of accumulation of evidence for decision making involved in various tasks (e.g., Ratcliff, 1978; Ratcliff & Starns, 2013).

As is common in numerous natural decay processes, one can assume the following rate of change:

$$\frac{dI}{dt} \;=\; -\alpha\,(t,x)\,I\,(t,x) \tag{9}$$

whose analytical solution is:

$$I\,(t,x) \;=\; I\,(x,0)\,\mathrm{e}^{-\tau(t,x)} \quad \text{with} \quad \tau\,(t,x) \equiv \int_0^t \alpha\,(t',x)\,dt' \tag{10}$$

The quantity $\tau(t,x)$ must be positive, i.e. $\tau > 0$, and possibly can be considered as "a psychological time interval."

This part of the equation describes an exponential decay of activation at the reducing rate. Many activations will decay only to some non-zero level. The reducing rate trend of the decay may model consolidation phenomena in memory – after initial rapid forgetting, the remaining activations do not change much in a long period of time.

The term $-\alpha\,(t,x)\,I\,(t,x)$ from the Eq.(4.6) models the evolution of the IPI $I(t,x)$ whose rate is controlled by the coefficient $\alpha\,(t,x)$. This means that the IPI tends to vary in time but not necessarily in a uniform way i.e. the process can run differently at different instants of time t as well as for different "objects" related to the variable x. Mathematically speaking, the coefficient a may also change sign and be negative in some domains of variables t and x, which would indicate a spontaneous growth of the IPI. This change could, for example, model a process of spontaneous concept recall. In any case, the functional dependence $\alpha = \alpha\,(t,x)$ remains arbitrary in the CDIE and can be specified either according to some realistic assumptions, or be a result of some experimental work. The α part of the decay term is used to model attention (see below).

Second, external local source of excitation affects the time evolution of the IPI. This is modeled using the source function $S\,(t,x)$ added to the right hand side of Eq.(4.1):

$$\frac{dI}{dt} \;=\; -\alpha\,(t,x)\,I\,(t,x) + S\,(t,x) \tag{11}$$

As mentioned, this external input may be described by a Gaussian function along the parameter dimensions (CS and TAS) to allow for the idea that one specific activated "object" may also intensify very closely related to the other ones to some extent. The $S\,(t,x)$ modifies the initial IPI both in time t and in CS or TAS dimensions. Clearly, such a process of "activation gain", requires $S\,(t,x) > 0$ i. e. a positive contribution to the IPI growth rate. In the reverse case if $S\,(t,x) < 0$, one can think in terms of a "activation drain," i. e. some process that intensifies the negative time derivative of $I\,(t,x)$ in Eq.(4.3).

In episode processing modeling, the input term, S, provides a counterpoint to the importance of the initial condition in a deterministic system. While this may be seen as a problem in some contexts, it can also represent an advantage for the system. For example, if a person has some initial distribution of knowledge, it can be manipulated by practice, for example. This may result in more adequate knowledge but also with bad training may lead to misconceptions.

Finally, we have a possibility for non-local correlation effects on the IPI in time and space which is a generalization of the equation first used by Amari who considered non-local space effects only.

$$\frac{dI}{dt} = -\alpha(t,x)\, I(t,x) + S(t,x) + C(I;t,x) \tag{12}$$

where:

$$C(I;t,x) = \int_{-\infty}^{\infty} dx' \int_{0}^{t} dt'\; W(t,t';x,x')\, I(t',x') \tag{13}$$

is the spatial and temporal correlation represented by the integration over x' and t'. The functional dependencies W are the weight functions determining influences of the temporal and spatial context at (t',x') on the selected item at (t,x). Their analytical expression has to be specified in the model either following some experimental results or by logical expectations about how the processes should run. For example, in reported simulations the model uses lateral inhibition at greater distances and excitation at smaller distances between items. In addition, small lateral activations at the x' and t' contribute less to the activation in x than greater activations (both negative and positive) while at the same time there is also a limit to the possible contribution of the large activations from other positions. This is achieved by using a Gaussian weighting function for distance contributions and a log function for intensity contributions from a site x' to the site x and from a point at t' to the point t. An issue with this model should be noted here. The second integral in Eq. (4.4) indicates that the system has some explicit memory of its past history. The weighting function along the time axis, W_t, covers a finite time interval of an episode modeled and is of an appropriate shape so that the system is potentially implementable in the brain. However, as observed above, this is an abstract model and I am neutral about these implementation issues. Note, too, that rather than using a typical constant time interval for change of the IPI, the CDIE introduces the attenuation rate coefficient that may depend on both time t and coordinate x which makes this model more general than above mentioned closely mathematically related models.

2.3.4 "Order in Time As Order in Space" idea from cognitive linguistics

How might temporal order of, say, list items be remembered? Despite the fact that philosophers and scientists such as Aristotle, St. Augustine, Zeno of Elea, Kant, Mach, etc. have asked and tried to answer questions about the nature of time and its relation to the mind, these remain largely unanswered. A psychological time has long been thought of as possibly distinct from the physical time (John Duns Scotus) and its relation with physical time is being researched.

A majority of episode processing in general, and specifically serial recall, experimental studies do not elaborate on what information might be included in a memory activation (information) and how interactions of information occur. However, if the activation contains many kinds of information, then any theory involving the concept of an event trace and experiments derived from that theory need to include specifics of evolution or behavior of any one information-containing part of the trace. With the development of mathematical models, these details have to be specified more precisely either by stating that there is no reason to think about a specific order remembering mechanism or, if an event trace carries information about both the position of a word in some mental space and serial order in a list, by specifying at least the interaction and evolution of these two.

2.3.5 Arguments for a distinct order information

The work of Keppel and Underwood (1962) comments on a phenomenon of intrusion in serial order experiments. Namely, the authors describe that in their experiments the items confused with each other were those with similar positions within their respective contexts. Say, for example, the third item from one list is confused with the third item from another, even when they are not otherwise similar. They explain that the design they have used was not suitable for examination of mechanisms of intrusion but suggest that this finding warrants further investigation. It seems obvious that some sort of mechanism for order processing may be useful in episode modeling.

Henson (1998) makes a distinction between three kinds of order information coding: serial chaining, positional chaining, and ordinal coding. First, the chaining models (e.g., TODAM, Lewandowsky & Murdock, 1989) store associations between items in a list from which the order may be reconstructed. Second, the ordinal models (e.g., the Primacy Model, Page & Norris, 1998; OSCAR, Brown et al., 2000.), assume that the information about the order may be reconstructed because the successive items in memory are stored according to their relative strengths. Namely, in these models it is assumed that each successive item is encoded as less activated than the preceding one. Finally, the positional coding models assume some type of a tag that marks the position in a context (e.g. Ladd & Woodworth, 1911; ACT-R, Anderson, 1993; Start-End Model (SEM), Henson, 1998; Perturbation Model, Lee & Estes, 1981; Positional Distinctiveness Model, Neath, 1993).

The need for the distinct order information has been pointed out in many recent arguments from several memory research areas. For example, Dennis and Humphreys (2001) in recognition memory and Hitch, Fastame, and Flude (2005) in serial recall learning argued that the item position information has to be included in models involving serially ordered items but that learning of a list involves probably more than just including a simple tag. Recently, Shankar and Howard (2010) presented a model with a separate mechanism, "the timing cells," which is capable of reproducing the sequence learned (TILT model).

Item-order-rank models developed taking into account the Lashley's suggestion that list items are recorded in distinct neural populations in the brain, Grossberg

(1978) suggested in his CN model that the spatial order is remembered by means of the Primacy gradient: the node, representing the first item is activated the most, and this encodes its position. After it is reactivated at recall, it is inhibited. Since these early works, it has been suggested that this mechanism alone is not enough to model more complex list processing, such as, when the same item occurs at two places in a list (Farrell & Lewandowsky, 2004 ; Grossber & Pearson, 2008, Silver & Grossberg, 2011). They suggested that something playing a role of "a tag" is needed to deal with the rank of an item in a list when items occur more than once. Importantly, in their Item–Order–Rank model of spatial working memory, Silver and Grossberg, (2011) incorporate influences of brain areas involved in counting into several other brain areas' processing of both spatial and temporal sequences, taking into account various electorphysiological findings and behavioral studies in monkeys and humans using various research tasks. In other words, the "Counting Cells" of the superior parietal lobe (Sawamura, Shima, & Tanji , 2002) along with several other brain areas involved in vision support both processing position in spatial and position in temporal sequences. The interaction between brain areas in this model (supported by experiments) are by no means simple and serial, but they seem to support an extremely close relation of time and spatial processing, which is in line with suggestions from other areas of cognitive science.

In the SIMPLE model (e.g., Lewandowsky et al., 2004), the authors make a parallel between their logarithmic scale for time and Weber-Fechner Law or Stevens Law of psychophysics. It is less parsimonious than a simple tag but it might be saying something about the mechanism of time handling by memory. This may look like continuity from the psychophysics that accounts for the data. However, while these laws relate to sensory transformations in psychophysics, the authors do not show why those would apply to the time dimension; we do not directly perceive time. As mentioned elsewhere, the model was later extended to include the within-list order position, of a form of a simple tag. This way, the model deals with both temporal distinctiveness and positional distinctiveness, namely, two dimensions each carrying information. (Lewandowsky et al., 2004). For now, I would like to note only the interesting way the authors explain how to position items on this log scale: "[...] the positions of the memories will appear compressed for the temporal perspective of retrieval (specifically, after logarithmic transformation of the temporal distances; right justified to "the present")." (italics not in the original text). This is an example of a Time-as-Space conceptual metaphor investigated in cognitive linguistics.

In sum, it seems that all theoretical accounts and models of episodic memory assume some kind of distinct mechanism that deals with the time passage during an episode and the order of events in it. Many of these accounts stem from the nature of implementation of the framework the authors work in and are, in a sense, limited by it. For example, the neural network models talk about groups of neurons that work in a certain way so that they can compute some kind of timing information. They often do not discuss in detail the phenomenon of episodic memory in more abstract terms, such as memory space dimensions. Often neither the nature of the time dimension is a primary focus of these models nor is the focus the interaction between this and the other dimensions.

In short, current research seems to suggest that a separate dimension or a mechanism that deals with something related to time needs to be included in models of episodic memory. More specifically, the order of stimuli in time is needed and it seems that the episodic memory requires more complex considerations about relations of memory dimensions in order to account for experimental data.

2.3.6 CDIE order information

The CDIE imports ideas from the field of cognitive linguistics to treat order information.

"The word [tempus] referred originally to space; the meaning 'time' is later, and came about in this way: the quarters of the heavens are thought of as corresponding to and standing for the parts of the day and year; east is morning, south noon, and so on." (Allen, 1880, p. 140; cited in Casasanto, Fotakopoulou, Boroditsky, 2010)

A Theory of Magnitude (ATOM; Walsh, 2003) argues that a shared brain mechanism handles space, time, and quantity (number) through treating those as magnitudes. However, the cognitive linguistic research seems to show that the time dimension rests upon the spatial dimension, and not vice versa. In other words, time – as more abstract concept than the space, with which we have more direct physical experience – may be conceptualized in terms of more familiar spatial terms (Talmy, 1988).

Recently, Casasanto and Boroditsky (2008) performed a series of experiments to investigate the idea that time and space are not only connected in language but that temporal information rests upon a more basic spatial representation. The evidence for the claim that space is a more basic concept than time in cognitive linguistics comes from the fact that in language there are more cases where time is presented as space than the other way around (e.g., Boroditsky, 2000, 2001; Piaget, 1927/1969; Tversky, Kugelmass, & Winter, 1991).

In the CDIE framework the order information (or any additional postulated information / dimension) is introduced as a separate mechanism. The idea that the position is related to intensity of an activation (such as Grossberg's Primacy gradient mentioned in more detail below) at first seemed most suitable and the simple CDIE simulations can accommodate it easily. However, the initial idea was abandoned because as soon as the simplest assumptions were replaced with just slightly more complicated ones – it was impossible to obtain correct order of item just based on their intensity. For example, as soon as the external inputs are not all of the same intensity, even though the reducing "encoding strength" was added, it was impossible to infer the order based solely on their residual activation in any way. The additional mechanism mentioned interacts (in a more or less complex way) with the first one, the CS dimension. Both of these can be hypothesized to be of a similar nature and to evolve in continuous time. The information that a specific word was also the fourth in a list, for example, may evolve in time so that it is very hard to retrieve it after some time period, in the correct position or overall. In the CDIE the two dimensions do not interact – they are modeled by separate differential equations but a more complicated system of two coupled differential equations can further model joint evolution of the two (or more, in principle) kinds of information a trace holds in an episode processor:

$$\frac{dI_1}{dt} = -\alpha_1(t,x)\,I_1(t,x) + \beta_1(t,x)\,I_2(t,x) + S_1(t,x) + C_1(I_1;t,x) \tag{14}$$

$$\frac{dI_2}{dt} = -\alpha_2(t,x)\,I_2(t,x) + \beta_2(t,x)\,I_1(t,x) + S_2(t,x) + C_2(I_2;t,x) \tag{15}$$

In principle, this kind of system has multiple types of possible solutions. In addition to types of solutions present when only one equations is solved, the system may have oscillatory solutions as well. On the other hand, if during any time period the evolution of the order information is such that the intensities almost do not change, one gets a situation where one can say that there is a "memory tag" for each word and practically does not have to specify it in more detail. In this sense, the ACT-R component that assumes memory tags, for example, becomes a maximally simplified version of the CDIE order information. In principle, however, it is obvious that decay, interference, etc. theoretically apply to this dimension too, unless one clearly explains why not. Indeed, just a quick survey of the cognitive linguistics literature shows that the two dimensions have different behaviors (e.g., Bjork & Healy, 1974; Henson, Hartley, Burgess, Hitch, & Flude, 2003). Future work with the CDIE model is aimed at investigating these issues in more detail. For example, if information from these two domains is conflicting, it may impair time processing (which is more abstract) more so than spatial processing. Indeed, this is what research in cognitive linguistics is finding (e.g., Matlock, Ramscar, & Boroditsky, 2005; Casasanto, Fotakopoulou, & Boroditsky, 2010). In sum, from the CDIE modeling perspective, it seems reasonable to treat the new dimension of episodic memory (e.g., in serial recall) as the space - order dimension that represents order in time. The simulations presented later will clarify more details of the model later.

2.4 Differential equations in modeling changes in systems elsewhere in science

Differential equations model dynamics of processes: most often, but not necessarily, the temporal changes. These equations are used in modeling changes in complex interactions of a system's characteristics and not mainly to describe the results of behaviors at one time point. As mentioned earlier, every mathematical model gives some insight into processes themselves but DE models typically at the very beginning of modeling take into account more realistic variables with corresponding changes, as well as their mutual interactions and thus may give additional important insights into processes of a system. In science, having the ability of taking into account both measured quantities and *changes in quantities* is more often than not advantageous.

There are many famous and widely used differential equations that model natural processes. For example in physics, Newtons laws from 17th and 18th centuries relate mass of a body, forces acting upon it, and its velocity and acceleration. The radioactive decay equation gives a relationship between half-live of radioactive material, spontaneous decay rate, total activity of the material, number of decaying particles in a sample, and specific activity of the material. Using this equation allows us to model how a specific sample of radioactive material changes in time. Einsteins field equations are equations relating Newtons gravitational constant with shape of time and space that are curved by energy and mass (the equation contain more than these elements but that is not crucial here). Another famous of DE are Schroedinger's equations (1926). They are used to probabilistically predict future states of a quantum system based on current states. In economics, Reverend Thomas Robert Malthus published his work on relationships between growing population rate, labor supply, and wages as they change in time, in 1798. In 1890, Gabriel Tarde studied how ideas and behaviors diffuse (change in time) in a society as the result of forces of imitation and innovation. The Sethi model (Suresh P. Sethi, 1983) gives the change in time of a relationship between advertisement and sales. In biological sciences DEs are also widely used. The famous Hodgkin-Huxley model contains several nonlinear ordinary differential equations and is used to model changes in behavior over time of so called "excitable cells" like heart cells and neurons.

I use another model, the Reaction-diffusion systems, to illustrate the above mentioned difference in mathematical models and to show what kind of a model is presented in this dissertation. The preceding examples illustrate that DE modeling nicely describes changes of some complex system by predicting their next state based on their current state and by using select characteristics of that system. These equations are used to simulate complex interactions of a systems characteristics (variables) and not mainly to describe the results of behaviors at one time point. As mentioned earlier, every mathematical model gives some insight into processes themselves but DE models at the beginning of modeling take into account more realistic variables (derivatives) and thus may give more insight into evolving changes in a system.

Consider a physico-chemical reaction: the diffusion model. It relates concentrations of substances that are distributed in space, transformations of those substances, and their spread in space, and describes the time changes of this system. The use of variations of this mathematical model is widespread in ecology, biology and so on. Here is a partial abstract of a study: A reaction-diffusion wave on the skin of the marine angelfish Pomacanthus, by Kondo S. and Arai R. Published in the journal Nature in 1995:

> The marine angelfish, *Pomacanthus*, has stripe patterns which are not fixed in their skin. Unlike mammal skin patterns, which simply enlarge proportionally during their body growth, the stripes of *Pomacanthus* maintain the spaces between the lines by the continuous rearrangement of the patterns. Although the pattern alteration varies depending on the conformation of the stripes, a simulation program based on Turing sys-

> tem can correctly predict future patterns. The striking similarity between the actual and simulated pattern rearrangement strongly suggest that a reaction-diffusion wave is a viable mechanism for the stripe pattern of *Pomacanthus.*

The authors used the reaction-diffusion wave model to simulate and predict changes in patterns on the skin of a fish. The equations they used are:

$$\frac{dA}{dt} = c_1 A + c_2 I - D_A \frac{d^2 A}{dx^2} - g_A A \tag{16}$$

and

$$\frac{dI}{dt} = c_4 A + c_5 - D_I \frac{d^2 I}{dx^2} - g_I I \tag{17}$$

and the fish looks like this:

Figure 4: Marine Angelfish
(Albert Kok. *Pomacanthus imperator.* [Digital photography]. Retrieved from Wikimedia Commons, http://en.wikipedia.org/wiki/File:Pomocanthus_imperator.jpg)

It is obvious that there are many characteristics of the system (fish) taken into account in this modeling. For example, the quantities "A" and "I" are concentrations of certain molecules, the "D"-s are diffusion rates, the "g"-s are decay rates, etc. The researchers were able to simulate the changes on the skin of the fish very nicely: Picture "d" is the fish skin pattern and the "g" is the simulation using the model.

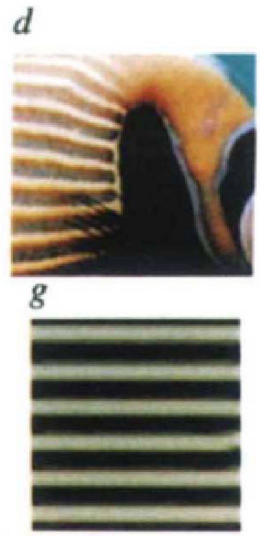

Figure 5: Fish stripes as they appear on the fish Pomacanthus (picture "d") and as obtained by process simulation using the system of equations 5.12 and 5.13 (picture "g"). (From Kondo & Arai, 1995, Nature. Author permission obtained. Publisher permission in progress.)

This work was more extensive than the points presented in this dissertation. I use this information to support following point: the pattern from the picture "d", say, along just the left vertical edge (axis) could have been described by a simple periodic function. In psychology, this is how forgetting curves are described. The number of correct recalls at each position in a list is plotted and the line is fitted through the resulting points. The points in that graph would correspond to, say, dark points along the left vertical edge of the picture "d". The resulting graph for the biological example here would look something like the following:

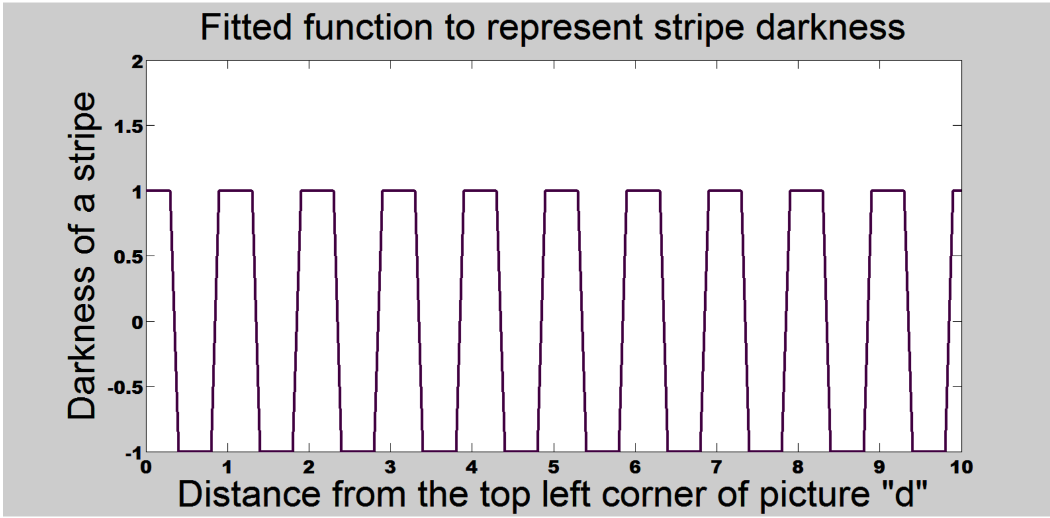

Figure 6: To capture abrupt change in brightness, a periodic function slightly different than the sine function can be used to fit the change of intensity of a dark color along one direction on the skin of the fish Pomacanthus. Variables such as decay rates and concentration do not play a role in this fitting.

On the other hand, picture "g" from the above is a result, (an emergent pattern), of the complex system of chemicals evolving according to the system of two given differential equations. In terms of scientific understanding of the process in a living tissue, the DE model and simulations seem to be a lot more informative, although the fitted periodic functions provide some hints about the underlying processes. The model presented in this dissertation, the CDIE is analogous to this dynamical approach. By using DEs and simulations, it tries to obtain the values for variables that can then possibly be fitted, as demonstrated above for the case of the "darkness of skin based on the distance from the upper left corner of the picture "d"." In other words, the forgetting curves graphs most often found in memory literature are analogous to the graphs in Fig. 3, while the model in this work produces the result of the process obtained in simulations in picture "g" (Fig. 5). Therefore, I argue, the DE modeling of forgetting in an episode processing system sheds new light on the processes themselves. This is *a complement* to other insights; they all model the same phenomena from different perspectives, ultimately they should not contradict each other, and they should be used to constrain and expand each other.

3 Time-Based (spontaneous) Decay in Episodic Processing

3.1 Spontaneous decay and the CDIE model

Presented first is an overview of the episodic memory research to show some current issues about time based decay and how have they evolved. Next, I review how current models treat decay, paying special attention to those models that are used to argue against decay. After mentioning ideas from related but different research areas about decay, I conclude that current paradigms cannot logically argue against decay, that most current models actually use decay but some without acknowledging it, and that there are other kinds of evidence for decay. I therefore, argue that the concept of time based spontaneous decay is useful in episodic memory research and has to be investigated further. Finally, I show some ways the CDIE addresses time based decay related experiments. This chapter should cast reasonable doubt, at least, on suggestions that the time based decay is not very important in forgetting. It is argued that:

1. Spontaneous decay should be precisely defined.

2. It should be discussed why verbal memory would have completely different dynamics than non verbal, as suggested by some authors.

3. Current experiments in episodic memory are not specific enough about how decay contributes to the time evolution of memories. This leads to vague predictions about the results of the operation of decay, which in turn only support weak arguments against it.

4. Current models of episode processing include terms analogous to time-based decay of information in a memory trace even when they argue against it.

5. There exist other reasons to consider decay, typically not considered in cognitive science and related areas.

6. Using the time based spontaneous decay issue, I argue that the dynamical systems approach using differential equations is not only useful for simulating episode processing but, even more importantly, is a new perspective in theorizing that introduces fundamentally new conceptualizations of episode processing mechanisms worthy of further investigation. Illustrations include demonstrations that decay may be important for distinctiveness of stimuli (which plays a crucial role in the Word Length Effect, Serial Position Curves, Chunking in memory, photographic memory, etc.) as well as for attention, false memory, and forgetting curves research.

The question whether spontaneous decay, in the psychological literature termed “time-based decay”, is at least a partial cause of forgetting in immediate memory, especially verbal memory, is unresolved (e.g., Brown & Lewandowsky, 2010). Interestingly, in non-verbal memory the operation of this decay is almost not contested (Oberauaer & Lewandowsky, 2011; Ricker & Cowan, 2010). Issues presented below present completely novel analyses and simulations, using DE modeling.

Current literature includes a large amount of arguments against the time based decay in forgetting. To illustrate some problems with these arguments, I specifically note two prominent issues, which result in extremely common but highly unsupported idea among scientists that the time-based decay does not play a major role in, notably, immediate memory.

The first concerns the definition of time-based decay. In current research on episodic memory, the time-based decay is defined more loosely than the general spontaneous decay definition in science and CDIE, at least at the first glance, for large majority of models. There are descriptive definitions of decay, such as recent definition of Lewandowsky, Oberauer, & Brown (2009): “Decay: refers to the notion that memory fades over time without an additional identifiable causal agent.” The decay notion must assume a compensatory process such as rehearsal whenever forgetting over time is absent.” At first this definition makes sense because it points out that processes such as interference may be the actual causes of forgetting. However, the notions of time and the idea of time as a causal agent need a lot more clarification. How does the time need to be conceptualized so that it alone may cause forgetting?

The second concerns the reports of findings, notably those arguing against decay. For example, Oberauer and Lewandowsky, 2011, in their abstract report “The article tests the assumption that forgetting in working memory for verbal materials is caused by time-based decay, using the complex-span paradigm.” and their conclusion is: “The authors conclude that time-based decay does not contribute to the capacity limit of verbal working memory.” Seeming disconnect between the cause of forgetting and the decay’s contribution to WM capacity is clarified in the paper: “The present experiments rule out decay as a major cause of forgetting in verbal WM, and therefore imply that an alternative explanation needs to be found for the cognitive-load effect”(p. 58). However, the title of the paper is “Evidence against decay in verbal short-term memory.” Therefore, the decay is not a major cause of forgetting in verbal WM but there is no support for the claim that there is evidence against its existence and role in forgetting, as the title would suggest. This is not the only paper with this kind of message, for example see: Neath and Brown (2012), Arguments Against Memory Trace Decay: A SIMPLE Account of Baddeley and Scott and Lewandowsky, or Oberauer and Brown (2009): “No temporal decay in verbal short-term memory.”

In the CDIE, the change of intensity of a memory trace depends on several processes. After initial activation of a single memory trace, if there are no other influences this activation will spontaneously decay in time. The rate of decay may be modulated both by the nature of the item and time, say, since the memory was formed. In CDIE model, when spontaneous decay is not present, the following happens – a complete saturation if not an overload of the field. Discrimination of stimuli is impossible.

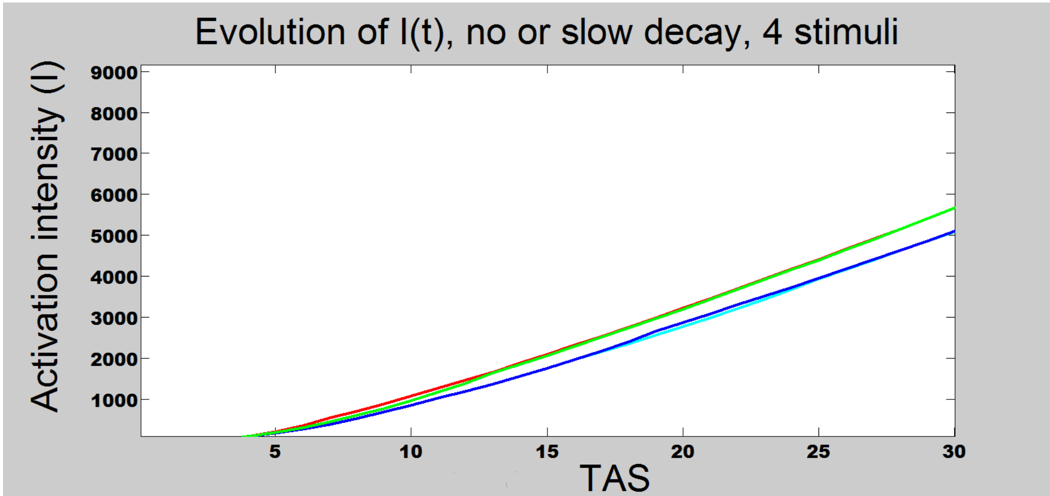

Figure 7: Minimal decay of for stimuli presented one by one to a field with a flat initial distribution. The stimuli are the same in every respect but their timing. This is not essential here but it will be in figures 28 and 30. There, the sole difference from this simulation is in the amount of decay in the system's dynamics.

If there is no decay operating, the intensity at every point in the field increases quickly over time, potentially reaching extremely large magnitudes. This problem is not necessarily present when decay operates. The decay is largely preventing extreme activations because it reduces higher intensities faster and lowers overall activations (and much more than lateral inhibition) so that new inputs do not increase the total activation of the system too much. Altmann & Schunn (2002) made the same points after using their Functional Decay Model to fit experimental data.

While this is the most straightforward result of CDIE simulations - a moderate decay is needed, there are other interesting points emerging from the CDIE simulations, some of which were reported elsewhere, about possible role of spontaneous decay in forgetting which include attention, simulations of the Word Length Effect, Serial recall, Forgetting Curves, Chunking: Temporal Gap Effects, and False memory. These will be illustrated and discussed in the second part of this chapter, after the history and current finding and debates are highlighted.

3.2 Background on decay research

In this section I present some classical experimental results about forgetting of episodes in general and more specifically in serial recall. The Trace Decay hypothesis is introduced with findings that are usually interpreted as supporting the hypothesis and some others which are negating it. It is later argued that the not-supporting statements are not logical conclusions of the experimental procedures used to argue against the time based decay.

What causes forgetting? The experimental memory research is said to have begun with Hermann Ebbinghaus. In 1885. he published results of experiments on himself involving learning and forgetting series of nonsense syllables. Memory, as expected, was degrading with time. Based on his results, Ebbinghaus concluded that a strong, well learned memory lasts longer than a low intensity memory and that the absolute amount of degradation of memory declines over time: After a sufficient time interval some memories are lost but those that survive after that time almost do not change any more. Similar forgetting trends, however, have been later found in almost any memory task on any time scale (see Wicksted & Ebbesen, 1991).

Ebbinghaus experimented with retention intervals that are today associated with the long term memory (the shortest retention interval being around 20 minutes). One could additionally ask what exactly happens with memory within a couple of minutes after learning? This is a short term interval of retention which is arguably supported by a separate memory store, the Short Term Memory (Atkinson & Shiffrin, 1968) or is involved in working memory (Baddeley, Thomson, & Buchanan, 1975; Baddeley, 1986) both being characterized by their authors as having a low span (the amount of material that can be recalled completely, in order of presentation, on 50 % of trials).

In 1958, John Brown put forward the Trace decay theory in order to explain both the origin of "immediate" short-term memorys very limited span and reasons why information is forgotten if the span is exceeded, without postulating two memory mechanisms, the short term and the long term one. He explained what he meant by memory trace (and I accept this concept for my work), decay, and short term forgetting (pg. 12, 13):

> The basic hypothesis of this theory is that when something is perceived, a memory trace is established which decays rapidly during the initial phase of its career. (By memory trace is meant only the neural substrate of retention, whatever this may be.) Some decay of the trace is assumed to be compatible with reliable recall - just as partial fading of print may be compatible with perfect legibility. [. . .] When a sequence of items is presented, the interval between the perception of each item and the attempt to recall that item will depend on the length of the sequence. If the sequence exceeds a certain length, decay of the memory traces of some of the items will proceed too far for accurate recall of the sequence to be possible. This length is the memory span. Thus the trace-decay hypothesis can explain both the origin of the span and why forgetting occurs when the span is exceeded (Brown, 1958).

To investigate decay over different time periods, in the experiment 1 reported in his 1958 paper, Brown presented one group of subjects with two sets of stimuli, the required set and the additional set, for subjects to read them aloud. The required set is the one that was recalled in the test phase. It consisted of one to four pairs of consonants and the additional set consisted of 5 pairs of numbers. All the stimuli were presented at an equal pace with the required set first and the additional set after that. Immediately after the last stimulus was presented, subjects were asked to write down, in correct order, stimuli from the required set. Another group of subjects had an empty but equally long additional set. The results supported the trace decay hypothesis-if the recall is delayed and there is some activity before recall to prevent rehearsal, then forgetting happens even if the number of stimuli is within the memory span.

Similarly, in 1959, Peterson & Peterson performed an experiment in which they have shown that trigrams that do not form words are forgotten in a third of a minute if they are not rehearsed. Their subjects were asked to remember a trigram and then to count backwards by three or four for several different time intervals (3, 6, 9, 12, 15, or 18 seconds) before the recall. The results illustrated that without the rehearsal forgetting increased with time (and not linearly). This parallels Ebbinghaus results on a longer time scale.

Based on these experiments, it seems obvious that without some counteracting activity memory traces decayed to the point where the recall of a learned material was not possible any more. However, it is important to notice the term decay might not be precise enough for communicating the details of this research. Indeed, lowering intensity of a memory trace may be caused by several mechanisms (spontaneous decay, interference, lateral inhibition, etc).

3.2.1 Literature on decay, rehearsal, and interference

Decay. Once a memory trace is formed it can be recalled if it has a sufficient intensity. This intensity gets lower with passage of time, if no "compensatory process" is involved (Lewandowsky, Oberauer, & Brown, 2009) and therefore the probability of correctly recalling an item also decreases. In effect, everything else being equal, the probability of correct recall decreases with time. In addition, any other measure that may be related to the intensity of a memory trace (e.g., reaction time) may change with time, too. Several different kinds of compensatory processes are used in research to show that time based decay does not show up even if these mechanisms are suppressed.

Rehearsal and refreshing. Rehearsal is assumed to improve memory trace negatively impacted by time based decay. Therefore, if the rehearsal is suppressed in some way, time based decay should be evident if it is happening. Unfortunately it is very hard, if not impossible, to come up with experimental paradigms that can clearly isolate time-based decay from other influences on evolution of activations.

i. Articulatory rehearsal. Articulatory rehearsal refers to subjects behavior during the process of learning new material. Most of the subjects continually repeat the material they need to remember. If the rehearsal is not overt, then often it is covert, sub-vocal rehearsal. The subjects deliberately repeat a list of words without saying it aloud. This kind of rehearsal is related to the phonological loop of working memory proposed by Baddeley & colleagues (Baddeley, Thomson, & Buchanan, 1975; Baddeley, 1986). While in the loop, the information, contained in a memory trace, remains activated and the probability of its recall is high. In order to monitor the time based decay processes of memory, rehearsal needs to be suppressed. The articulatory suppression (AS) is therefore one intervention that may be implemented in experiments where one wants to examine time based decay of a memory trace. Peterson et al (1959) used counting backwards to fill the period after learning and the results showed that this hurt recall. Baddeley et al., (1974, 1975) used slightly different articulatory suppression methods during parts of the procedure or during the whole procedure. For example, a word was supposed to be repeated on purpose or a sequence of random numbers was supposed to be generated in order to prevent the rehearsal of the target material. All paradigms using articulatory suppression show that the memory is lowered if the rehearsal is suppressed, supporting the idea of time based decay.

ii. Attentional refreshing/rehearsal. It is possible to imagine that some other mechanisms in addition to vocal and sub-vocal rehearsal may have a role in overcoming time based decay and thus need to be controlled for. Cowan (1992) argued that within the pauses between the recalled words, subjects use the time to mentally scan entire memory set, which is an attention-based refreshing process that could counteract time based decay. In other words, the attention is focused on a word in recall while that word is being recalled but traces of all other words decay during that time. However, when the recall of the word is over, the attention may be used to counteract decay of the other words. The authors note that this mechanism resembles the suggested memory scan mechanism coming from the recognition memory research. It is also possible that this refreshing operates only on one item, not the entire memory set (Johnson et al., 2002; Raye et al., 2002). Waugh & Norman (1965) reported it this way: “In the colorful terminology of one such subject (Waugh, 1961), the most recent items in a verbal series reside temporarily in a kind of “echo box” from which they can be effortlessly parroted back.”

Interference: Proactive and Retroactive. In addition to time based decay and rehearsal, changes in memory traces may be due to influences of other stimuli such as those from previous trials (Keppel & Underwood, 1962), often called the proactive interference, or from the material used to suppress the rehearsal (Waugh & Norman, 1965), called the retroactive interference. Further, currently activated context of a memory trace may interfere with it and thus impede the recall (Anderson, Bjork, & Bjork, 1994; Watkins, 1978) or even create false memories (e.g., Roediger & McDermott, 1995). Proactive and retroactive interference were initially studied in relation to long-term memory. Both refer to effect that one learned material has on another one. In retroactive interference what is learned later influences the memory of previously learned material. In proactive interference current learning influences how the next material will be learned (e.g., Underwood, 1945; Postman 1961). One may ask whether similar interactions also happen in short-term retention phenomena. Here, again, it is clear that experimental paradigms are likely to fail in isolating the time-based decay from other processes in order to make solid inferences about its operation.

i. Proactive interference in serial recall. Keppel & Underwood (1962) pointed out a methodological issue concerning most experiments involving time related to short term memory- most of them (excluding Brown 1958) had used a single item that was supposed to be remembered. In order to provide a continuity with the long-term memory research, they decided to examine the role of proactive interference, and not decay, in short-term memory when lists of stimuli were presented. Their design included three retention intervals (3, 9, and 18 s), and they measured accuracy of recall in case where the critical trigram was preceded by various numbers of other trigrams. Their results showed that when there were multiple items to be remembered, the items that came before the target item interfered with the target items recall. This constitutes evidence for the role of proactive interference.

ii. Retroactive interference. Waugh & Norman (1965) designed experiments to investigate their assumption that both time based decay and interference play a role in forgetting over the short term. The authors presented subjects with 16 digits, either at a pace of one per second or four per second. Subjects were instructed not to rehearse a digit after the next one appeared. At the test, subjects were shown one of the digits and were asked to say which one has followed it in the study phase. Their results show the rate-invariant probability of recall, supporting retroactive interference.The following is the explanation the authors used in interpretation of results:

They interpreted the results to be evidence against decay and for interference.

iii. Decreased distinctiveness. Bjork & Whitten, 1974. suggested that the retrieval from the long term memory may have been led by the ratio between the temporal differences of the list items and the delay period between the study of those items and the recall. In other words, the items that were studied longer time ago were less distinctive from each other (and this constitutes interference) than the recent items. This reduced distinctiveness may have mimicked the effects of time based decay. This is a variant of the temporal distinctiveness hypothesis later used by several authors to argue that the apparent loss in memory is not due to time based decay (e.g., Turvey, Brick, & Osborn, 1970; Brown & Chater, 2001).

3.2.2 Summary of the studies and conclusions

The cited studies represent brief overview of research on forgetting in episodic memory, with more specific details about the forgetting over short time and serial recall. The research begun by testing systematically observations that what one learns does not stay equally available for recollection over the periods of hours, days, or moths after learning. Soon after this initial research, the research on rapid forgetting across a short term interval was developed to explain the nature of this forgetting and apparently limited capacity for remembering during this period. One logical explanation for this was put forward - the time based decay of a memory trace. The memory trace of certain intensity is formed and it then decays if no other influences occur. One possible mechanism to counteract time based decay is to rehearse or refresh the material that needs to be remembered and this mechanism was explored in all phases of episodic memory paradigms. Early studies focused on preventing rehearsal and refreshing in hope that this would expose time based decay. At first, the results were encouraging- the decay hypothesis seemed to be well supported. However, other researchers then suggested other mechanisms that could mimic decay, such as the interference. Some of them introduced these mechanisms without specifying their connection to the time based decay and some made very specific claims about the non-existence of this decay.

While it seems uncontroversial that different kinds of processes may cause some activated information ("activations") to be short-lived, the question about roles of spontaneous decay in memory remains unresolved.

First, one reason for this clearly is the fact that it is extremely difficult to come up with a paradigm that would provide crucial evidence against decay in episode processing. Some of the findings described in literature offer explanations for forgetting which are alternatives to time-based decay explanation. These are not necessarily arguments that show evidence against decay (nor were they necessarily meant to be that, notably Keppel & Underwood 1962). For example, the distinctiveness hypothesis in recall research does not necessary exclude decay as a reason for poor recall; logically, decay may still be present even in situations where the distinctiveness improves recall. As Brown (1958) pointed out "It is important to recognize that such failure of discrimination, i.e., confusion between responses, cannot be regarded as a primary cause of forgetting. Failure of discrimination presupposes forgetting of that which determines which of the responses is correct. It is thus a possible effect of forgetting, however caused, but is not itself a primary cause of forgetting."

Even if decay was not found in the research, it seems obvious that one could speculate about additional reasons why this would be the case. Ideally, one would use a paradigm that includes a single memory trace isolated from all other memories and other influences and its intensity would need to be measured over different time delays. The fact that this procedure is hard to turn into a real experiment is a problem if one wants to gather evidence for decay, but at the same time it is also a problem for those who want to argue against it. Therefore, as long as is not very clear what the existing procedures measure, they should not be the basis for rejecting the trace decay idea and for discouraging decay research. In sum, currently there are no experimental episodic memory paradigms that may logically conclude that time-based decay is not operating in memory in episode processing. On the other hand, there are experimental results that support the time-based decay ideas and do not exclude the possibility that the forgetting is very complex and involves other factors, too.

Second, very descriptive definitions of decay used in experiments pose a large problem for measurement interpretations of results. These definitions capture the intuitive notion of decay but do not provide more specific mathematical description, which may prove to be detrimental for thinking about forgetting. They do not say anything, even descriptively, about, say, how decay of one activation (say, representing a memory) influences other activations and in what time frameworks any interactions may happen. In principle, this means that completely opposite predictions based on this definition may be possible. For example – it may be argued that the decay of a memory helps other memories by not interfering with them but also that it may inhibit other memories because of its strength. This means that interpretation of experimental results may be extremely misleading. In addition, some features of several models, some of which are implemented in neural networks, do have mathematically defined at least some (mathematically) similar functions but do not straightforwardly connect them to or designate them as decay. More definitions of decay seem to be needed in order to achieve more specificity in experiments, simulations, and their interpretations. The mathematical definition presented here is intended to contribute to this goal.

In sum, the following points about decay based on current literature should be noted. The spontaneous decay should be more precisely defined. It should be discussed why verbal memory would have completely different dynamics than non verbal, as suggested by some authors. Current experiments in episodic memory are not specific enough about how the decay contributes to the time evolution of memories. This leads to vague predictions about the results of the operation of decay which in turn allows only for very weak arguments against it. Current models of episode processing include terms analogous to time-based decay of information in a memory trace even when they argue against it. Using the time based spontaneous decay issue, I would like to suggest that the dynamical systems approach using differential equations is not only useful for simulating episode processing but, even more importantly, it is a new perspective in theorizing that introduces fundamentally new conceptualizations of episode processing mechanisms, such as time-based decay in this case, to be further investigated.I return to more details about specific models and arguments later in this text.

3.2.3 Decay in other mathematical models of cognition: present but may be hidden.

In this section I will briefly introduce some current models, especially those that are most often used to argue against decay and examine how they treat time based decay. This will include a survey of the matter of what information might be present in a memory trace and therefore possibly might decay, too.

TILT model

As mentioned earlier, the TILT model (Shankar & Howard, submitted) assumes time based decay of the memory trace. The information about when an item was presented becomes fuzzier because of this decay and because of that it is easier to make mistakes in recall. Note that the order information in this model is implicit. The context of a stimulus is explicit and the entire history is recreated from the context vector when a cue is present. This then allows for the order information to be inferred. As mentioned earlier, it is not clear, however, how much of the fuzziness to expect for a first item in a serial position paradigm, for example. It is expected that this model contain details that are parts of the CDIE model, which assumes that most agree that memory is a dynamic system. For now it is important to notice that the model assumes time based decay and assumes an additional mechanism to keep track of time (and therefore of order). It seems that without this additional mechanism, it is not possible to extract the order information from memory.

SIMPLE model

SIMPLE (Scale Invariant Memory, Perception, and Learning) model rests upon three main assumptions (Brown, Neath, Chater, 2007):

(a) episodic memories in multidimensional psychological space are located along a dimension representing temporal distance from the point of retrieval, (b) the retrievability of an item is inversely proportional to its summed confusability with other items in memory, and (c) the confusability of items along a temporal dimension is given by the ratio of the temporal distances of those items at the time of recall.

If two items are distinct only along the temporal dimension, the farther away they are from each other, the more distinct they are and this translates into their better recall (all other things being equal). This part of the SIMPLE model mathematically resembles relations among stimuli along one axis in the CDIE model. As the authors note, their model is similar in using the distinctiveness to a range of other models used not only in serial recall research (e.g., Murdock, 1960; Baddeley, 1976; Bjork & Whitten, 1974; Crowder, 1976; Glenberg & Swanson, 1986; Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999; Neath, 1993a, 1993b); Gallistel, 1990; Gallistel & Gibbon, 2002).

In the SIMPLE model but not in the CDIE model the time dimension is logarithmically transformed. Theoretically, however, the logarithmically transformed time scale represents a built in exponential time based decay of distinctiveness of items. The justification for this transformation is not necessarily the time based decay but as I argue latter, these authors do not offer convincing arguments for any other mechanism; on the contrary, they mention that this is a valid interpretation of the reduced distinctiveness. It further follows that the other models that incorporate the distinctiveness logic may actually also have time based decay built in into some initial assumptions.

In the SIMPLE model, the end of time scale is “right justified” to the present time and so it is not flexible when moving towards the beginning of the scale/time. The authors explain their use of this scale by referring to psychophysical laws but without much justification for that decision. One consequence of this is that it is hard to understand what guides the transformation in different cases of episodic memory. (I discuss this scale in more detail in the text later) Finally, in this model as well, and possibly in select other models, the mechanism that can keep track of the order position is separate from the dimension along which, say semantically different, the words are arranged. The interactions between dimensions are not clearly specified as they should be, I argue.

TODAM model

TODAM (Theory of Distributed Associative Memory; Murdock 1982, 1983, 1995; Hockley & Murdock, 1987; Lewandowsky & Murdock, 1989) is a neural network model that contains the “forgetting parameter” alpha that seems to be the time based decay of existing memory. Moreover, this parameter is a function that, as explained by the authors, enables TODAM to model the forgetting curves. TODAM2 is an improved version of TODAM and *the constant* Alpha remains.

$$M_{new} \,=\, \alpha \, M_{old} + [...] \tag{18}$$

This is analogous to the Eq.9 of the CDIE model: it is a spontaneous, time based decay, which is constant, unlike in the CDIE model. The order information is obtained by reinstating the context of an item and this makes the model similar to the TILT model. The model uses the time based decay with the coefficient alpha and needs a separate dimension/mechanism to infer the order information. Nonetheless, the shown part of the Eq.18, is not acknowledged as the time based decay.

SOB model

In this neural network implemented model (Serial-Order in a Box, Farrell & Lewandowsky, 2002) each new item encountered in an ordered list of stimuli is encoded gradually with less and less strength, provided that it has at least some similarity to already encoded items. This produces the "primacy gradient" that models serial position effects observed in experiments. The main quantity describing this is postulated "energy" of the system that changes over time course of the task. This change is, importantly, set by the authors to fit the data. This opens the question whether the authors can claim that this change may not include any time based decay interpretation. It does not seem that this matter was of any importance in the initial purposes of the model but it becomes important in the context of the debate about decay. In regards to the order information, it also seems that with this model, an additional mechanism is needed to extract the order information from the states, such as the strength, of items after learning.

In addition, some very recent papers by these authors include decay but do not clearly say it in titles or abstracts. For example, Oberauer, Lewandowsky, Farrell, Jarrold & Greaves, 2012 have a paper titled "Modeling working memory: An interference model of complex span" and the following is their abstract:

> This article introduces a new computational model for the complex-span task, the most popular task for studying working memory. SOB-CS is a two-layer neural network that associates distributed item representations with distributed, overlapping position markers. Memory capacity limits are explained by interference from a superposition of associations. Concurrent processing interferes with memory through involuntary encoding of distractors. Free time in-between distractors is used to remove irrelevant representations, thereby reducing interference. The model accounts for benchmark findings in four areas: (1) effects of processing pace, processing difficulty, and number of processing steps; (2) effects of serial position and error patterns; (3) effects of different kinds of itemdistractor similarity; and (4) correlations between span tasks. The model makes several new predictions in these areas, which were confirmed experimentally.

In the text the authors explain their model SOB-CS in which the time-dependent Hebbian Antilearning (Anderson, 1991) operates "to remove irrelevant representations" (equations 10-12, page 787.) This is exactly the same diffusion process, the spontaneous decay, as described in other areas of sciences, in the CDIE equation and many other connectionist models (e.g., Grossberg, 1978). To be fair, the authors have a paper “Modeling working memory: a computational implementation of the Time-Based Resource-Sharing theory”, where Oberauer and Lewandowsky (2011) introduce time based decay when making a model of complex span task, the Time-Based-Resource-Sharing (TBRS) model of Barrouillet, Bernardin, and Camos, 2004.

OSCAR model

The Oscillator-based memory for serial order by Brown, Preece, and Hulme, (2000) is a sophisticated model of serial order that accounts for a nice range of experimental results. Similarly to some other models, it takes into account the context of items and at recall the OSCAR model that the context is reinstated. Importantly for this dissertation, the model contains multiple time scales. This is also the assumption for the model presented in this dissertation. Further, the accuracy in recall is directly related to distinctiveness of items, which will also be shown later for the CDIE simulations. Complexity of sources of forgetting is readily acknowledged:

Adaptive Resonance Theory (ART)

This theory developed by Stephen Grossberg and Gail Carpenter is mentioned here, briefly but with a very good reason. First, it is based on early works of Stephen Grossberg on list remembering (Grossberg, 1969) which included differential equation describing a single memory almost identical to CDIE model's, which is not surprising if one treats the activation or excitation of a trace as physics would- using a widespread diffusion of energy model. This was not often the case in psychology and cognitive science, which is why this theory is mentioned here. Further, the theory now contains three memory “stores”, with slightly different dynamics. All are modeled by differential equations. The first difference is that in this work, the attenuation parameter, Alpha, is a constant in Grossberg's and his colleagues' works while in CDIE it it is not. This opens up possibilities for the CDIE to model some attentional phenomena which have already been addressed in the ART. Second, the ART is implemented in connectionist networks, while the work here relies on analytical and numerical methods for simulations. Naturally, each approach informs the other this way. Finally, the CS dimension in CDIE is a continuous one, while in list modeling Grossberg uses discrete items. The main point important for now, however, is that the decay is used, and in many ways, to examine learning of episodes.

These are not, of course, the only articles and models in this area of research or the most representative ones in many ways but they well illustrate the following two points. The first is that the time based decay is very possibly implicitly or explicitly present (but not in main messages of papers) even in theories that (seemingly) argue against the need of its further examination in some memory research. Sometimes the decay seems to be even essential for these theories but is named “exponential decrease,” “gradient”, etc.

The other point is that order information and its dynamics need to be explained in more detail in episodic memory models, such as serial recall models; all of the accounts seem to postulate a separate mechanism that handles this information. In an episodic memory generally and in serial recall specifically, memory trace needs to carry information about the nature of the item itself but also the items position in temporal context of other items. Therefore, these are two qualities which may potentially be quantified and modeled. Moreover, they both may have similar nature and non-trivial evolutions and interactions in time. Therefore, if it is reasonable to argue that a memory trace decays in time, then it is also reasonable to look for more specific information about decay of order information. This is another reason why the time-based decay research needs to, and does continue.

None of the existing models is capable of (or necessarily intended for) taking into account together all of the factors and interactions mentioned in the above survey of experiments and models to look into trace evolution except for the CDIE model. In a sense, the CDIE model is a generalization of existing models of episodic memory. To illustrate my point I present the following analogy: each of these models is looking at the two-dimensional map of a surface of a sphere and it describes some parts of that surface well but eventually it runs into problems because it does not take into account the fundamentally new property from their perspective - sphericality. What is needed in cognitive science is not the decision which model explains the most data but a way to combine these models in novel ways to learn more about episode processing. I suggest that that the CDIE model offers one way to approach this phenomenon. It is hard to think that any researcher would deny that memory is a dynamic system. Neural networks are dynamic systems and between models using the CN and the CDIE the only difference is a perspective for describing memory, nothing more radical.

Taking into account multiple interactions simultaneously at a more behavioral level and simulating the whole process of episode processing offers new insights about its mechanisms, especially into the step by step evolution (as opposed to only encoding or only retrieval phases). A dynamical systems approach is not just describing the results of the time passage for processes but it offers more information about how these results may arise from the processes. This adds a lot more detail into hypotheses about the processes. Further, dynamical simulations, such as those with the CDIE or CN models, may show some counter-intuitive behaviors and explanations because they typically take into account more interactions than a researcher can follow only by theorizing about a phenomenon. The latter is typically the basis for the model assumptions. Specifically, it seems that computer simulations of models are a better way to understand influences of time based decay and other factors, because of the problems related to separating the the spontaneous decay from other mechanisms. This might open possibilities for devising new experimental paradigms for this research.

3.2.4 More reasons to consider decay

The above mentioned paradigms seem to be confused about what is really being measured. They appear to be finding some time-based decay. Are there other arguments for thinking about the concept of time based decay of memory or for abandoning it completely, the arguments not often used in cognitive science? In what follows, some such arguments for decay are addressed.

Even before the cognitive term "memory" was generally allowed into scientific psychology, physiologists and behaviorists studied stimulus-response associations (Thorndike, 1898; Pavlov, 1927; Watson, 1919). The strength of the association was changed as the consequence of learning, or forgetting of what was learned if the learned association was not used any more. The strength of association in this research relatively closely resembles the cognitive concept of a memory strength. Something has been strengthened beyond some basic level by some effort. If left alone after that, it spontaneously returns to the base level.

Hebb (1949) thought of an idea being in short-term memory when a specific pattern of neurons was active above some resting level. In the absence of some additional activation, this pattern decays to a resting state. Closely related ideas exist in neural networks research. Long term memory is related to the strength of connections between units while the short term memory (or working memory) is represented as a current activation of a subset of units. The connections that are not activated in some way are not used in tasks. Those connections used at one point in time, but not in another, decay.

Examples like these only show the usefulness of the concept of time-based, spontaneous, decay. It is economical to temporarily add activation to a part of a system and let the system return to the base level on its own as opposed to having the same system in much higher activation state and temporarily put in an effort to lower the activation of its parts. This principle seems to be often found in nature. Instead of examining further why this might be the case, it seems reasonable to say that there has to be a very specific reason why to assume that this principle would not be appropriate for perception, attention, and memory research. The research mentioned above does not seem to provide this reason. Moreover, several theories of memory that do assume two or more independent or partially related "episodic processors" or memory stores, each of which may have different dynamical properties, notably at least the long term and short term memory stores (e.g., Waugh & Norman 1965; Reyna & Brainerd, 1995), are capable of accounting for a massive range of data.

As Brown, Neath, and Chatter (2007) and Brown and Lewandowsky (2010) argue, forgetting curves along many time scales seem to be very similar to each other which may implicate similar mechanisms, albeit on different time scales, operating in traditionally separately investigated phenomena such as long term memory and short term memory recall or recognition. The CDIE model can easily accommodate these ideas and it is currently being used to provide a precise theoretical definition of the term spontaneous decay, and to examine the various consequences of decay for both TAS and CS dimensions. Decay here has a complex and non-intuitive influences on total dynamics, and simulations show that without decay, or with too much decay, the model is not able to capture forgetting trends from psychological experiments. The above discussion may be convincing for some researchers to continue exploring the role of time-based decay in cognition.

3.2.5 Conclusions about time based decay in research

The concept of time based decay does not seem to be wrong or useless in memory research. Many natural phenomena seem to include it and there is no reason to assume otherwise in cognitive science. On the contrary, it may serve as a concept that brings the diverse research areas of cognitive science closer together and moves cognitive science closer to more traditional sciences. Further, the existing paradigms are not sufficiently sophisticated to conclusively argue against the time based decay and actually seem to find some evidence for it. It appears that decay may be operating in addition to any other suggested mechanism that may have similar effect on memory. There is no argument or reason why this could not be the case. The fact that many authors report decay in the lower amount than expected may serve as a guide for further research. Typically, the research mentioned above that reports little evidence of decay is examining forgetting over the course of several seconds up to well shorter interval than two minutes or beyond. Longer time intervals may produce higher values for decay. As mentioned, forgetting curves along many time scales that seem to be very similar to each other may implicate similar mechanisms, albeit on different time scales, operating in traditionally separately investigated phenomena such as long term memory and short term memory recall or recognition.

The previous part of this chapter argued for the following:

1. The spontaneous decay should be precisely defined. 2. It should be discussed why verbal memory would have completely different dynamics than non verbal, as suggested by some authors. 3. Current experiments in episodic memory are not specific enough about how the decay contributes to the time evolution of memories. This leads to vague predictions about the results of the operation of decay which in turn allows only for very weak arguments against it. 4. Current models of episode processing include terms analogous to time-based decay of information in a memory trace even when they argue against it. 5. There exist other reasons to consider decay, typically not considered in cognitive science and related areas.

and partially for the following point mentioned in the beginning of the chapter:

6. Using the time based spontaneous decay issue, I argue that the dynamical systems approach using differential equations is not only useful for simulating episode processing but, even more importantly, it is a new perspective in theorizing that introduces fundamentally new conceptualizations of episode processing mechanisms to be further investigated. Illustrations include demonstrations that the decay might be important for distinctiveness of stimuli (which plays a crucial role in the Word Length Effect, Serial Position Curves, Chunking in memory, photographic memory, etc.) as well as for attention, false memory, and forgetting curves research.

In order to continue on this last point, I now turn to several CDIE applications that all include time-based decay, to further illustrate the CDIE itself and its role in examining the time-based decay in episode processing.

4 CDIE application on spontaneous decay issues: Decay is important for Distinctiveness of stimuli

4.1 The Word Length Effect and Distinctiveness analysis.

The classical Word Length Effect (WLE) is the finding that words that take shorter time to articulate are remembered better than the words that take longer time (Baddeley, Thomson, and Buchanan, 1975). This finding has been taken to represent a strong evidence for time-based decay (e.g., Cowan, 1995): Longer words have more time to decay until recalled and fewer long words may be rehearsed in the phonological loop (Baddeley, 1975)-hence their recall is worse. Note that this is the case when the stimulus sets contain either all short words or all long words.

Neath, Bireta, and Surprenant (2003), performed a set of four experiments to examine the WLE. The procedure they used is presented in more detail here to illustrate general ideas of the paradigm. The authors used four different sets of stimuli that varied only in time for articulation and measured the word length effect. In all four experiments, the following design was used. Subjects were asked to silently read each presented word. They were presented with 20 lists each, 10 of which were short-item lists and 10 were long-item lists. Each list contained 8 items. Each successive word was presented on a screen immediately after the previous one was removed. At the test all the words from the list were shown on the screen and the subjects were asked to click on them with a mouse in the order of presentation, at the self-paced speed. The authors compared their word sets on the following measures and concluded that the long and short lists they used differed only on the articulation time: mean normal pronunciation time in milliseconds, concreteness, printed familiarity, imageability, meaningfulness, number of phonemes, number of syllables, Paivio frequency, British National Corpus frequency, standard frequency index.

In experiment 1, the words were the following. Short: BISHOP, DECOR, EMBER, PECTIN, PEWTER, TIPPLE, WICKET , WIGGLE, and long: COERCE, FRIDAY, HARPOON, HUMANE, MORPHINE, TYCOON, VOODOO, and ZYGOTE. The long words were recalled more poorly than the short ones which is the classical word length effect.

In the experiment 2, the following words were used. Short: BULLET, CABIN, CARROT, DEVIL, LADDER, PICNIC, TICKET, ZIPPER. The long words: BABY, BALLOON , CRAYON, ORANGE, SIRLOIN, SPIDER, TOWER, and VACUUM. This experiment showed the reverse word length effect.

In the experiment 3, the words were as follows. The short words: BUTTON, CANDLE , PENCIL, POCKET , SHOVEL, SPIDER, TRACTOR, and WHISTLE . The long words: BRANCHES, CANOES, CURTAINS, NECKLACE, NEEDLE, PEBBLES, ROBOT , and STATION. In this experiment, both sets of words were recalled equally well.

In the fourth experiment, the words were as follows. Short: ACROBAT, ANIMAL, DAFFODIL, GENTLEMAN, MEDALLION, PHYSICIAN, UMBRELLA, VEGETABLE. The long words: AUTOMOBILE, INFIRMARY, MACARONI, NEWSPAPER, PERFORMER, PROSECUTOR, VOLCANO, and WHOLESALER. Here too, the authors did not get any difference between the two item lengths.

I presented all the stimuli here to illustrate that it is not at all clear that these words are sufficiently different along the important dimensions; namely, the effect size reported in the paper seems small at the first glance, even though it may be statistically significant. Indeed, Jalbert, Neath, Surprenant (2011) call this phenomenon the time-based word length effect and it seems that this effect is limited to a specific set of stimuli (the original set by Baddeley et al., 1975, which is almost always used to replicate the effect).

There is another kind of the WLE that occurs when the words are NOT matched on the number of syllables or the articulation rate. It is named the syllable-based word length effect. Consequently, this introduces a larger difference in the duration of words. This effect seems to be much less stimulus set specific, as expected.

In the simulations presented in this dissertation the syllable-based word length effect is examined (time units represent a syllable). These simulations therefore shed light on many experiments using words of different number of syllables: Baddeley et al. (1975), Bireta, Neath, and Surprenant (2006), Baddeley, Chincotta, Stafford, and Turk (2002), Watkins (1972), Avons, Wright, and Pammer (1994), LaPointe and Engle (1990), Tehan, Hendry, and Kocinski (2001) (for more information on specific tasks used see Jalbert et al., 2011).

These analyses lead directly into the discussion of trace distinctiveness and its role in remembering ordered information. I conceptualize the distinctiveness as a complex measure and recall as being directly related to distinctiveness.

The following three sets of figures are the results of the CDIE simulation of memory traces of words of three different lengths: 1 unit of time (labeled 1;0;0), 2 units of time (labeled 2;0;0), and three units of time (labeled 3:0;0). In all three of them the break between stimuli is 0 time units. At the beginning, each lists contains the same kind of stimuli- either Long, Medium, or Short. Nine stimuli per list are simulated. Note that the stimulus 4 (Cyan) is in this simulation unusually close to stimulus 3 (Blue) along the conceptual dimension. Other stimuli are relatively far away from each other so their interactions are similar to each other. Because of its unusual interaction with the third one, the fourth stimulus does not perfectly follow general patterns visible in these figures. In all figures, several lines and peaks of the same color- belonging to the same stimulus - are visible. This is because the stimulus intensity was plotted at each time step for the entire evolution time (along the x-axis). This is convenient here because this way it is clearly visible how long the stimulus lasts at which strength of intensity. Multiple peaks of different colors, that are visible at the same time location indicate that these other-than-the-highest peaks exist in the vicinity of the highest one, along the CS dimension.

For the first figure in every pair, the following is the case: On the vertical axis (height) is the simple intensity of a trace. On the horizontal axis is the position in time (farthest left position being the earliest time). Stimuli were presented in the following order: Red, Green, Blue, Cyan, Magenta, Black, Gray, Orange, Purple. This is obvious from the figures.

For the second figure in every pair, the following is the case: The vertical axis represents the simple intensity of activation, just like before. The horizontal axis represents the position in a list. For example, the position number one in the and the red color refer to the same stimulus. The peak intensities for every position are the average of the three highest points for that stimulus from the first pictures in each pair. This choice, to use the three highest points is arbitrary but since the threshold above which the intensity has to be to be registered in another field is of the same nature, the exact determination of these details needs to be done in a context of a specific task.

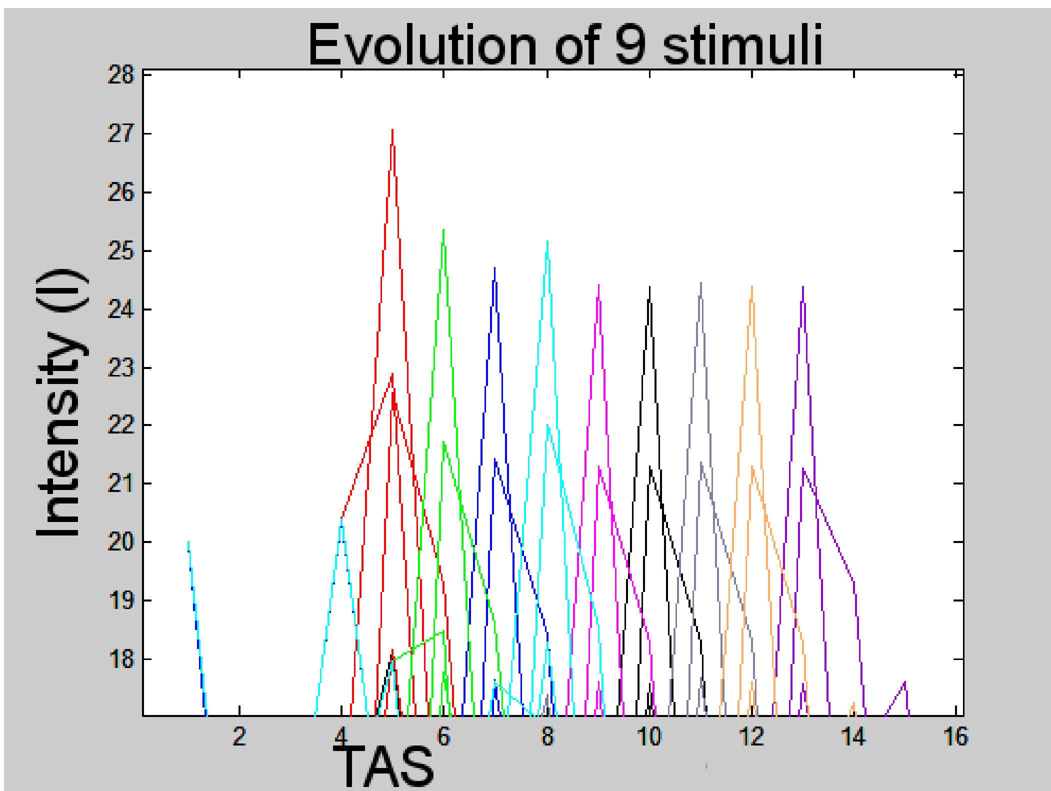

Figure 8: Nine stimuli of duration one are presented in time (Red, Green, Blue, Cyan, Magenta, Black, Gray, Orange, Purple) and their activations (intensities) at each time step. The primacy effect is pronounced. Several lines and peaks of the same color- belonging to the same stimulus- are visible because the stimulus intensity was plotted at each time step for the entire evolution time. This way it is clearly visible how long the stimulus lasts at which strength of intensity AND multiple peaks (of various colors) that are visible at the same time location, indicate that these other-than-the-highest peaks exist in the vicinity of the highest one along the CS dimension.

This and similar figures show a profile of the "2D" field. It can be thought of as a side view of Figure 2.but it is not exactly that - this is showing the entire evolution of each stimulus along the CS dimension. All time-series for each stimulus (many lines of the same color) from a graph of the form of Figure 2. are shown from the perspective of the TAS dimension.Notice that the inputs, and the resulting intensities of all activations are Gaussian both in the TAS and the CS dimensions.Along the TAS dimension, there is one Gaussian input at each order position (1, 2, 3, etc).Also note that in these figures intensities above some prescribed threshold are shown, so we are only seeing the strongest activations. Finally, recall that intensity is taken to be directly related to probability of recall in these models. Since all the stimuli in this case differ only in their order, and the fourth one is the only one that is significantly close to the third one along the CS dimension, the primacy effect is obvious, because overall the first input has the highest peak, and is thus most recallable. I argue later that it is not only this intensity that needs to be taken into account when making CDIE forgetting plots.

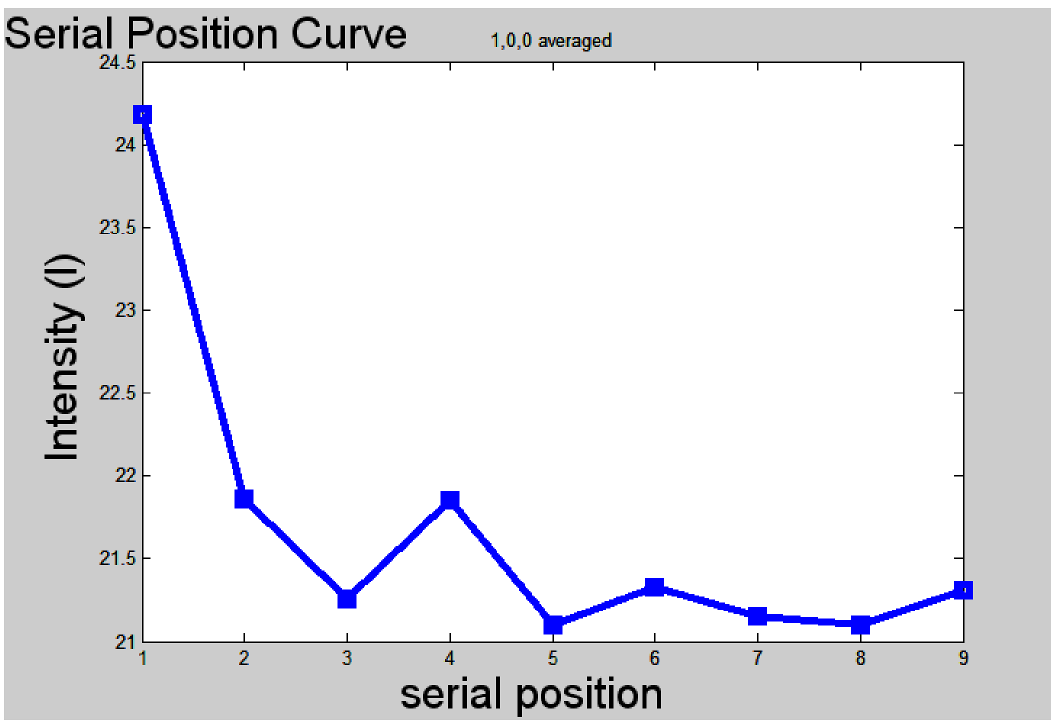

Figure 9: For the same stimuli, the serial position is on the x-axis, and the activation of each stimulus on y-axis. For this figure, the three highest peaks from the Fig. 8 for each stimulus are averaged. Note that since the intensity units and thresholds are arbitrary, this is not problematic. Very similar general shape of the curve is obvious from Fig. 8 if just the highest peaks are connected.

The following set of two figures represents stimuli that last 2 units in exactly the same form as Fig. 8 and 9.

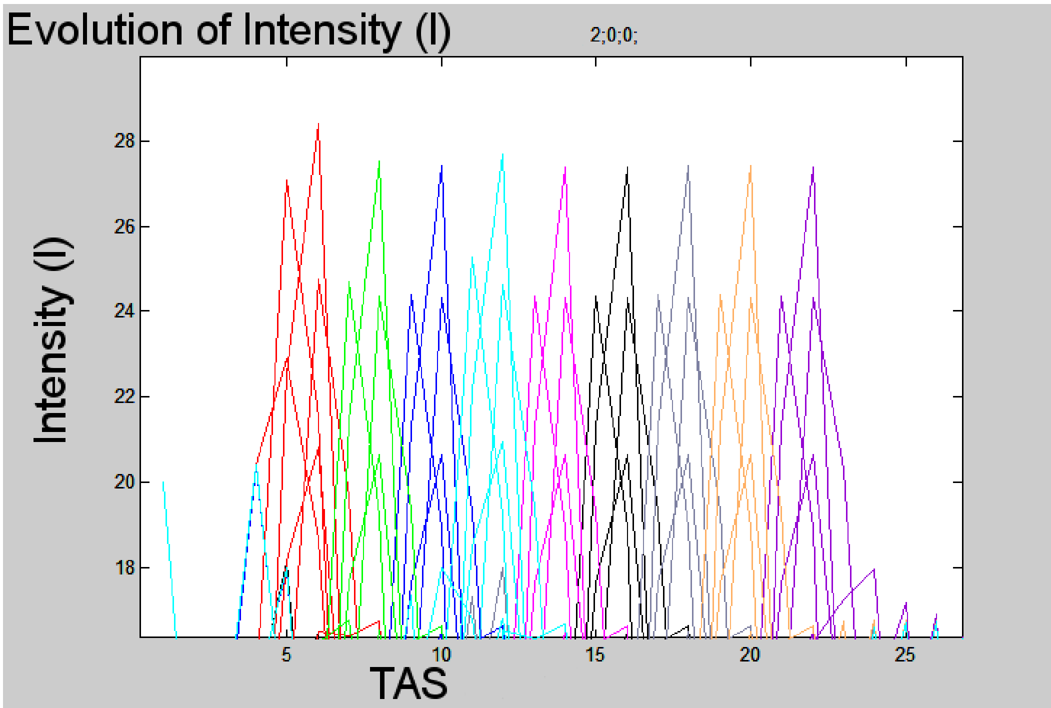

Figure 10: Duration of stimuli is 2 time units, the break between them is always 0.

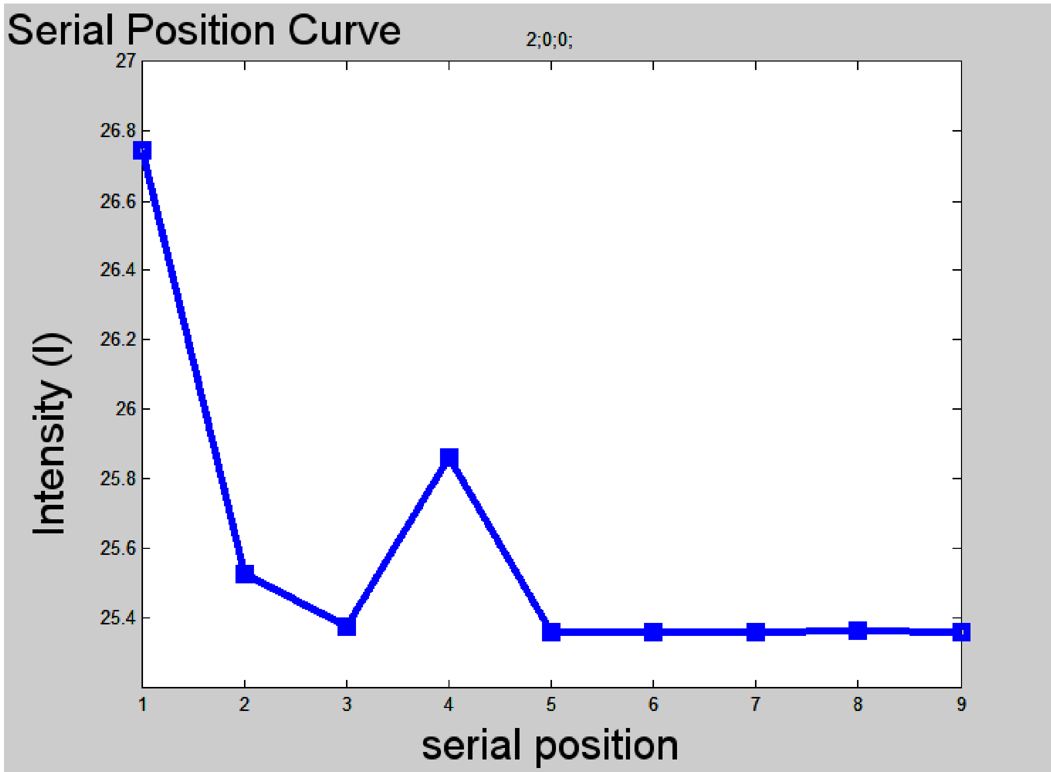

Figure 11: Duration of stimuli is 2 time units with zero duration break.

Finally, the following set of two figures represents stimuli lasting 3 units, again, in exactly the same form as the previous four figures.

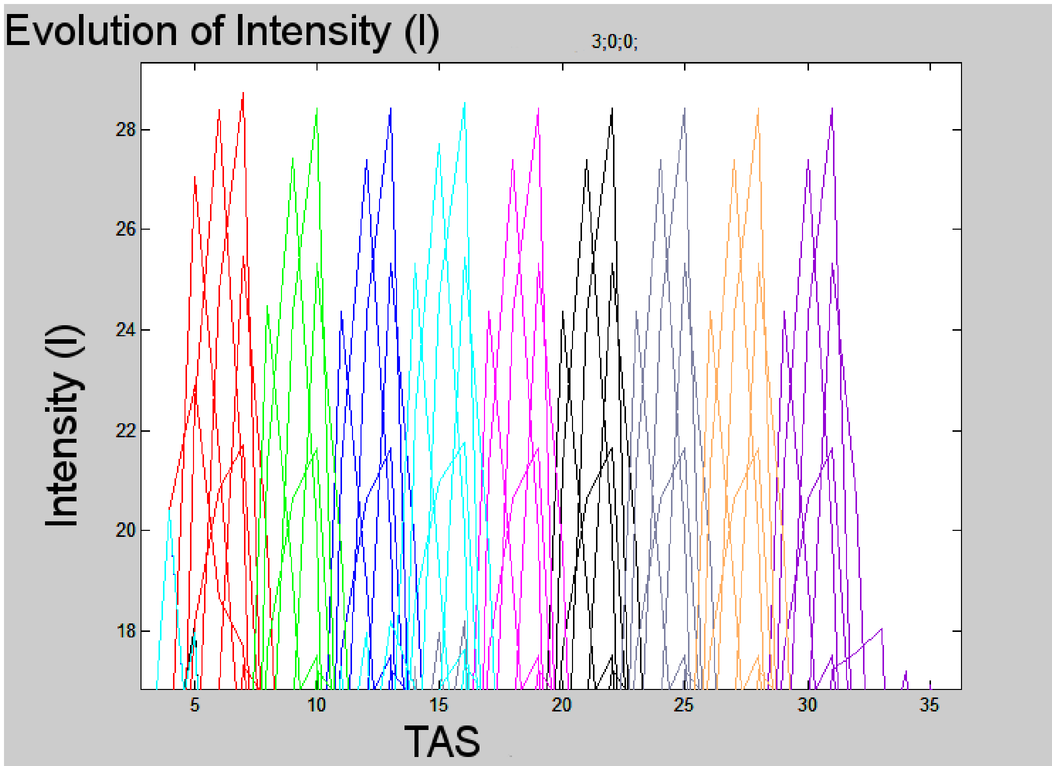

Figure 12: Duration of stimuli is 3 time units and all breaks last 0 units.

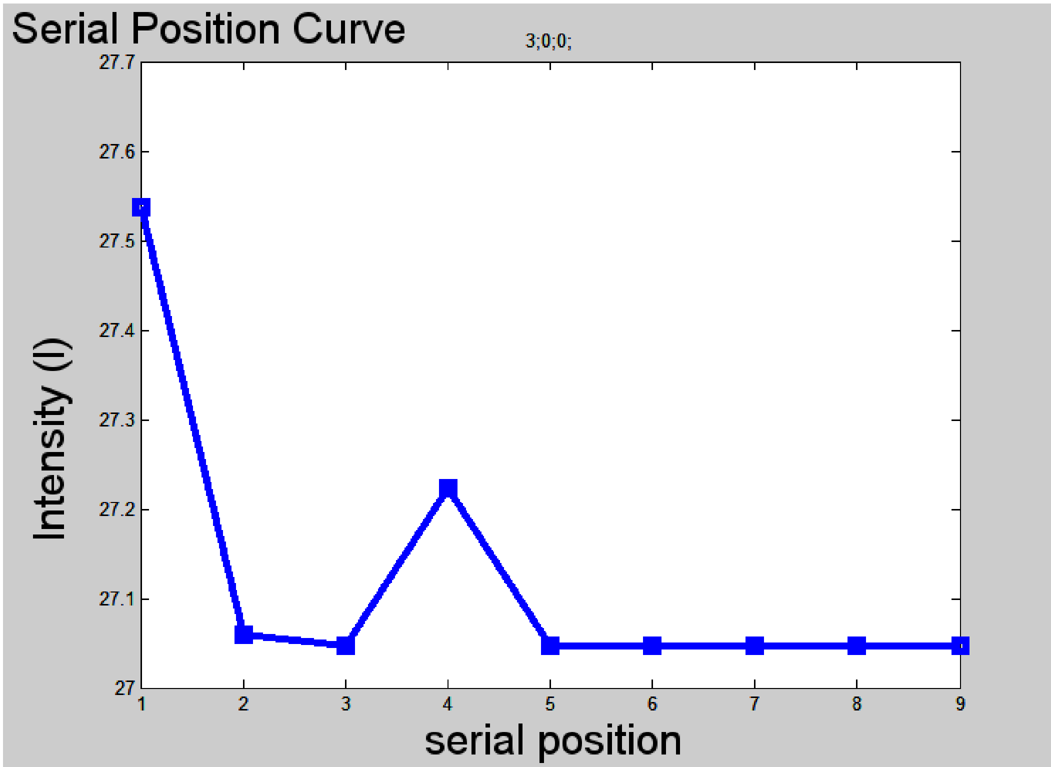

Figure 13: Duration of stimuli is 3 time units and zero duration of breaks.

When the graphs from Fig. 9, Fig. 11, and Fig. 13 are put together in one figure:

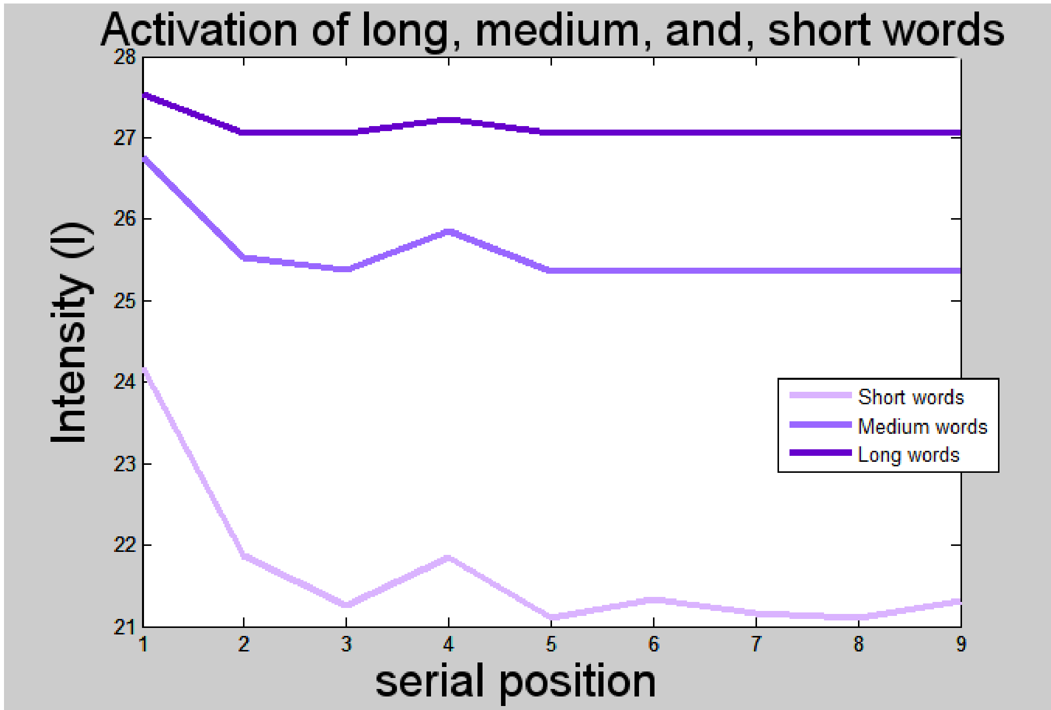

Figure 14: Long words add more total activation to the field but at the same time the activation of all of them seems to be very similar compared to short words. Everything else being equal, this adds distinctiveness to words in the short-words lists. Note, however, in experiments rarely everything else is equal.

Next, it is useful to examine what happens with the depth of the gap this is important for the distinctiveness of traces - if the break between the stimuli is not 0. This is examined for the stimulus duration 2 and 3, while the lists still contain only stimuli of the same length. The following four figures show these situations. These are only some of possible combinations of stimulus and break durations but they seem to be representative of a wide range of possibilities.

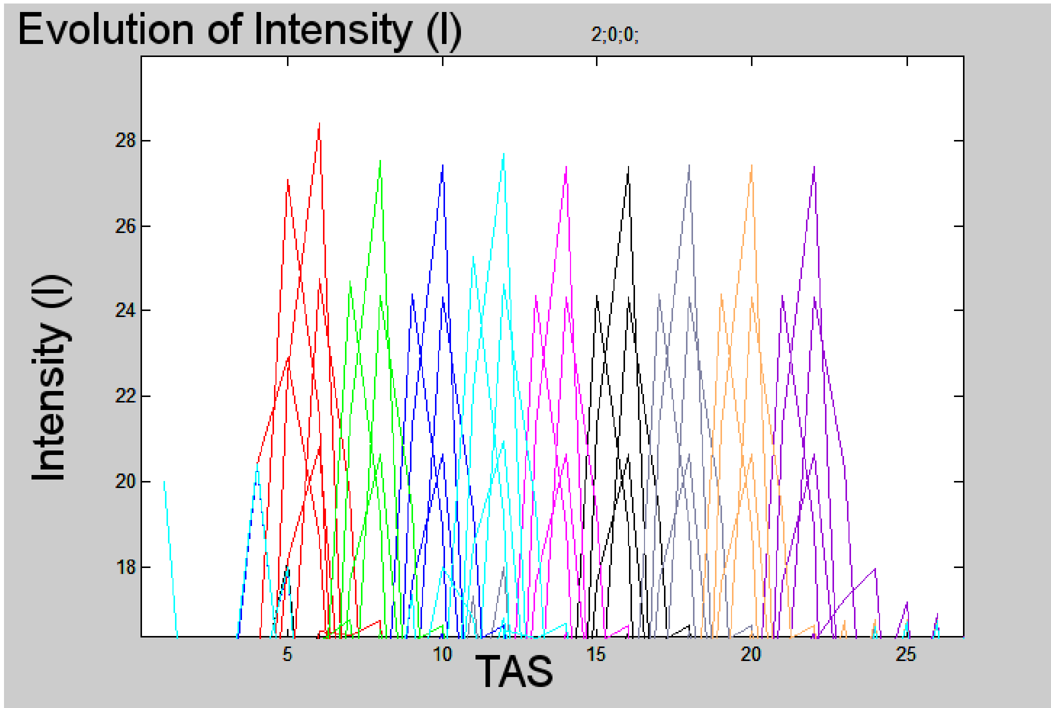

Figure 15: Order and duration of two-unit of stimuli do not change from Fig. 15 to Fig. 18. All the pauses between stimuli are the same within each of the four subplots. The pauses here are zero units. The depth of gaps between stimuli increases from Fig. 15 to Fig. 18 as a consequence of the slower rate of presentation of stimuli.

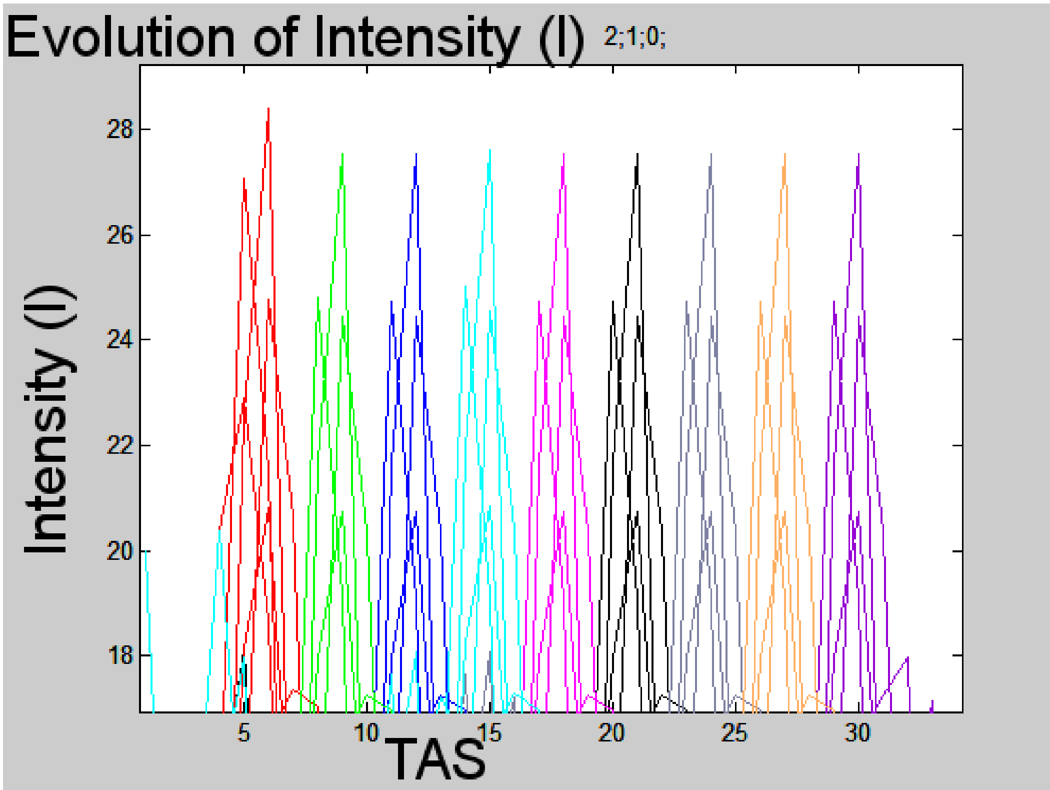

Figure 16: Order and duration of two-unit of stimuli do not change from Fig. 15 to Fig. 18. All the pauses between stimuli are the same within each of the four subplots. The pauses here are one unit. The depth of gaps between stimuli increases from Fig. 15 to Fig. 18 as a consequence of the slower rate of presentation of stimuli.

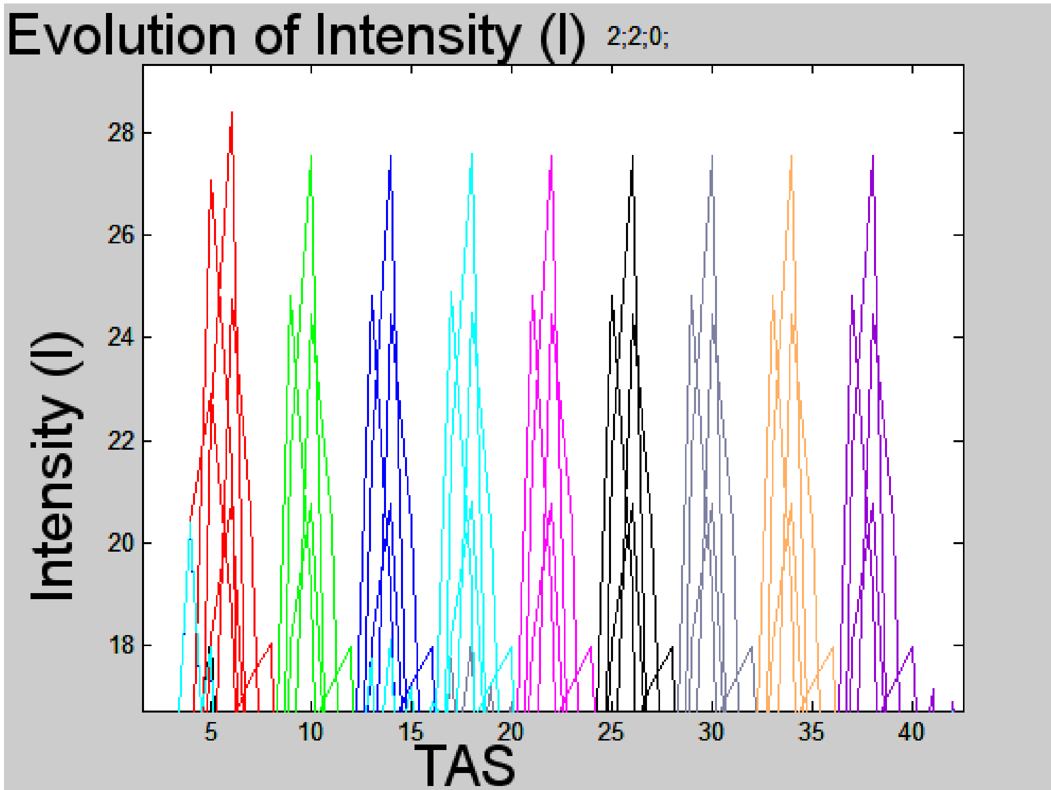

Figure 17: Order and duration of two-unit of stimuli do not change from Fig. 15 to Fig. 18. All the pauses between stimuli are the same within each of the four subplots. The pauses here are two units. The depth of gaps between stimuli increases from Fig. 15 to Fig. 18 as a consequence of the slower rate of presentation of stimuli.

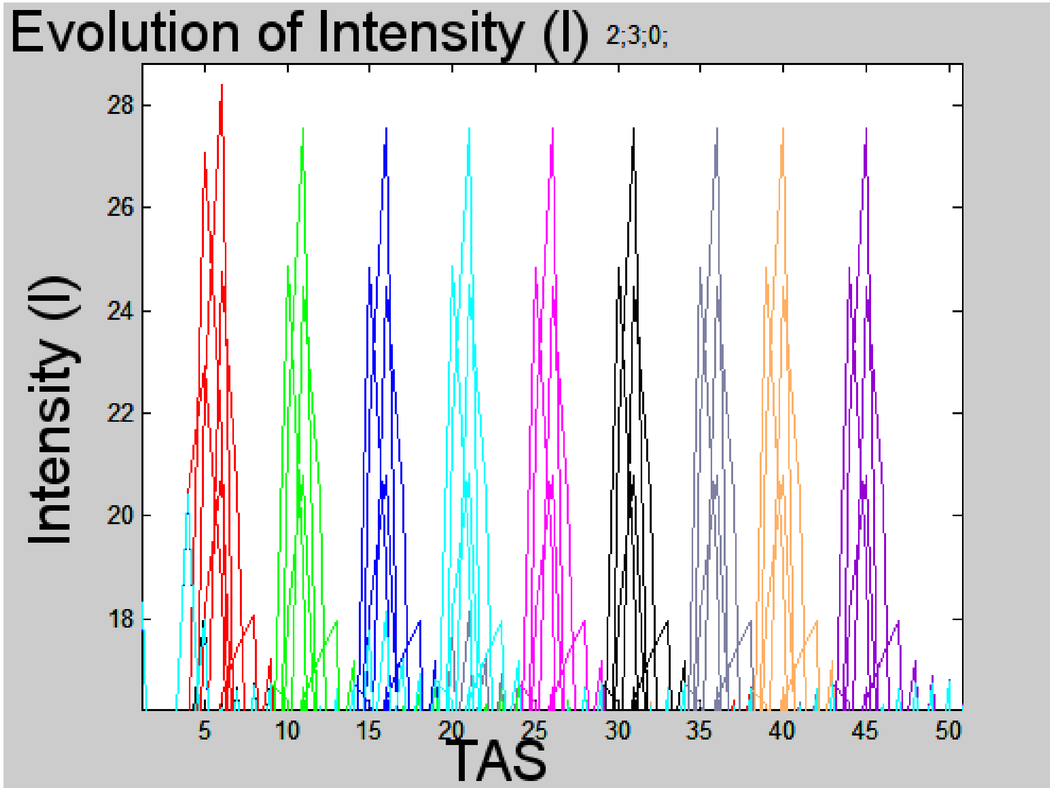

Figure 18: Order and duration of two-unit of stimuli do not change from Fig. 15 to Fig. 18. All the pauses between stimuli are the same within each of the four subplots. The pauses here are three units. The depth of gaps between stimuli increases from Fig. 15 to Fig. 18 as a consequence of the slower rate of presentation of stimuli.

The following is a set where stimulus duration is 3 units and the break goes from 0-2 units. The trend in gap depth is the same as in previous simulations.

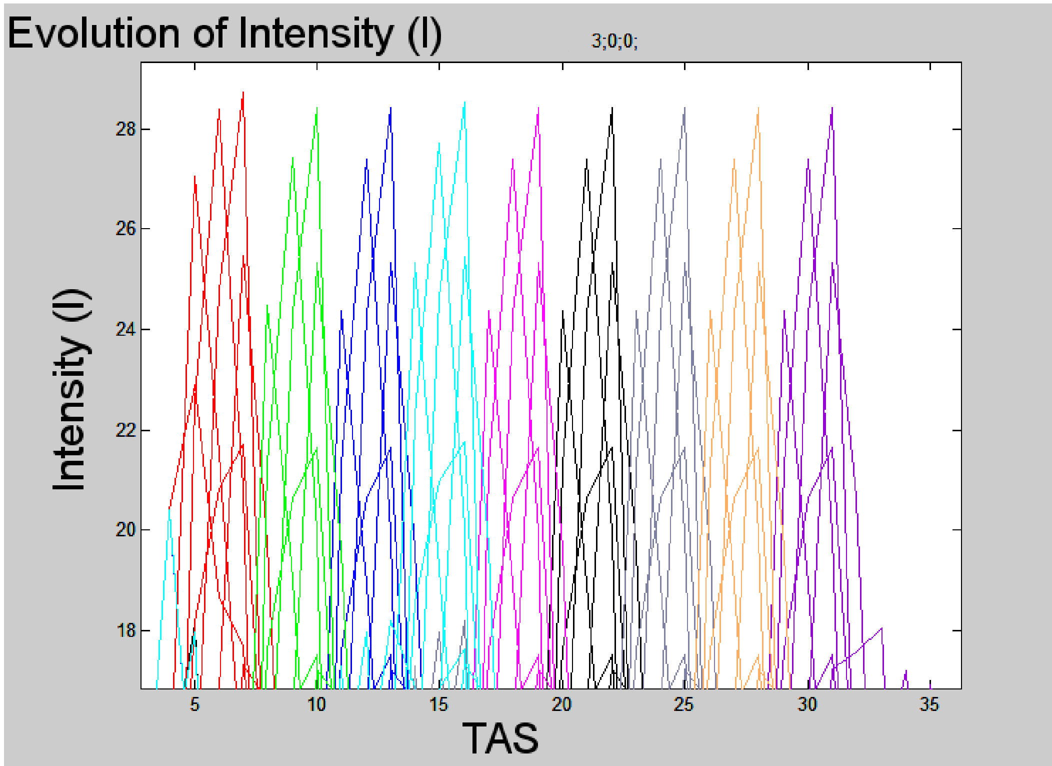

Figure 19: Order and duration (3 units) of stimuli does not change from Fig. 19 to Fig. 21. All the pauses between stimuli are the same within each of the figures. The pauses here are of duration zero units. The depth of gaps between stimuli increases from Fig. 19 to Fig. 21 as a consequence of the slower rate of presentation of stimuli.

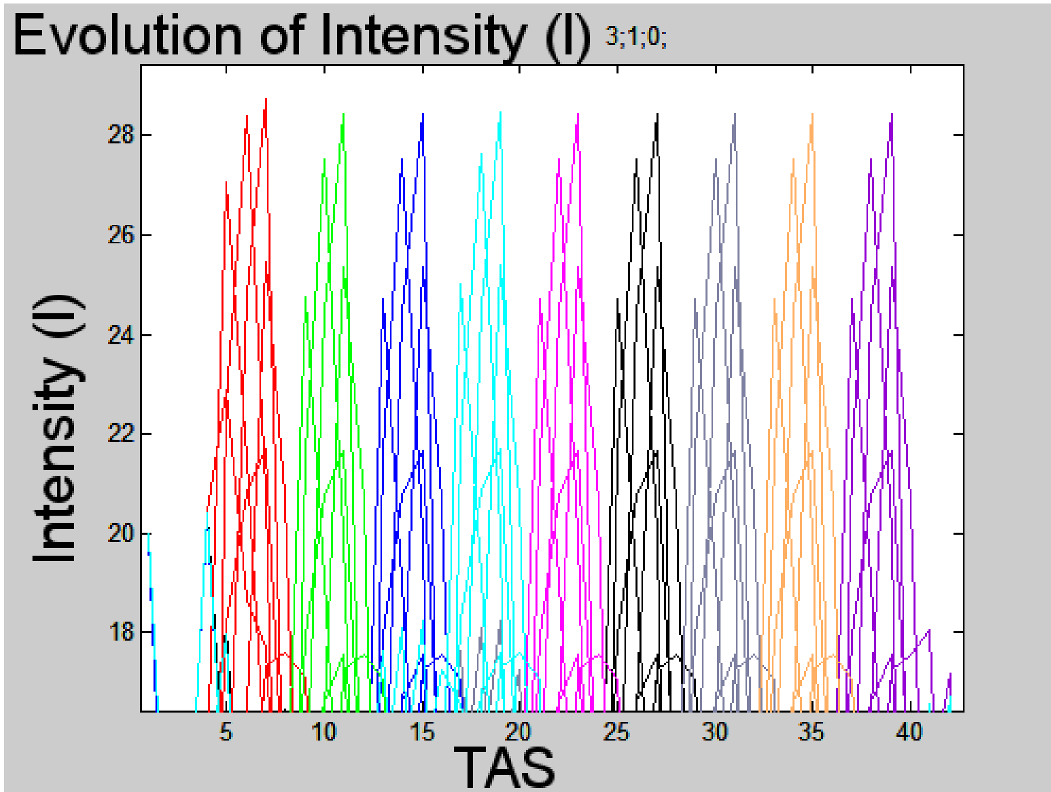

Figure 20: Order and duration (3 units) of stimuli does not change from Fig. 19 to Fig. 21. All the pauses between stimuli are the same within each of the figures. The pauses here are of duration one unit. The depth of gaps between stimuli increases from Fig. 19 to Fig. 21 as a consequence of the slower rate of presentation of stimuli.

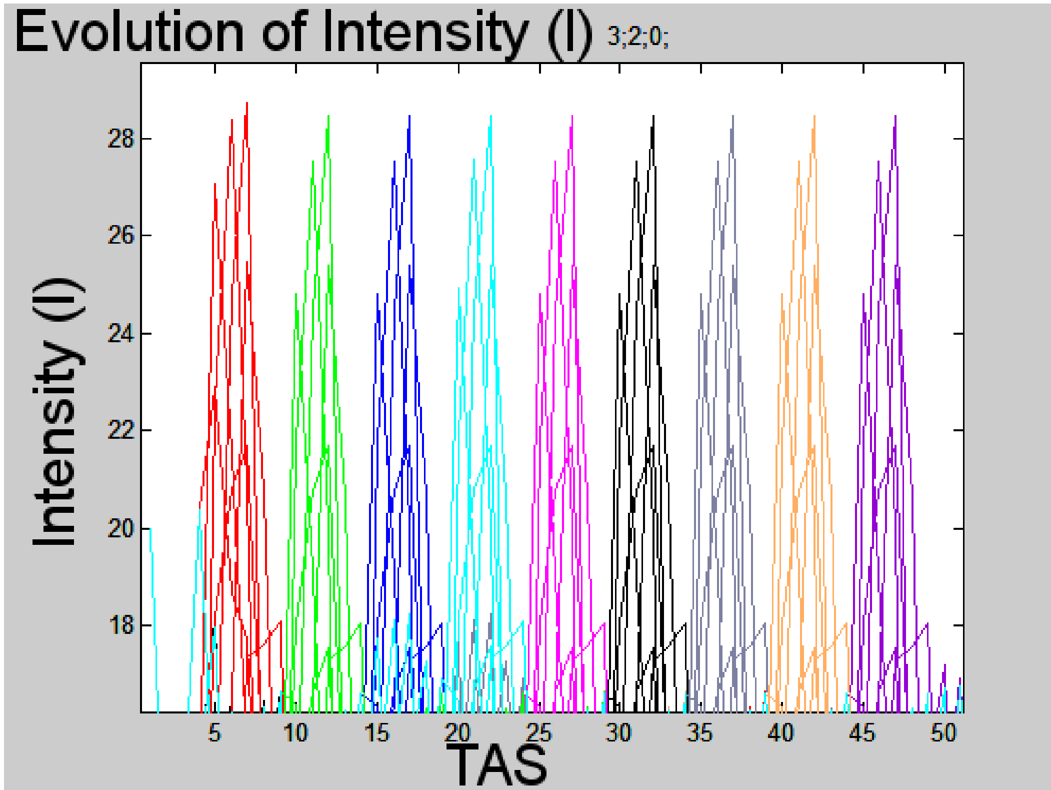

Figure 21: Order and duration (3 units) of stimuli does not change from Fig. 19 to Fig. 21. All the pauses between stimuli are the same within each of the figures. The pauses here are of duration two units. The depth of gaps between stimuli increases from Fig. 19 to Fig. 21 as a consequence of the slower rate of presentation of stimuli.

As visible from the previous two sets of simulations (Fig. 15-18 and Fig. 19-21), the depth between stimuli is getting larger (lower and lower activation) as the break between the stimuli gets larger, in each of these cases. This is in line with thinking about the gap as being responsible for the psychological construct of distinctiveness. It is important to notice that these figures only illustrate the distinctiveness along the time-as-space dimension but the other dimension is mathematically unchanged. Therefore, there are (at least) two separate kinds of distinctiveness that need to be taken into account when comparing stimuli. In the literature this is reflected in a debate about whether the word's length (its width along the time-as-space dimension) or its orthographic, phonological, semantic, etc, neighborhood sizes (the width along the conceptual dimension here) is responsible for various finding in the WLE research.

The following figure (Fig.22) shows a mix of short and long words, beginning with a short one, with a 0 break between stimuli. The advantage in distinctiveness and height for the first stimulus is clear when compared to other stimuli of the same duration, but the duration of long words adds the activation to the traces of long words, reducing the depth of gaps leading to the next short word. Gaps between stimuli are variable, overall, and will play a large role in distinctiveness.

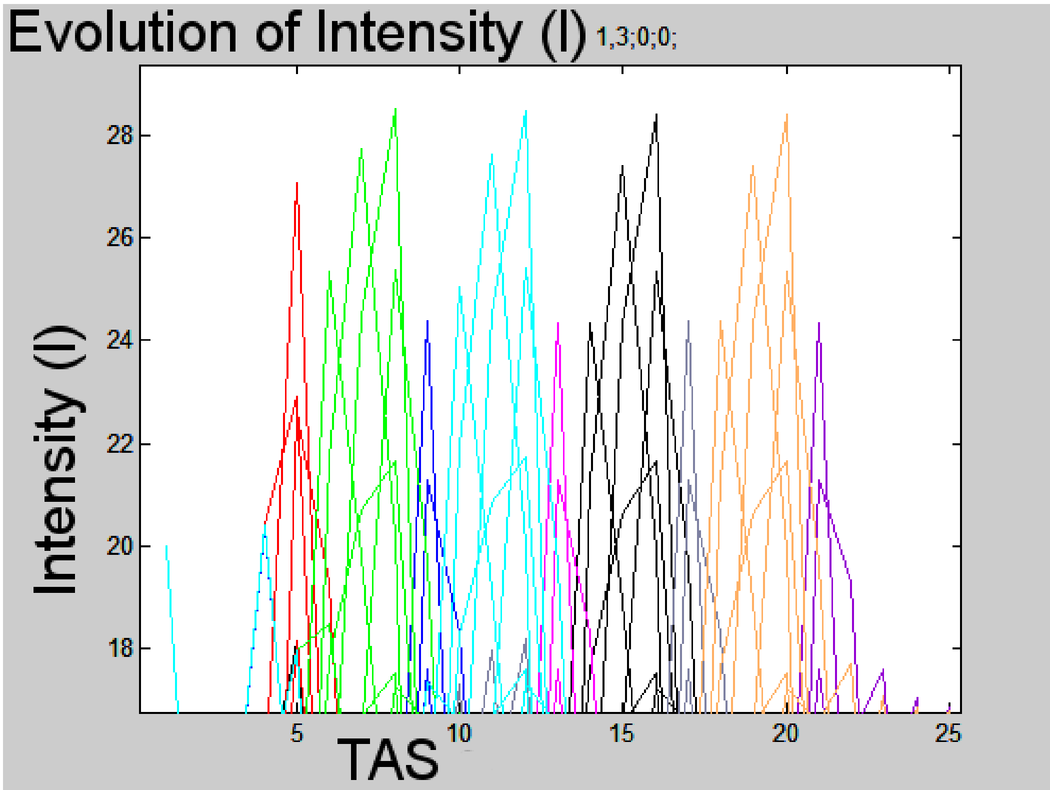

Figure 22: Five short stimuli (red, blue, magenta, gray, and purple) and four long ones are presented with no breaks between them. More explanation is in the text.

Fig. 23 shows the same stimuli with the break between them being 3 time units. It shows the same trend as Fig. 22.

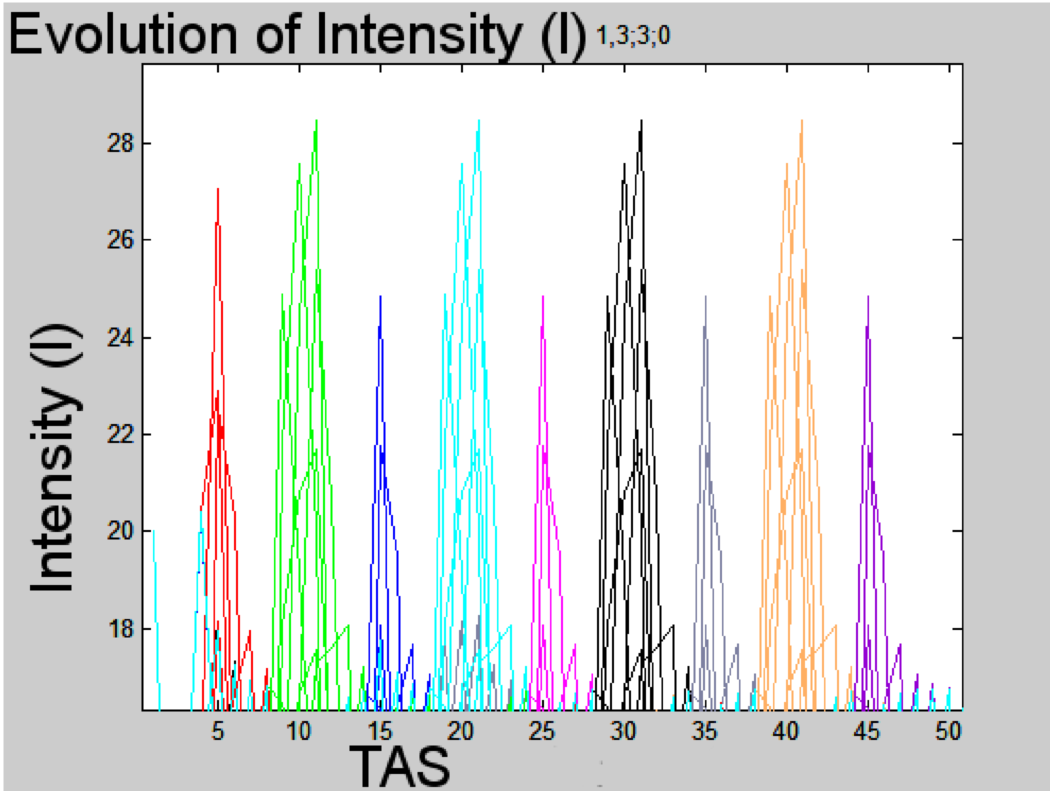

Figure 23: Five short stimuli (red, blue, magenta, gray, and purple) and four long ones are presented with 3-unit breaks between them. More explanation is in the text.

From previous nine figures, several interesting observations can now be made.

First, the serial position curves (SPC), which are discussed later in this dissertation, seem apparent from the trends of activations, assuming that the more activated the trace is, it is also easier to recall. The Primacy Effect (PE) is very pronounced and robust. The recency effect appears relatively weak and seems to be decreasing in strength proportionally to the word length. These results are present in simulations with a wide range of parameters (not presented here). This result is in line with experimental data, the trend that, with the following insights, lead to a discussion of the assumption about the simple activation's role, made in the above interpretations of recall probabilities. As often observed in the literature, these simple activations are not enough to explain recall probability but do play a role.

Second, the peak activations for long words are much higher than for the short words. If the more activated the trace is, it is also easier to recall, then the longer words should be easier to recall, which is the opposite of the WLE as defined above! Of course, this, too, necessitates re-examination of the assumption that the larger activation mostly alone means better recall.

Third insight, therefore, is extremely important: the gap between stimuli in the

Figures 9, 11, and 13 may be an important feature of results of simulations. This gap is related to the issue of the distinctiveness of traces, often brought up in literature on serial position curves.

These simulations show that if the activations and gaps get recorded in a two dimensional field which gets reactivated at recall and read out from the left to the right, along the TAS dimension, the longer stimuli almost always get interrupted by the shorter ones, before their spontaneous decay "erases" them. This makes both less unique, but it causes the short one start at the higher level. It is hard to say whether this equally affects both stimuli; this is likely highly variable based on parameter variations and input order. The short input, however, gets more intense in time and is also consequently relatively well separated from the next, the third one in this case, because of the (longer) break duration relative to its (shorter) durations.

The following figure (Fig. 24) takes into account only the activation of an input and the difference in the peak and the following gap in calculating distinctiveness for the mid-list items.

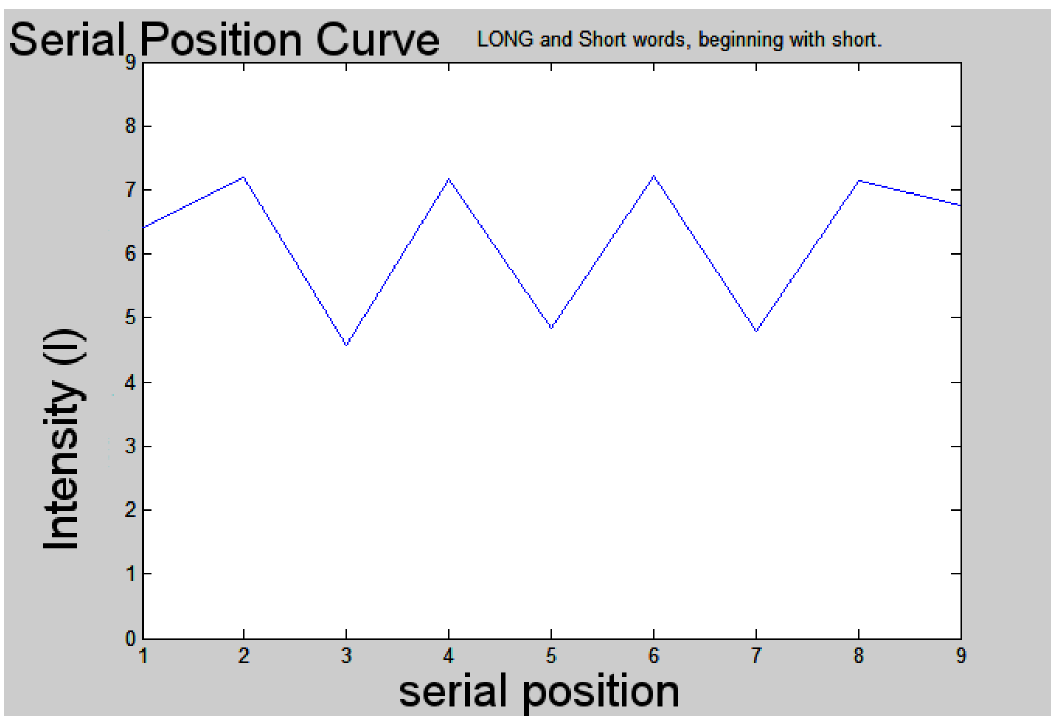

Figure 24: A simple calculation of distinctiveness: the intensity difference between the peak and the following gap.

The first and the last inputs, which are both short words in this case, are often better recalled by people than the mid list short words. According to this figure, however, the long words, should be recalled slightly better than the short words, which is not always the case in experimental situations. Two additional influences on distinctiveness should be thus important and evident here. First, the longer the input lasts (the wider the input is along the TAS dimension), the more activation are added to its *surroundings* along the conceptual neighborhood. This also reduces its distinctiveness if larger neighborhood gets recorded due to its above-threshold activation. This is the case even if short and long words are modeled by the same Gaussian inputs; it seems that longer words should occasionally be modeled by wider Gaussian inputs, which may further reduce their distinctiveness (but not necessarily always, as discussed below).

4.1.1 Intermediate discussion of distinctiveness

The distinctiveness, *as considered above*, which is always at least partially dependent on the activation of peaks, especially in the multiple fields models, is related to stimuli that are close neighbors in time. It should be clear, however, that while, say, the fourth stimulus is at its peak, all of the previous ones are also somewhat active. This additionally reduces the distinctiveness of the fourth one at recall, to some degree. In this case, the general trend is that the later stimuli have more competition from early ones as the series of inputs get longer. The simulations above have all the inputs of the same shape and height. This is a simplification of the reality but it is useful for isolating effects of other mechanisms of the model. In reality, the shapes of inputs along the conceptual dimension vary from one concept to another. This means that the heights of peaks in previous figures may follow different trends. However, it is reasonable to assume that the general rules discussed above will still hold - the first stimulus is always very active because it arrives to an undisturbed field. The intensities then diminish and this makes later stimuli somewhat less distinct - or harder to recall. The gaps between stimuli influence their distinctiveness, too. The presence of remains of previous stimuli additionally influence each stimulus, making later stimuli generally harder to distinguish from the rest of the stimuli. The first stimulus is not affected by preceding ones (since none precede it), and the last stimulus is unaffected by any following ones (since none follow it). Now the insights from the time-as-space dimension (and the temporal distinctiveness, e.g., Crowder's Telephone poles analogy, 1976) are generalized to the other dimension of the model - the space of concepts. Is seems reasonable to assume that similar rules apply to the conceptual dimension as well. For example, the words that have wider "neighborhoods" may need to have wider Gaussian inputs in simulations and are possibly less distinct from each other. Indeed, the literature on WLE adds evidence for this idea.

Words differing in one letter from a target word constitute its "orthographic neighborhood." Words that differ in one phoneme are the closest "phonological neighborhood" of the word. Semantically close words are the "semantic neighborhood" of the target words. The larger the neighborhood size, less distinct the word seems to be from its neighbors (for semantic neighborhood see Hunt & Hunt & Worthen, 2006; Elliott, 1980; Jacoby and Craik, 1979; Zechmeister, 1972; for phonological see (Coltheart, Davelaar, Jonasson, and Besner, 1977; Stone, Vanhoy, and Van Orden, 1997; for othoghraphic see Jalbert et al., 2011; Cortese, Watson, Wang, and Fugett, 2004; Glanc & Greene, 2007). Typically in word recall paradigms, the stimulus set is chosen to have words well separated along all of these dimensions. When this is not the case (on accident or on purpose), the experiments show that the recall is indeed influenced by all of the items mentioned above.

For example, on the one hand Luce, Pisoni, and Goldinger (1990), show that the words with larger orthographic neighborhoods are harder to recall and take longer time to identify from speech. In recognition memory, the word frequency effect is finding that low frequency words, which typically have smaller orthographic neighborhoods (Glanc and Greene, 2012), and I would argue smaller semantic ones, are easier to recognize probably at least partially due to their larger distinctiveness in a small neighborhood (Kinsbourne and George, 1974; Malmberg, Steyvers, Stephens, and Shiffrin, 2002). On the other hand, Jalbert et al., 2011, argue that shorter words, in WLE paradigms, also typically have larger neighborhood sizes (as a confound) and their increased probability of recall is due to their larger context-based availability so that this is the reason for the WLE, not really the duration of words.

Keeping in mind the simulations and the complexity of the distinctiveness, these findings seem to be in line with the CDIE analysis. Importantly the reduced distinctiveness may come from the fact that more durable stimuli disturb the field more and generally have smaller variations in peak activations of stimuli (Fig. 16.5), regardless of their orthographic or semantic neighborhoods.

The following figures take into account the complexity of distinctiveness, to which I return in detail again in the sections "Distinctiveness Revisited", and illustrate that the results of CDIE nicely fit to the experimental data. The long words are modeled with wider Gaussians (along the CS dimension, meaning larger orthographic / semantic / phonological etc. neighborhoods) than the short words in one simulation (Fig.26). Then this is compared to the simulation with the same Gaussians (Fig.25). The last figure shows classical way of presenting results: on the y-axis is the probability of recall and on the x-axis is order in time.

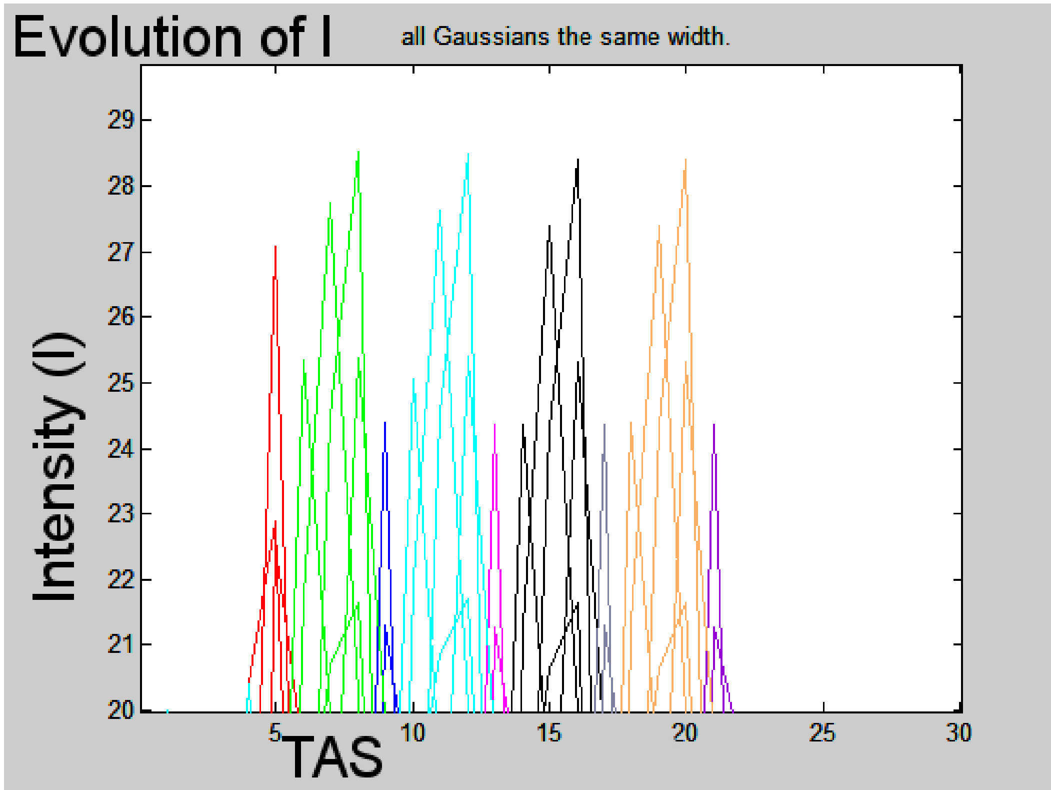

Figure 25: All the inputs are the same width along the CS dimension. Compare with Fig. 26

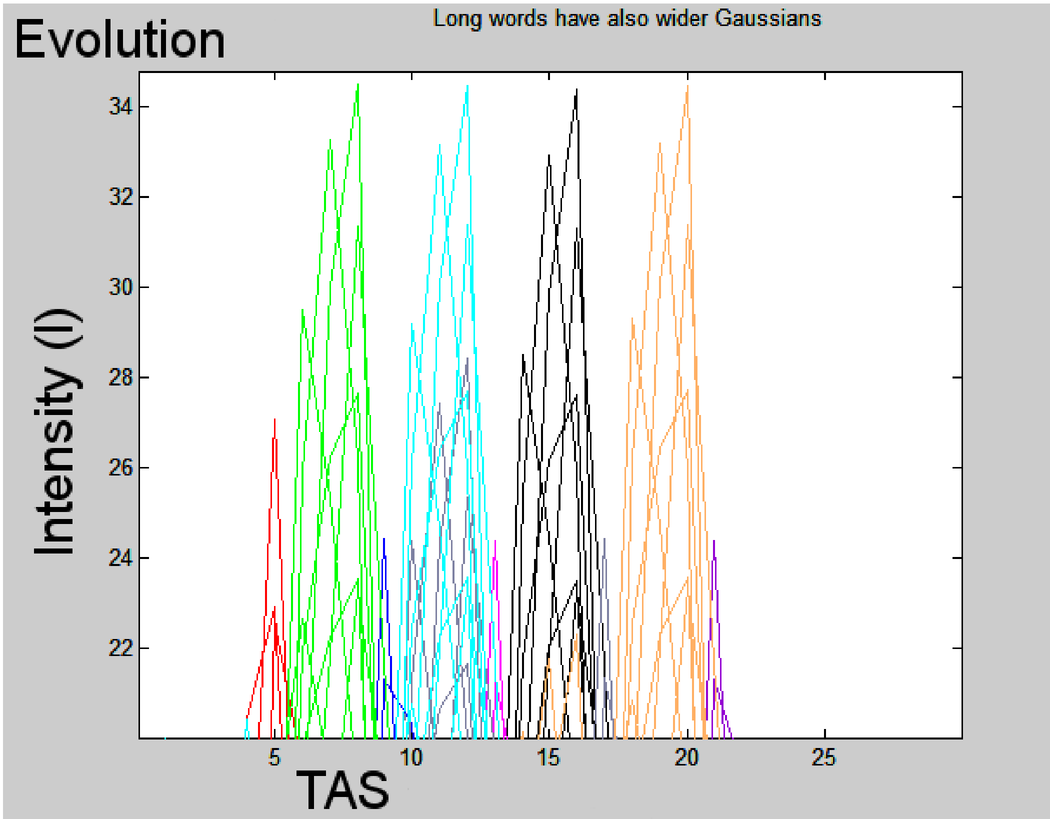

Figure 26: Longer words have wider Gaussians along the CS dimension than short words, and compared to long words in Fig.25.

Wider Gaussians of longer words add four main kinds of problems for distinctiveness.

- These words *may* have larger conceptual and other kinds of neighborhoods - more similar words that are highly activated, to be confused with even in the absence of other stimuli; this decreases distinctiveness.

- These words interfere more with the other words, making themselves and the following ones, the mid-list words, confusable with each other.

- The wider Gaussians add more energy, or heightens the peaks; this may increase the distinctiveness.

- Lastly, if the overlap between two concepts lasts long enough, it might produce a large non-input peak - a false memory. In any case, the mid-list items, and the long words especially, suffer from other influences in their evolution, which may or may not be compensated by their larger simple activation compared to shorter words. The lack of these issues for the first and the last stimulus, gives them an additional advantage (as well as to very isolated concepts along the other dimensions, if they exist). All of these complex interactions may result in better recall of longer words and the overall poor recall of mid-list words (Fig.27, below).

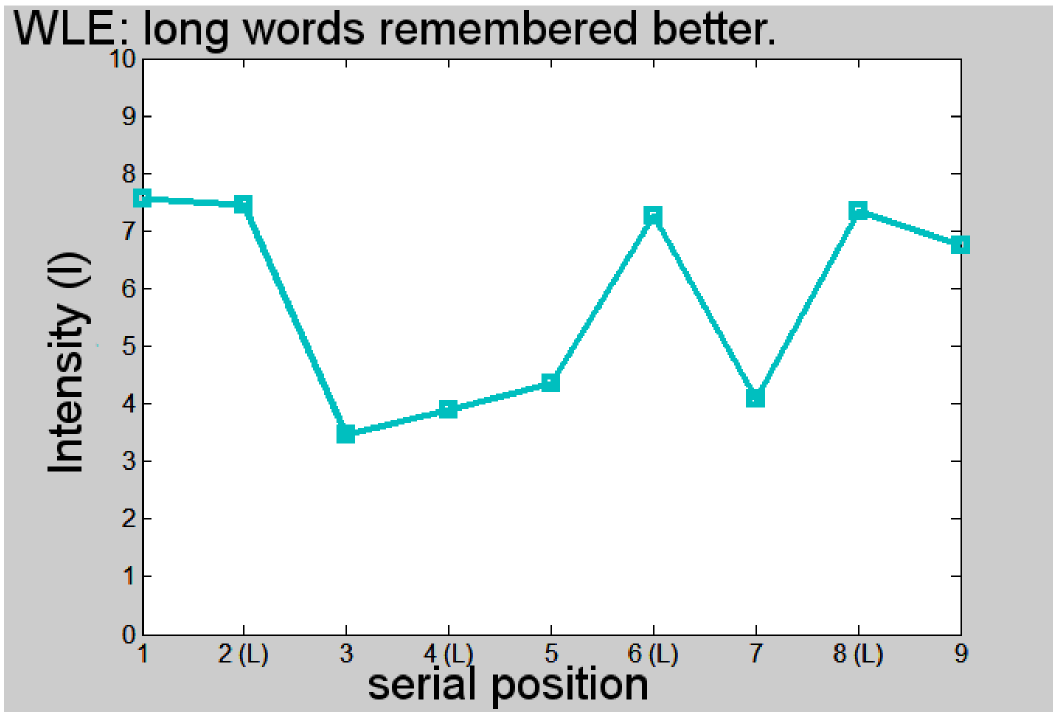

Figure 27: WLE in the CDIE fits well with experimental data for lists containing both long and short words. The long words' probability of recall is reduced due to their larger neighborhoods but their large peaks offset this, as illustrated in Fig.24.

In sum, in the CDIE model, the WLE can be nicely modeled and various results from experiments seem to logically follow from the complex nature of the process modeled in these simulations. We might see this as the greatest advantage of any new, different, kind of modeling - there is great potential for unifying many findings, not only within the WLE research, but across many episode processing phenomena.

For example, the False Memory (FM) phenomena are tightly related to the distinctiveness issues: When two Gaussian inputs are so close, the activation between their original peaks may end up being larger than any of the original peaks which simply yields a false memory (see the False Memory part of this text). Additionally related to the issue of gaps between items are issues of the distributed learning (e.g., Baddeley and Longman, 1978) and chunking in memory (see the Chunking part of this text). Finally, I would like to suggest, that the insights related to the high activation of long words may be useful in reasoning about the depth of processing in learning (Craik and Lochkart, 1972) - longer processing during learning may be analogous to the longer words development- the activations of many concepts for a longer time become high, which makes them, along with some delineation work later, good candidates for recall.

4.2 Amount of decay and the WLE

At this point, it would be useful to review the Phonological Loop (Baddeley and Hitch 1974), a concept often associated with word length effect. It is a part of the authors' theoretical account of working memory and it is assumed to be a place where stimuli are temporarily stored; it is a short term memory store. According to the authors, each item in this store decays in time, unless it is refreshed. The refreshing takes time, however, so if the list of traces is too long, some traces will be lost before they can be refreshed. Therefore, this short term memory store is of a limited capacity. Related to the WLE, this means that longer words during the refreshing allow for more time for other words to decay (Baddeley, Thompson and Buchanan, 1975). In addition, the task, of course, requires the words to be recalled - evidenced by saying them, for example. Longer words take longer time for the process to complete, which allows additional time for decay to operate (Cowan, Day, Saults, Keller, Johnson, and Flores, 1992).

I would like to argue that it is important to compare this idea of decay to a different one, present in a neurological phenomenon called the Transient Global Amnesia. The cause of the phenomenon is not completely clear but seems to be vascular. As the result of a temporary deprivation of blood, the patients seem to have an episodic memory lasting for several minutes and then suddenly disappearing. This constant "resetting" of memory may last several hours and then everything seems to return to normal (Gazzaniga, Ivry, Mangun). Memory literature on this phenomenon is sparse and sometimes anecdotal . More research is needed to shed light on episodic memory.

In line with this phenomenon is the research finding reported by Cowan et al. (1997) in which the memory for a tone was found to suddenly dramatically drop some time between 5 and 10 seconds. These examples of decay have different meaning than the decay as used in relation to the WLE. Here, the entire memory record, possibly containing many individually decaying memory traces, is lost after some time. This distinction is important to notice because the time limitation of a short term memory may be caused not by decay times and rates of each trace, as mentioned earlier, but by, say, bio-physical properties of tissues involved in this process (the CDIE model would argue for this); perhaps the neurological substrate of the STM is limited in its total activity by the rate of energy use or temperature, so that it simply "resets" itself on occasion to be reused. This is in line with ideas about roles of hippo campus and the brain cortex in research on Connectionist Networks (CN). O'Reilly and Munacata (2000) nicely show how they model hippocampal structures as those that have limited-duration role in memory while the cortical structures then accumulate information more slowly. In sum, the limitation of the STM does not necessarily mean that the decay operates on individual memory traces. Of course, neither does it mean it is absent. Some authors, as mentioned earlier, however, do relate immediate memory's capacity to items' decay rates in a very simple manner. The CDIE analysis of complex interactions suggests that this may be an inappropriate simplification of the processes and the interpretations of the experimental results may need to be reinterpreted. As mentioned earlier, many recent (or recently updated) models seem to begin including some kinds of time-based decay mechanisms in order to be able to account for experimental data.

4.2.1 Very fast spontaneous decay in CDIE

This section explores the role of spontaneous decay in more detail, as defined in the CDIE model. It is a spontaneous decay of every memory trace on its own that happens even in absence of any other trace present. In the literature, the WLE has alternative explanations both involving spontaneous trace decay and those not relying on it at all. Brown and Hulme (1995), suggested that long and short words have different spontaneous decay rates. In particular, the short words decay slower and hence remain activated longer, which makes them easier to recall. Alternately, Neath and Nairne (1995), do not use decay in their account of the WLE - long words have more parts and this leads to greater probability of wrongly combining them during the task. As we have seen, the CDIE suggests a lot more complex mechanisms in this task. For example, longer words have not only longer time to decay, but also longer time to build (higher) intensities. Longer words add more activation to the entire field, and while some of that activation inhibits other parts of the field, it may also give "a jump start" to stimuli that follow, and so on.

We have seen what happens with the WLE if the decay is almost non-existent. The following figure shows the opposite situation - when the spontaneous decay of each memory trace is extremely fast. On the vertical axis (height) is an intensity of an activation at specific point in CS-TAS space. On the horizontal axis is the position on the TAS dimension (same as time, here), the most left position being the earliest time. Stimuli were presented in the following order: Red, Green, Blue, Cyan. In figures, each color shows the entire evolution of a specific CS position. This is obvious from figures. This works the same way for short and long inputs, and with various breaks between stimuli (not all simulations shown here). Without decay, the entire field becomes almost evenly saturated (if not, say, overheated) without possibilities to differentiate among stimuli and with a very large decay, every trace is like the first one, which annihilates the primacy effect. The last stimulus, naturally, decays completely, too, which reduces the basis for the recency effect. This is like perfect memory.

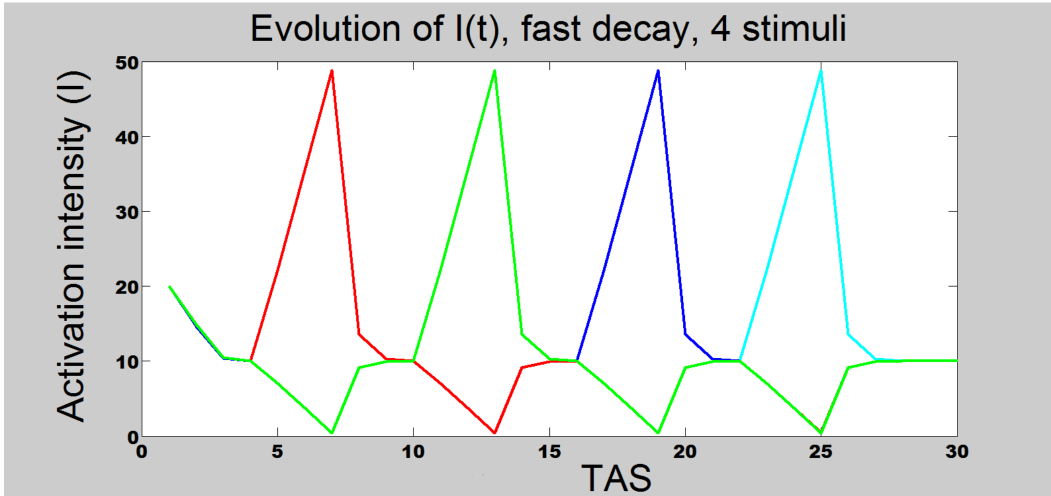

Figure 28: Very fast decay: The four stimuli of same duration are presented on at a time. The stimuli are identical to those in figures 29 and 30, but the decay speed of the system is different.

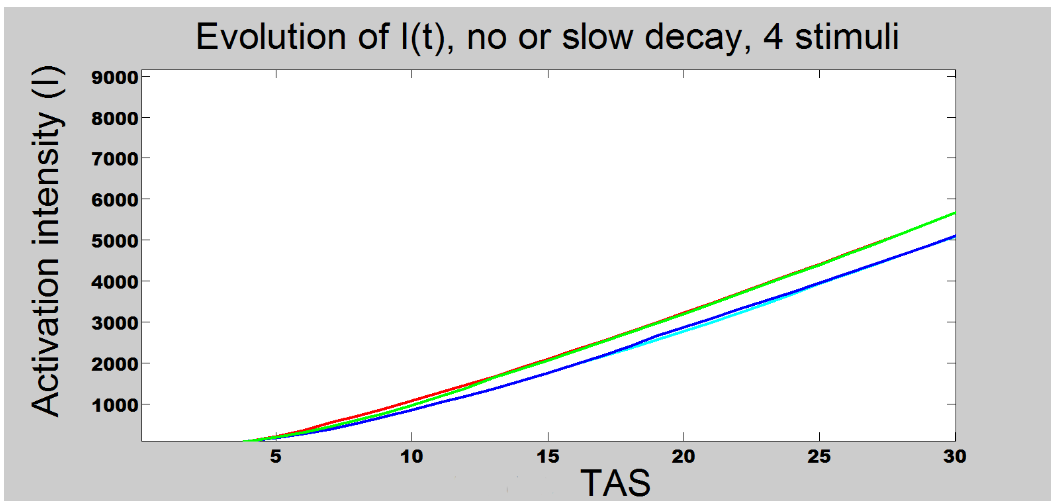

Figure 29: Very slow decay: The four stimuli of same duration are presented on at a time. The stimuli are identical to those in figures 28 and 30, but the decay speed of the system is different.

It soon becomes clear that some moderate decay is needed for obtaining differences among traces.

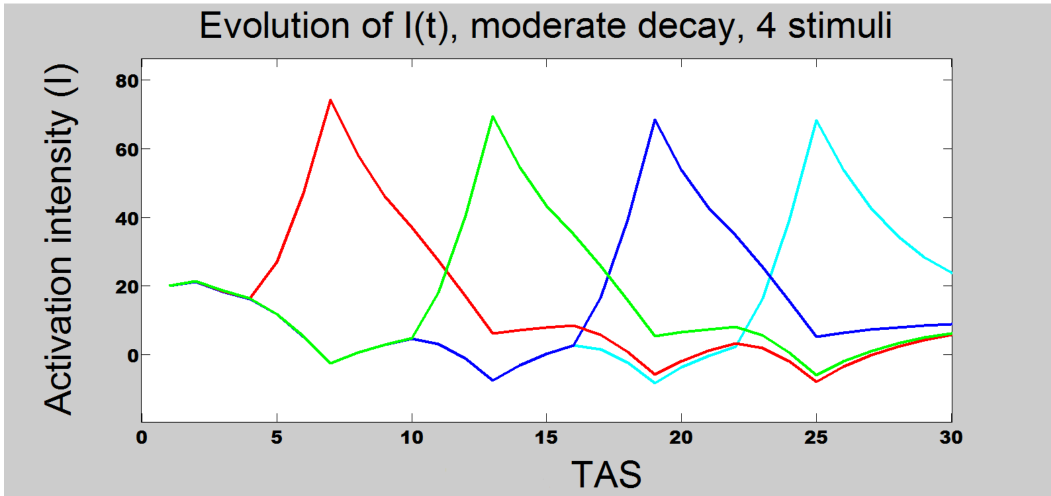

Figure 30: Moderate Decay: The four stimuli, exactly the same as those in figures 28 and 30 (Red, Green, Blue, Cyan,) and their activations (intensities) at each time step are presented. The primacy effect is obvious. Prolonged distinct activation of the last stimulus might contribute to the recency effect. The stimuli are identical to those in figures 28 and 29, but the decay speed of the system is different.

4.2.2 Perfect memory, Decay amount in CDIE and the Learning Rate and Penalty Terms in Connectionist Networks

In CN literature, it has been shown that feedforward networks often over-fit data- they learn to generalize patterns but begin to learn some accidental regularities in the data. Several ideas how to restrict this have been developed. One of them is an introduction of the Weight Decay term in learning algorithms. This term is conceptually analogous to the learning rate in weight updating process. Even thought the CDIE does not model this level of a complex dynamical system, the role of decay is very similar. Consider the following passage from: A Tutorial on Support Vector Machines for Pattern Recognition, Christopher J.C. Burges, Bell Laboratories, Lucent Technologies; 143 (c) Kluwer Academic Publishers, Boston. Manufactured in The Netherlands.:

> Roughly speaking, for a given learning task, with a given finite amount of training data, the best generalization performance will be achieved if the right balance is struck between the accuracy attained on that particular training set, and the “capacity” of the machine, that is, the ability of the machine to learn any training set without error. A machine with too much capacity is like a botanist with a photographic memory who, when presented with a new tree, concludes that it is not a tree because it has a different number of leaves from anything she has seen before; a machine

> with too little capacity is like the botanists lazy brother, who declares that if its green, its a tree. Neither can generalize well. The exploration and formalization of these concepts has resulted in one of the shining peaks of the theory of statistical learning (Vapnik, 1979).

In CN, Hinton, (1987) introduces the weight decay as the penalty term to reduce overgeneralization. In the CDIE simulations in this dissertation, The Gardener's Lazy Brother situation is analogous to the situation where the decay is completely absent - the memory traces are indistinguishable, everything is the same. The Botanist With a Photographic Memory situation is almost literally the one with too much decay- every trace is like the the first one, everything is completely separated, a photographic memory. The optimal performance is stuck only when increases and decreases in interactions are appropriately balanced. As mentioned, many working memory models are being updated to include moderate decay to achieve above mentioned balance.

For example, these and many more simulations analyzed together during the model development (e.g., Čadež & Heit, 2011; Čadež, Čadež, & Heit, 2010), yielded a suggestion that the item distinctiveness is a complex phenomenon and the decay analysis might be an important part of it. In the literature on the Word Length Effect researchers have investigated distinctiveness of items in time or in some variant of CS dimension but only more recently there have been suggestions that maybe much more than one dimension has to be taken into account when considering the distinctiveness (e.g., Jalbert, Neath, & Surprenant, 2011.; Larsen, Baddeley, & Andrade, 2000.).

I now turn to the closer examination of the Serial Position Curves in serial memory paradigms.

4.3 Serial Position Curves (SPC) in CDIE and Distinctiveness analysis.

4.3.1 The Primacy Effect

From Figures 9, 11, 13 and so on, it seems clear that the earlier stimuli typically arrive to less disturbed field. Depending on parameters of the field and inputs, these effects may be more or less pronounced but they seem to very robust. At this point it is worth repeating that the simulations here use stimuli that are identical in everything but the time of presentation and duration (for the WLE). Their initial intensities and shapes are the same. Even with this setup, the early stimuli have the advantage. This is completely in line with the Primacy Effect phenomenon in experimental findings (e.g., Postman and Phillips, 1965). In many recall paradigms there is a tendency for the first few items to be better recalled than the rest of the list (Ebbinghaus, 1913; Deese and Kaufman, 1957).

There are many ideas about the origin of the primacy effect. The early items are rehearsed the entire time and this is why they are recalled better (e.g., Rundus, 1971). However, for example, Howard and Kahana (2002) note that if primary items during presentation are also the first recalled, then they should not have the mentioned longer rehearsal time. Therefore, the rehearsal cannot be the only basis for the effect. Furthermore, Brown, Neath & Chater, (2007) suggest that the first item is very distinct from the following ones because it is not preceded by anything. Kamp, Forester, Murphy, Brumback and Donchin (2012), in electrophisiological research found that the first words and the "isolates" - a larger font word - seem to have similar electrophysiological memory effect and are recalled better as compared to other words in a list. This is interpreted as the contribution of distinctiveness to the PE. In research that does not directly investigate memory but deals with ordered information upon which other processes are built, the equivalent to the Primacy Effect is present: Anderson and Hubert, (1963), suggest that diminishing attention produces the primacy effect in memory which results in the primacy effects in personality impression formation.

The CDIE simulations would not disagree with these and many other existing ideas, but they would first add more insight in how many, if not all, of these phenomena are interrelated in the dynamics of memory. It should be noted here that in this dissertation, I did not use the temporal changes in effects of attention (which could be modeled using the decay term, as noted in the model explanation); the spatial effect were used only in one simulation, to explore false memory (later in this text).

4.3.2 The Recency Effect

If a person is asked to recall a list of words in any order, the last item or so also has an advantage in recall (e.g., Murdock, 1962). Glanzer, (1972) suggested that when people begin the recall the last few stimuli are still active in a Short Term Memory store and this is why they are recalled first and easily. Other researchers (e.g., Bjork and Whitten, 1974; Baddeley and Hitch, 1974; Crowder, 1976; Howard and Kahana, 1999) have shown that the Recency Effect is less robust than the Primacy Effect, that probably it is not entirely dependent on the STM, and that it it is present over different time scales along with the PE.

The CDIE model thus far suggests, in line with other research, that the Recency effect seems to be less robust and may be due to several different processes. The following two figures show evolutions of four stimuli in time, with slightly different parameters than in WLE simulations. The first figure shows shorter stimuli than the second figure. The main observation here, as well as in all of the WLE simulation with moderate decay, is that some number (five, in these figures) time units after the last input, the last input is still relatively highly active, although it did get reduced. There was not much interference along any dimension during this period - which is not the case with any other input. This is what makes this one additionally distinctive from the rest of the inputs.

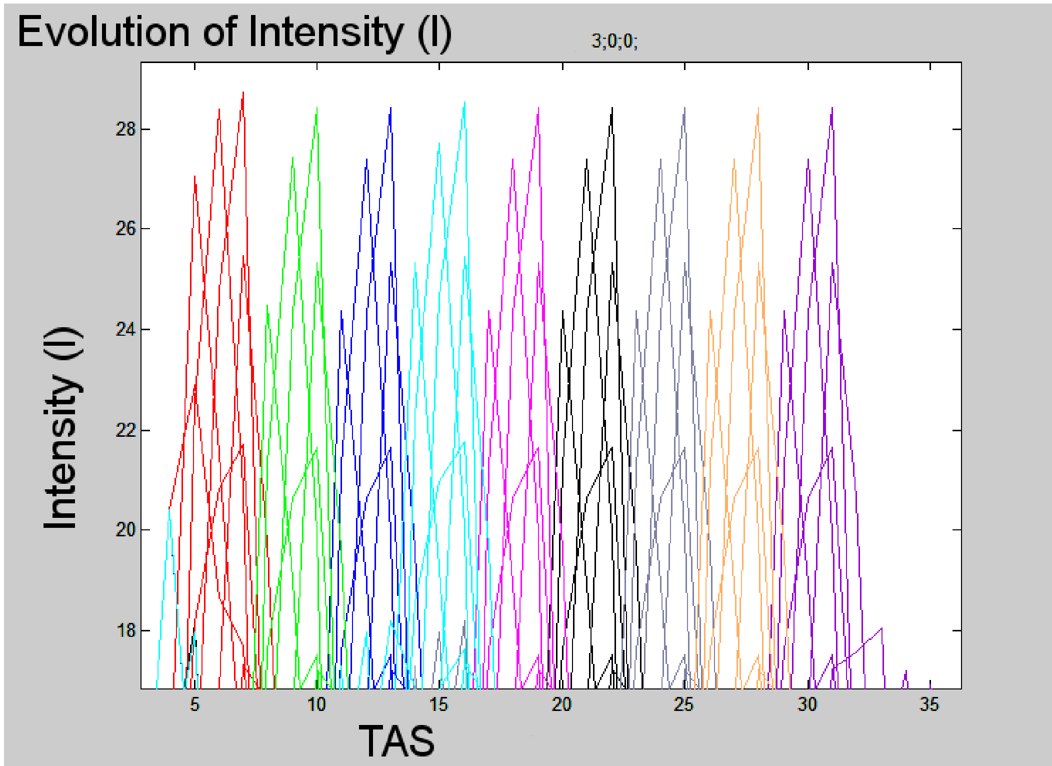

Figure 31: (same as Fig. 19.) Compare the last and the second to the last stimulus' evolution after they reach their peaks. The break between stimuli is 0.

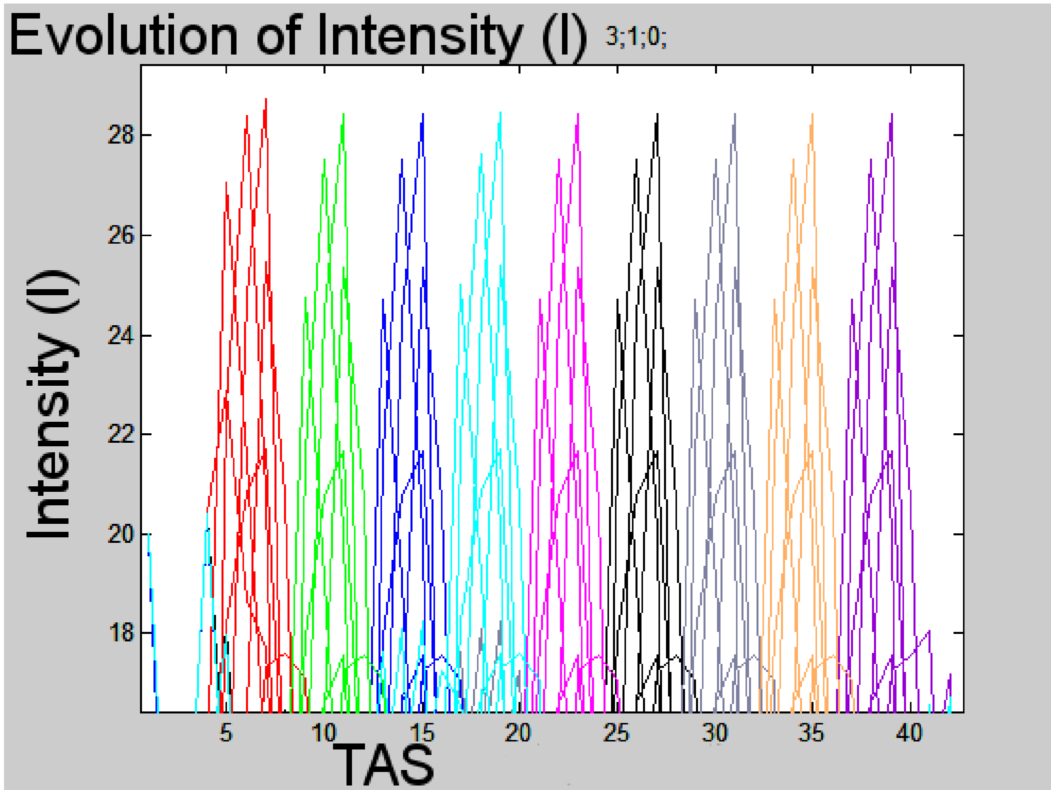

Figure 32: (same as Fig. 20.) Compare the last and second to the last stimulus' evolution after they reach their peaks. The break between stimuli is 1unit.

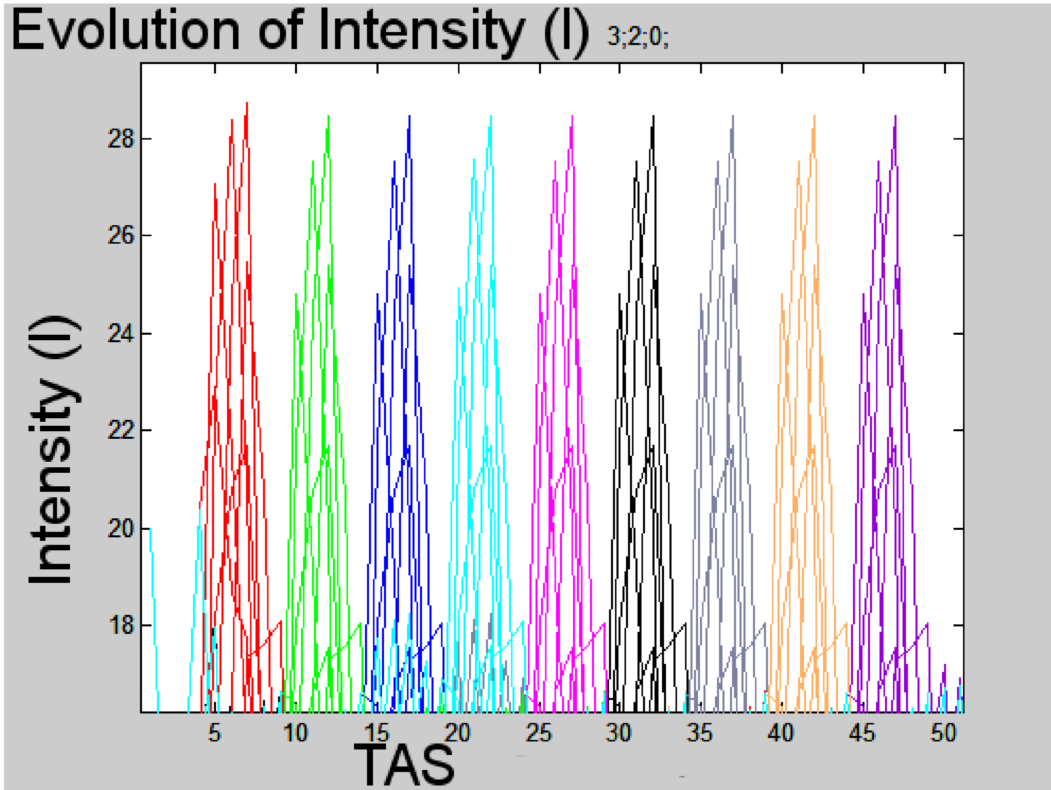

Figure 33: (same as Fig. 21.) Compare the last and second to the last stimulus' evolution after they reach their peaks. The break between stimuli is 2 units.

Figures 31, 32, and 33 show different stimulus length and presentation rate configurations and they all show that the last input, just like the rest of them except the first one, becomes more distinct from the previous ones as the break increases. Unlike the rest of the inputs even after some time after it reaches its peak activation, this input is still relatively high in activation while at the same time any other input is very low. The last one is, during this prolonged time, in fact, not extremely confusable with any other one. This, along with other influences on its distinctiveness may be the basis for the Recency effect.

In sum, the Primacy and the Recency effects have different bases in CDIE simulations, as is suggested by other authors as well. In addition, with the decay completely absent (very slow), as we have seen, the discriminability is completely lost for every input so neither the Primacy not the Recency effects are present. The same is true when there is very fast decay.

4.4 Distinctiveness Revisited

Above, I have discussed some general trends in dynamics of serial order memory evident from simulations using the differential equation of the CDIE model. When all inputs are modeled as identical Gaussians, the first stimulus is always very active compared to others because it arrives to an undisturbed field. The intensities then diminish and this makes later stimuli somewhat less distinct from each other, or harder to recall. The gaps between stimuli also influence their distinctiveness. The presence of the remains of previous stimuli further influences each stimulus, making later stimuli generally harder to distinguish from the remaining stimuli. In all previous simulations with nine inputs it is obvious that the fourth stimulus violates general trends. As mentioned before, this occurs because it is very close in concept space to the third stimulus. Because it comes after the third one, the fourth stimulus is influenced by the third one but the reverse influence is not necessarily symmetrical. In general, previous stimulus influences the activation of next one more than the other way around. The first stimulus is clearly not affected by preceding ones and the last one is unaffected by any following ones. In the part of this text about chunking, I argue that every time the subject thinks that the list is over, or a chunk of the list is over, the subject adds a longer break after the last input, regardless of whether the break is physically there. Finally, the insight from the time-as-space dimension is generalized to the other dimension of the model - the space of concepts. Notably, this dimension does not have the strong directionality of the time-as-space dimension, although it in principle it might have some directional biases.

Similarly to the spontaneous decay issue in memory research, based on these simulations it seems that the distinctiveness of stimuli is a more complicated concept than the ones currently used. Very often the distinctiveness is described as relatively simply diminishing quantity that changes inversely with the number of stimuli and/or with time - and often with real time and not psychological time- since the stimulus was presented. It seems quite reasonable to assume these relationships but the CDIE simulations reveal a more complicated story.

A successful serial recall directly depends on distinctiveness of a stimulus. The relationship between successful recall and the mentioned phenomena contributing to the distinctiveness may be roughly described from previous analyses:

- Distinctiveness in time-as-space is influenced strongly by directionality of time.
- Distinctiveness in general is directly proportional to the product of an intensity of an activation and its duration as the strongest one. The duration part here takes care of the influence of later stimuli to the older ones.
- Distinctiveness is directly proportional to the depth between peaks in a two dimensional space (along both of the two CDIE dimensions). It seems reasonable to take into account only the immediately surrounding peaks in this estimation, although other options are possible.

- In any parameter dimension the number of previous inputs is inversely proportional to the distinctiveness of any stimulus because of the residual activation of those during the evolution (proactive interference).

- Distinctiveness of a stimulus has two (or more) components, the two dimensions in CDIE model: distinctiveness along the concept dimension and along the time-as-space dimension.

A practical question, of course, seems to be whether it makes sense to try to precisely quantify all of these influences. The answer is: probably not. This issue is, however, not the main theme of this article, and it was only used to point towards a novel perspective DE models might open in research and to illustrate novel insights the CDIE model has produced already. Following examples continue this illustration of the CDIE model.

Insights in the processes in the complex system, however, tell a fuller story and show cases when this simplification is not appropriate: when conceptual neighborhoods of items are not all the same, when the initial intensities of items differ from each other dramatically (as in, say, words with high emotional valence), when items are conceptually very close, when items are very close in presentation time, etc. Some of these situations are modeled elsewhere in this work.

As mentioned before, figures 10, 12, and 14 above, do not take into account all the useful information about distinctiveness and therefore probability of recall from their respective simulation results presented in figures 9, 11 and 13. When additional adjustments are made according to the proposed simplification the well known shapes of curves are nicely identifiable.

First I compare serial position curves for long and short words. These are used for the convenience only and typical paradigms that examine serial position curves and forgetting of lists of words aim at keeping everything constant for inputs except their timing. Figures 25 and 26 are differ from Figures 9 and 13 only in that the fourth input (cyan) is not extremely close to the third one (blue) in conceptual space; along this dimension, all stimuli are completely distinct from each other.

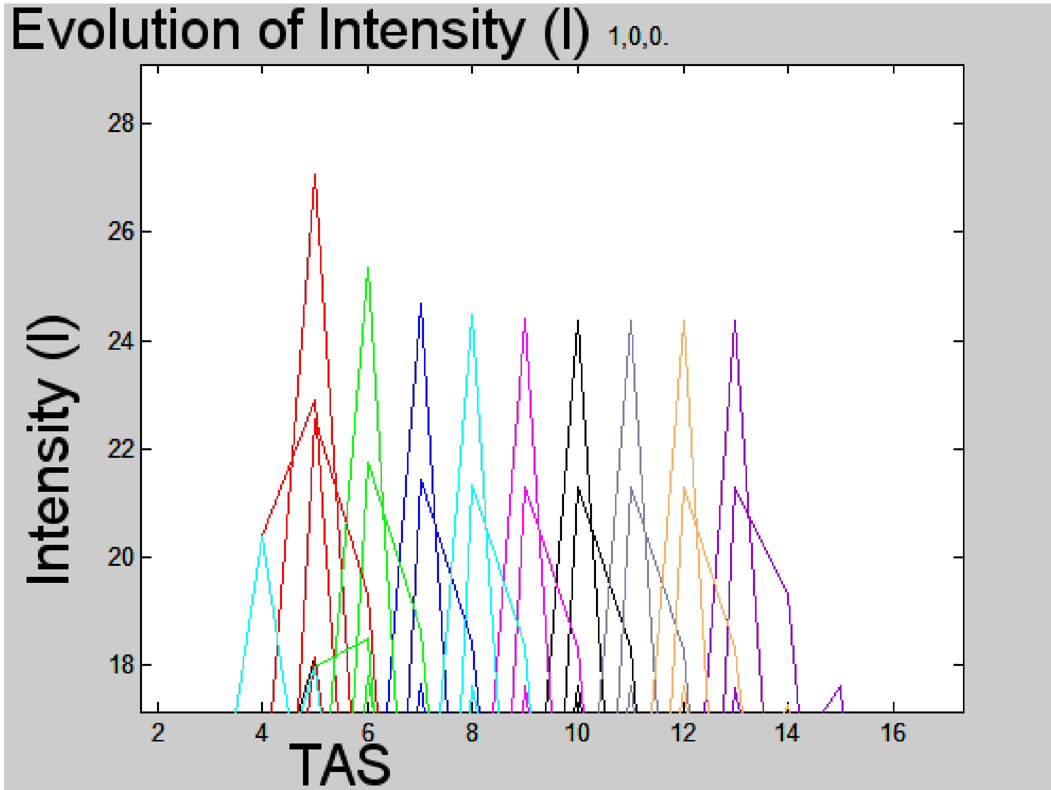

Figure 34: Serial presentation for short words (duration=1)

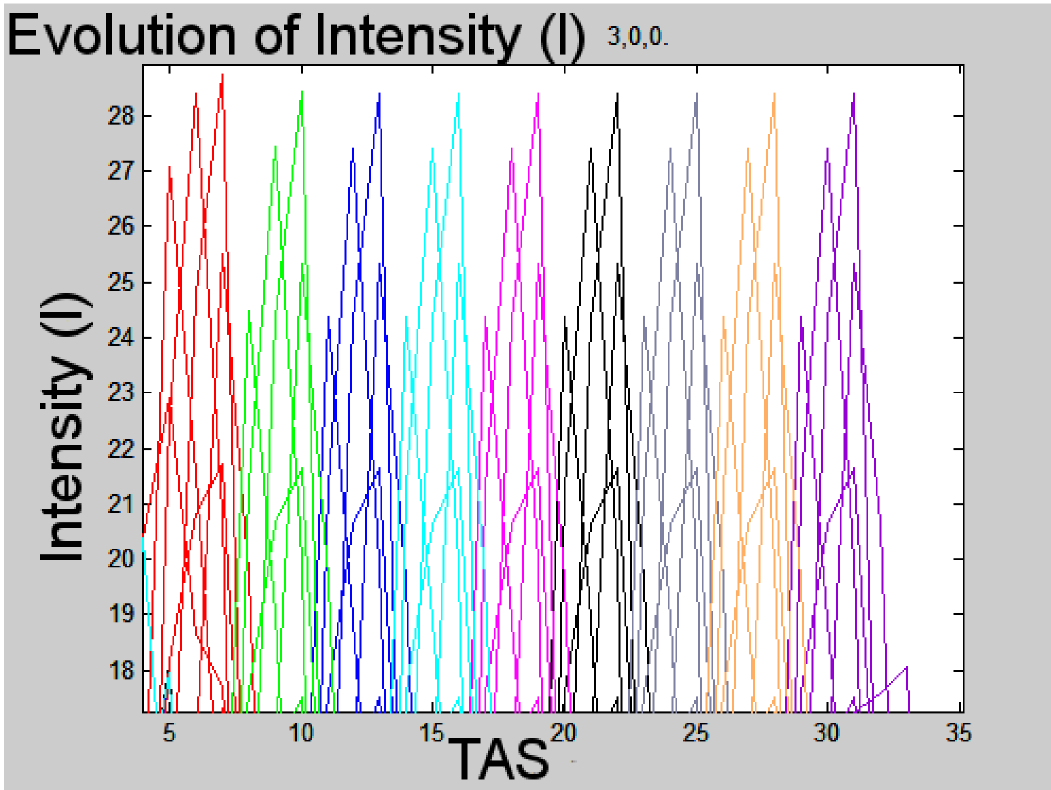

Figure 35: Serial presentation for long words (duration=3)

Corresponding classical representation of recall probability vs. serial position the following two figures (36 and 37). The Recall probability is obtained by using the actual activations. "*I*" of the peaks, as a difference in activations of a peak and its successive gap - the position when the next stimulus is equally as active as the previous one and the two are indistinguishable. This is only an approximate measure but it is sufficient to illustrate serial position curves for lists of all long and all short words (corresponding to figures 10 and 14, after the appropriate adjustments).

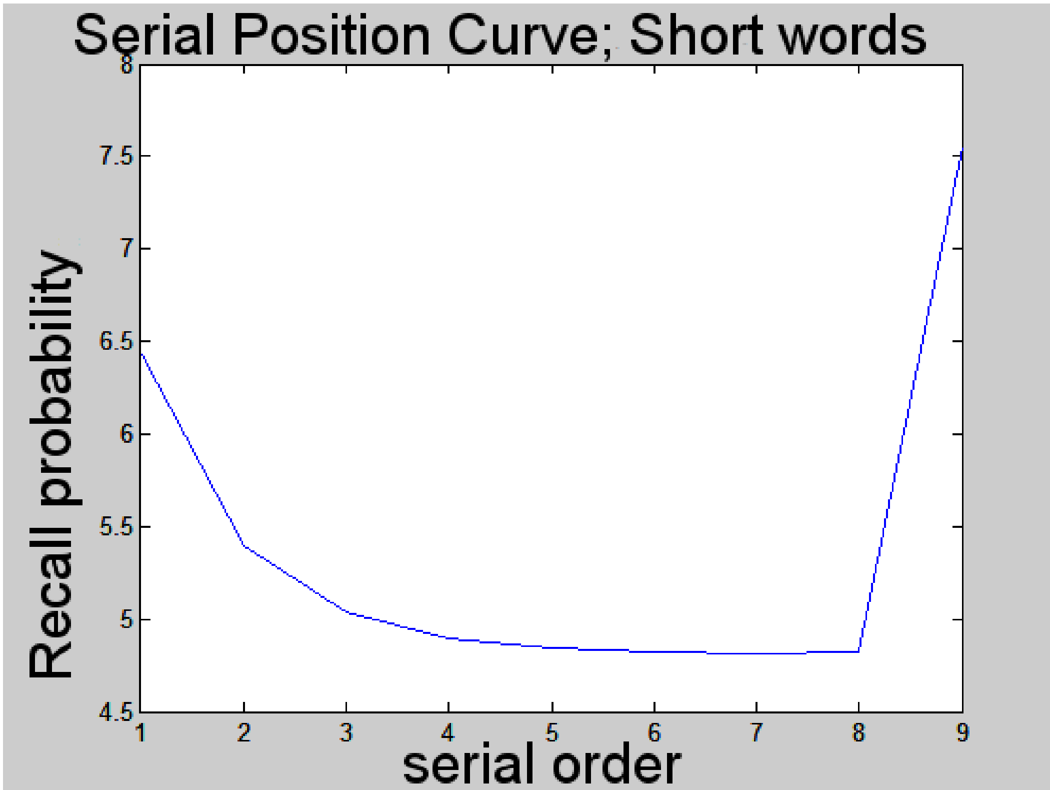

Figure 36: Serial Position Curve for short words (duration=1). Item presentation position is on the x-axis.

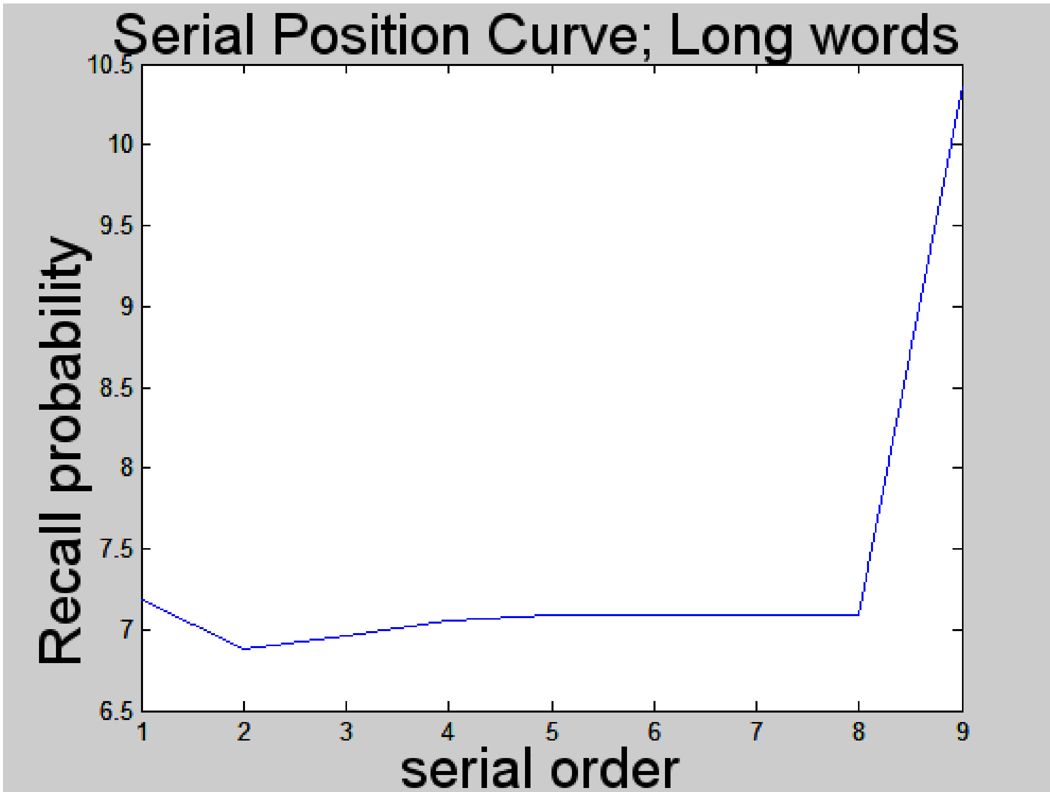

Figure 37: Serial Position Curve for short words (duration=3). Item presentation position is on the x-axis.

When comparing all simulations above, a relatively simple calculation for distinctiveness seems to be sufficient to obtain the serial position curves if the inputs are modeled by identical Gaussians and only differ in its time of presentation. This is true for a large range of ratios of input and break durations (not all shown here). In all of the simulations the first stimulus reaches higher peak activations because it arrives to the undisturbed field. After the first one, the differences between inputs decrease in all senses of distinctiveness mentioned above. In Figure 28 it is clear that the first three inputs vary considerably more than the remaining inputs. The only reason the last stimulus is different in these figures is that its distinctiveness takes into account that there is no more stimuli after it. This would give it an advantage for some time if it was read out relatively soon after its activation; inevitably, an input would come later (wherever these are registered) and the uniqueness of the “last one” would be lost.

Another trend should be noticed. As the duration of pauses between inputs gets larger compared to input duration, the peaks get more distinct, obviously (and this trend has a limiting break duration after which the break itself does not make a difference), but only temporally. Their distance along the other dimension of the model stays the same (at least for some time - I would argue that this is what changes in Learning and Conceptual Change), but in other sense of distinctiveness they are increasingly similar. The recency effect is not present.

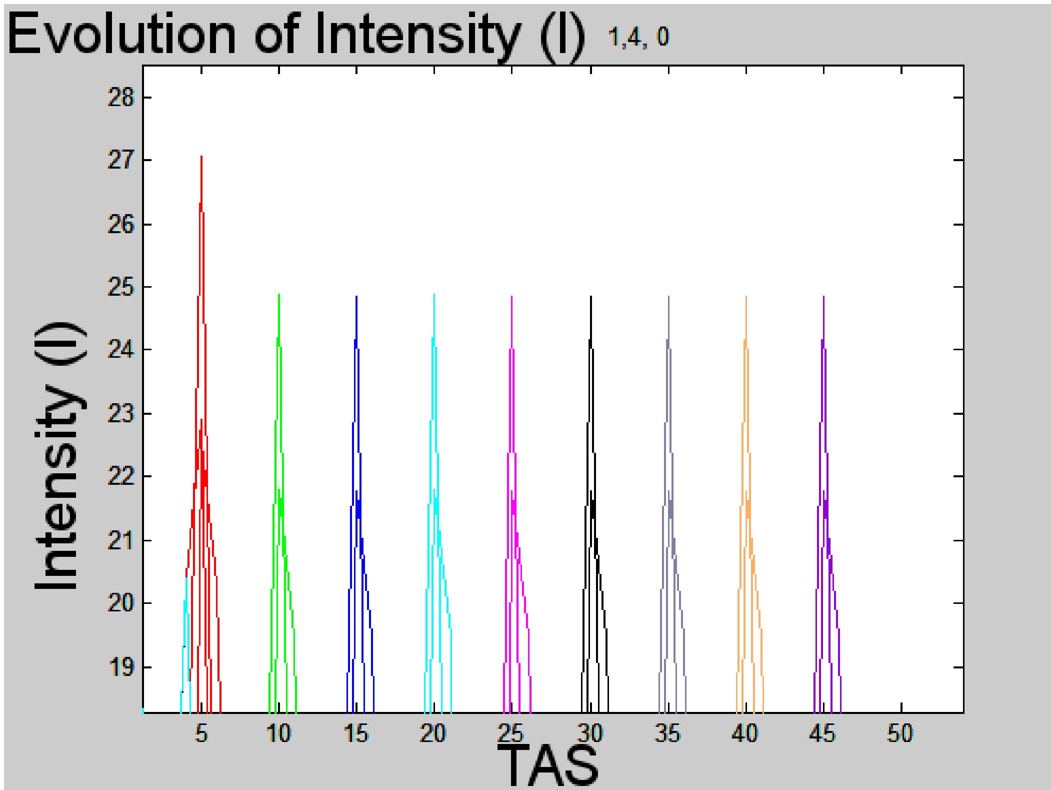

Figure 38: As the stimulus duration – break ratio gets smaller, stimuli get temporally very distinct but in terms of other features of their shape, they become almost identical. The recency effect is not present because of this.

5 Chunking and temporal gaps effects

Anderson, Bothell, Lebiere, and Matessa (1998) presented subjects each with several ordered lists of numerical items to remember. They later tested them on forward and backward recall of lists. Two subgroups of subjects received spatially but not temporally different presentations of stimuli. For both groups, spaces on a screen for each digit in every list were displayed at the beginning of learning activity so that the subjects immediately knew the length of the list. However, for one group, the lists were segregated, visually only, into sublists. The authors measured the recall speed and accuracy at the test. Accuracy was only affected by list length and not segregation. Speed of recall clearly showed segregation, with latency increasing whenever a new group was being recalled, both in forward and backward recall.

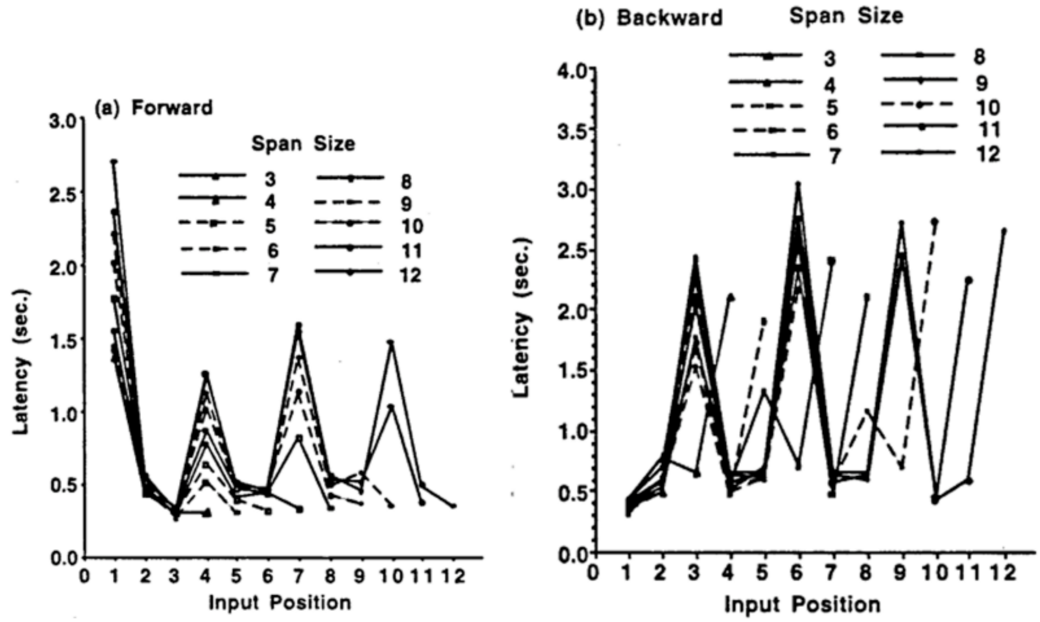

Figure 39: Forward and backward latencies from the experiments in Anderson et. al., 1998

This experiment was simulated in CDIE. Each item is represented as a Gaussian input above the x-axis with specific intensity (z-coordinate) at specific positions in time (y-coordinate) and space (x- coordinate of the mean of the Gaussian). The standard deviation of a Gaussian different from zero implies that each input may produce memory traces for some number of items very similar to the item presented. The effect of serial presentation of stimuli are simulated by adding one Gaussian input at the time, at chosen times, and then continuing simulating the time evolutions of all activations represented by their intensity.

This list memory experimental paradigm and this simulation assume that no two items on a list are extremely similar or dissimilar to each other. In the simulation, this is reflected in positions (both temporal and spatial) and shapes of Gaussian inputs (very small SD). The items were placed on the x-axis so that the distances (differences in means) between any two spatial neighbors are the same. In addition, the widths of items (SD) were the same for all items. These two choices pertain to equality of items in terms of their spatial position. The equal starting intensity, "I", for all the items means that no specific item on the list is an especially strong or weak memory.

The intensity is assumed to be directly proportional to both the accuracy of recall and to the speed of recall (e.g., ACT-R, SEM model in Henson, 1998). It is further assumed that both the maximal intensity and temporal coordinate of each input are registered by mathematically or physically separate system. The order information and spatial are obtained the same way.

In the Anderson et al. experiment the items were presented at the constant rate (equidistant along the time dimension) even though they were visually segregated. The visual segregation was modeled by adding a blank stimulus - a time delay at the position of the visual break. The following seems to justify the decision to do this: Anderson et al. report that the result for the unsegregated condition were qualitatively similar to those in the segregated condition. Therefore, either the segregation did not happen for the subjects in the segregated condition or it did actually happen for those in the unsegregated condition. The verbal reports on temporal chunking (e.g., Deese & Kaufman, 1957) in rehearsal together with results of this experiment made it more reasonable for us to simulate the latter alternative.

The blank stimulus seemed to be a good choice because the apparent segregation in list learning happened even in the absence of any external dividers. The time delay functioned well as this internal divider. The same blank stimulus was added to the beginning and the end of each list under the assumption that all chunks are equivalent. It is important to note here that these two delays do not affect logic of an explanation of serial position effects.

To get the overall serial position effects, the only assumption made is that all the stimuli in the list are reasonably close to each other, belong to the same x-space. Their belonging to the same list accomplishes this. The first stimulus arrives at the resting system and changes it for the remaining stimuli. This in turn changes the influences of those stimuli on the evolution of the first one and so on.

Any other stimulus but the first one arrives to the field that is most of the time more inhibited than at the beginning (due to the fact that the lateral excitation region is smaller than the lateral inhibition region). The interactions are complicated but the non-primary stimuli are thus generally of lower intensity than the primary one and their influence to their context is weaker. At the end of the list, the last stimulus, by the same logic, does not have any following stimuli to influence its evolution, which in this case also means less lateral inhibition arriving to it. The last stimulus, therefore, develops under substantially different conditions than any other before it.

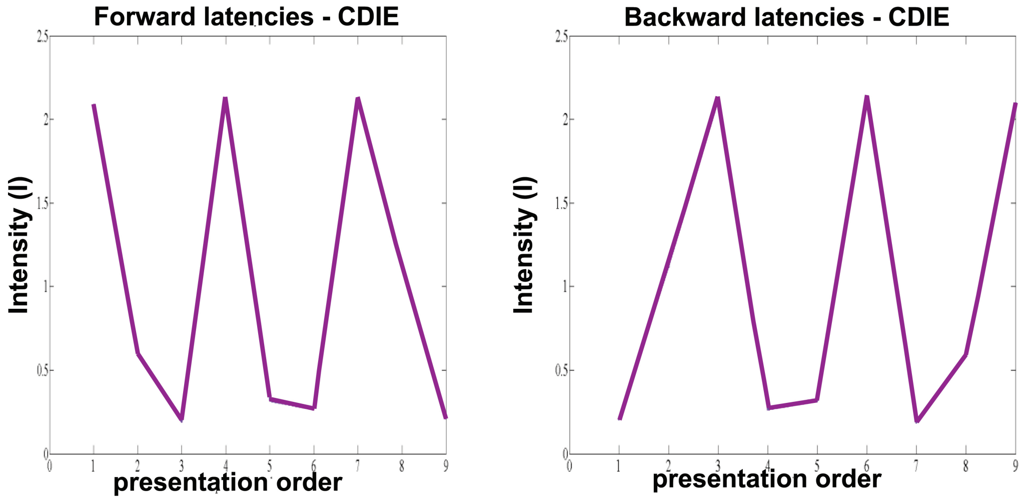

Figure 40: CDIE simulation nicely captures the results from Fig. 40.

In Fig. 40 are the results of the CDIE simulations of response latencies in forward and backward recall which fit well with the experimental data. In the dynamics of the complete task, the serial position effects are combined with the following: The pattern of changes in latencies seen over two chunks of stimuli repeats over any number of successive pairs of chunks. This is a result, for the most part, of the added break: the longer latencies of the beginning items of the chunks are given by the time needed to arrive to that item from the previous one or to the beginning of the task. Then this time is combined with the intensity of the item which gives rise to latency time inversely proportionate to it. This is the total latency in retrieval. In the simulation this was enough to replicate the results of the experiments and no rehearsal was involved. Therefore, one can conclude that the rehearsal may not be necessary to obtain positional effects in serial recall. Chunking may help memory by reducing detrimental interference of otherwise close stimuli, as predicted by the distinctiveness analyses above.

5.1 Conclusions from this section

The section argued for the following: First, the spontaneous decay should be precisely defined for episodic memory or any other research. Existing qualitative definition, although intuitive, allow for predictions that widely vary and may even produce opposite predictions. As a result, almost any experiment that is taken to show minuscule importance of spontaneous decay in memory may be reinterpreted or the procedure may be criticized for haven been inappropriate for a given test. In addition, the literature on time-based decay is often used to describe any kind of reduction of some intensity or activation. However, the reduction is most likely only partially caused by spontaneous decay and actually may include a lateral inhibition, various interference effects, etc. Interestingly, this complexity is often acknowledged. The precise definition might help reduce seeming contradiction as any mathematical model, it simply introduces rigor that clarifies hypothetical constructs of theories as well as experimental procedures and interpretations of results.

Second, as evident from the large amount of studies mentioned in the section, the verbal short term memory seems to be the main area in which the spontaneous decay is being attacked. Not many author, however, offer comments on this phenomenon, besides simply noting that other proposed immediate memory paradigms are actually finding spontaneous decay. Therefore, it seems that at lest some comments on biological or similar bases for this are in order.

Third, during presenting a number of studies it has been pointed out that they for various reasons do not seem to allow for the conclusion that the spontaneous decay is an unimportant phenomenon in the corresponding experiments. Of course, some of these studies were not intended to argue this, on the contrary, they argued for decay. It has also been illustrated that some authors seem to under-represent the role of the time-based decay.

Fourth, a good potion of mathematical models that are used to argue against the spontaneous decay, after closer examination seems to actually include some concepts that are very similar to decay. The authors do not comment on how are those different than decay. Moreover, some newer models and newer versions of older models, as well as the works reported about them seem to be more clearly including decay at various points. I believe that the mathematically well formulated notion of decay not necessarily the same as CDIE' s may greatly improve, and stimulate, the research on spontaneous decay episode processing by clarifying the result of model simulations.

Fifth, there seem to exist various insights from outside the traditional psychological and cognitive science research that may be used to inform research on the role of spontaneous decay in memory, attention, etc. Again, a precise definition or multiple ones may be useful here.

Finally, the CDIE model, as a model of a dynamical system that uses one possible differential equations approach to think about and simulate episode processing seems to have been useful in gaining insights about several classical areas of research. Notably, at this point, the role of decay in distinctiveness and the role of distinctiveness in recall have been examined from this different perspective and it produced arguably interesting insights about them. In the text that follows, more work along this line is presented.

6 The role of non-constant time-based decay in attention and false memory

In memory research, a rehearsal is sometimes assumed to improve memory that would otherwise be negatively affected by time-based decay (Baddeley, Thomson, & Buchanan, 1975; Baddeley, 1986). It is further possible to imagine that some other mechanisms may have a role in overcoming time-based decay. Cowan (1992, 1995) argued that within the pauses between the recalled words, subjects use the time to mentally scan the entire memory set, which is an attention-based refreshing process that could counteract time based decay. In other words, attention is focused on a word in recall while that word is being recalled but traces of all other words decay during that time. However, when the recall of the word is over, attention may be used to counteract decay of the other words. The authors note that this mechanism resembles the suggested memory scan mechanism coming from the recognition memory research. It is also possible that this refreshing operates only on one item, not the entire memory set (Johnson et al., 2002; Raye et al., 2002).

The intuitions about attentional refreshing just mentioned fit elegantly into the relation of attention and the spontaneous decay term of the general equation of CDIE.

Consider again spontaneous decay in the natural sciences. It is a case of some quantity' s time evolution in the absence of external influences or context (Eq. 9):

$$\frac{dI}{dt} = -\alpha\,(t,x)\,I\,(t,x)$$

In the space and time dependent attenuation case and, for example, where we have two inputs, the function $I(0,x)$ is basically a sum of two Gaussians (Fig. 41), each with a specially chosen amplitude and variance:

$$I\,(0,x) = \frac{I_{01}}{(\sigma_1\pi^{1/2})}^{-(x-x_{c1})^2/\sigma_1^2} + \frac{I_{02}}{(\sigma_2\pi^{1/2})}^{-(x-x_{c2})^2/\sigma_2^2} \tag{19}$$

In this example, the attenuation function α is:

$$\alpha\,(t,x) = \alpha_0\,[x-(1+bt)x_c]^2 + k_\alpha \tag{20}$$

This is a parabolic function (Fig. 12) whose location of the minimum moves in time at the rate defined by the coefficient b. The initial location of the minimum is at $x = x_c$. The coefficient τ is:

$$\tau \equiv \int_0^t \alpha\,(t',x)\,dt' = \tau_a\left\{(x-x_c)^3 - [x-(1+bt)x_c]^3\right\} + k_\alpha t \tag{21}$$

where $\tau_a \equiv \alpha_0/(3bx_c)$.

When these equations are used, the following patterns emerge:

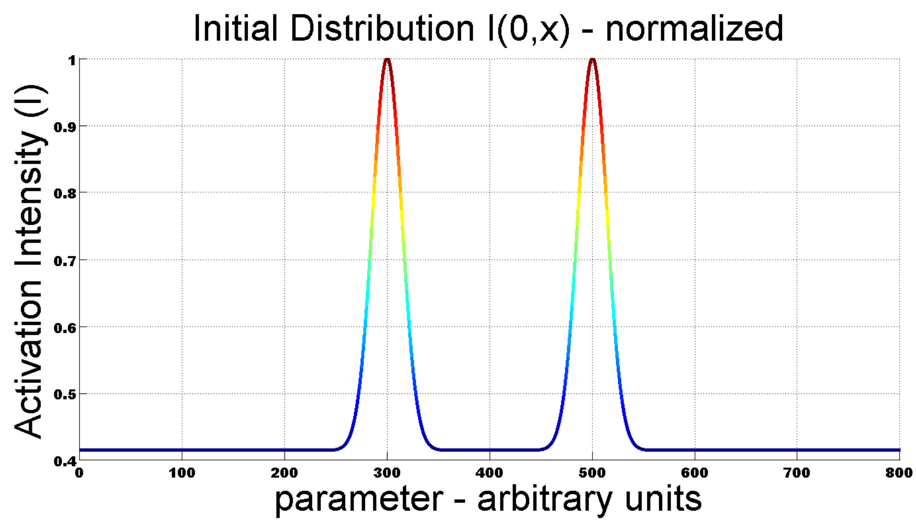

Figure 41: Two Gaussians representing two concepts, in this case, which are activated in the interaction field. Notice the area containing x-positions with low activations between the two peaks of activation. This indicates that the concepts are easily distinguishable and that no other concepts are significantly activated. This distribution of activations is not a necessary one, as discussed later in the text and depicted in Fig. 45

Fig. 41 depicts two activation peaks at points $x = 300$ and $x = 500$. These are two concepts that are, in this case, at such a distance from each other that they minimally influence each other. This is depicted by the lack of any significantly highly activated overlap or peaks between the original peaks.

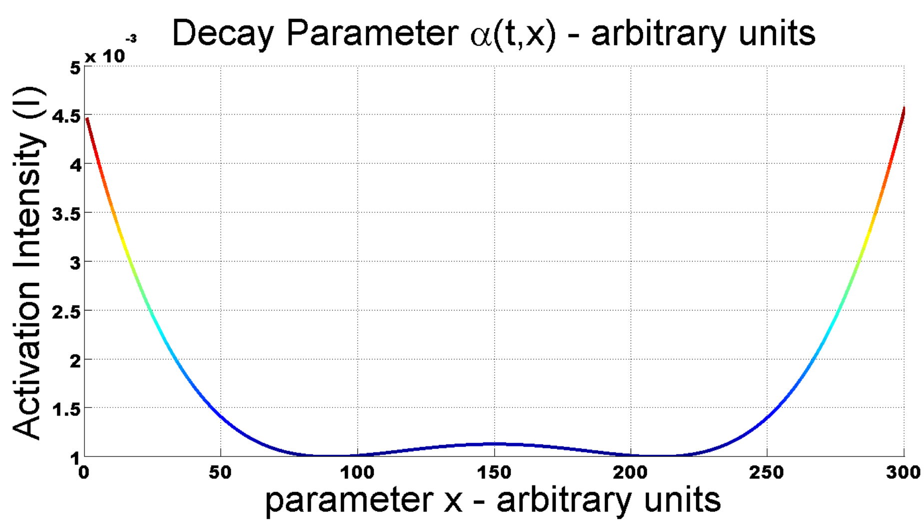

Figure 42: The shape of the attenuation function Alpha, $\alpha(x, t)$, along the x - axis. At each point, "x", the intensity of α is different indicating that some concepts are more potentiated or suppressed by it. (For more discussion about the consequences of this, see discussion below.) In addition, as explained in the text, the position of the minimum of the function α changes in time (not depicted in this figure).

At some point in time the attenuation function, simulating attention, may be introduced in the interaction field already containing the two Gaussians . The minimum of this function is to the left of both of the maxima / peaks, so that it asymmetrically influences them (Fig. 43 and Fig. 44). Note that the results are not a simple sum of these two Gaussians and other activations, but the sum obtained according to the CDIE equations for the interaction field. In addition, it should be noted that the sign of the attenuation function in principle may be positive or negative, so that it will either add or reduce activations of points in the field, depending on the purposes of modeling. When the entire process of attention acting upon the initial activations is simulated for some time period, the activation patterns shown in Fig. 43 and Fig. 44 emerge. The time evolution "flows" throughout the figures, from the top to the bottom. At early times, the two original high activations (marked red) are visible, as well as the relatively inactive region between them. However, as the time passes, the attention adds activation to the left peak more than to the right one. This interaction is combined with all other ones in the interaction field the entire time of the evolution. Because of these interactions, including the changing position of attention function in time, the highly activated positions of the field change. For example, at time $t = 650$, the right initial activation peak is completely suppressed – the concept is no longer activated an possibly non-recallable. On the other hand, at time around $t = 1600$ there is high activation at the parameter position $x = 400$, and the original peaks are only somewhat (possibly subtreshold) activated. This may represent the situation when a false memory appears due to the attentional process. Note that Figures 14 - 16 below show another, different possibility of false memories formation – the two initial Gaussians are close enough to begin with, so that the area between them becomes highly activated forming a false memory at that position. In this case, this "false" peak may be higher than the original peaks and the attention may play a role in suppressing this "false" peak. This nicely simulates the results of warning subjects about the possibility of forming false memories when remembering lists of related words, so that they pay attention and suppress them (see text below).

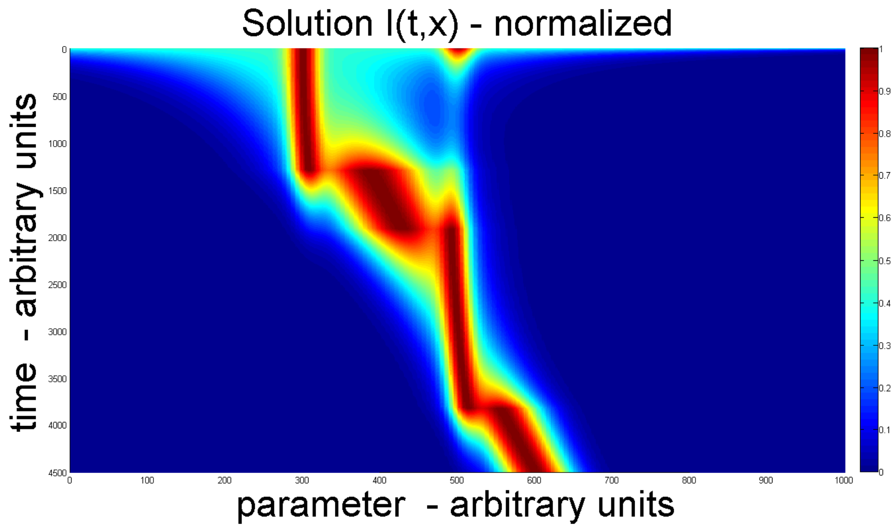

Figure 43: This image represents the normalized solution of the development of the CDIE sum of the two Gaussians. The distribution of activation along the x-axis changes as time progresses (from the top towards the bottom of the graph). Notice the emerging large activations between the original peaks of activation that appear after the attention has been applied for some time.

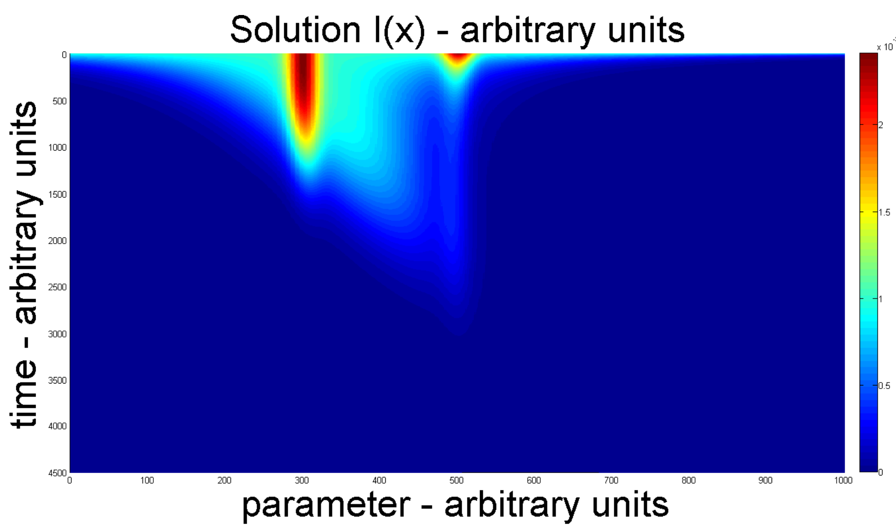

Figure 44: This image depicts the non-normalized solution of the development of the CDIE sum of the two Gaussians. In general, the process of normalization of intensities, "I", will change the contributions of points' intensities to each other in many ways due to the complexity of the system, as described by the CDIE equations, as compared to non-normalized solutions. This results in different evolutions of the same initial distribution.

As with any data set, simulations can be run with or without normalization of all intensities, "I", for example. The choices about this are made by authors based on theoretical assumptions or modeling needs (e.g., Reynolds & Heeger, 2009). In these simulations both are used, as figures show, and *importantly,* these are illustrating the difference that may be obtained based on the mentioned choices. In the case with normalized data, the final distribution does not even resemble the initial distribution here and may shorten the duration of traces in the field! Note that for the untransformed data, as well as, the original peaks of activation are not clearly preserved, albeit they are modified to the lesser extent in these simulations. This supports intuitions, for example, that from what one processes after some time, we cannot clearly see what that person has initially processed.

As mentioned above, the attenuation function of the decay term may be used to model the effects of attention in *suppressing* false memories. The following CDIE simulations show the simplified Deese - Roediger - McDermott (DRM) false memory paradigm (Deese, 1959; Roediger & McDermott, 1995). The typical finding in DRM experiments is that if participants are presented with a list of related concepts, they often falsely remember an additional related concept as having been presented in the list. The CDIE nicely captures this by presenting two closely related concepts as overlapping Gaussians, even when the TAS dimension is not present. When the overlap is significant, the middle point between the two original maxima reaches higher activation than any of the two originals. This may represent a very intense false memory. The question now may be how do these three memories (two true and one false) evolve? The results of the CDIE simulations are shown next.

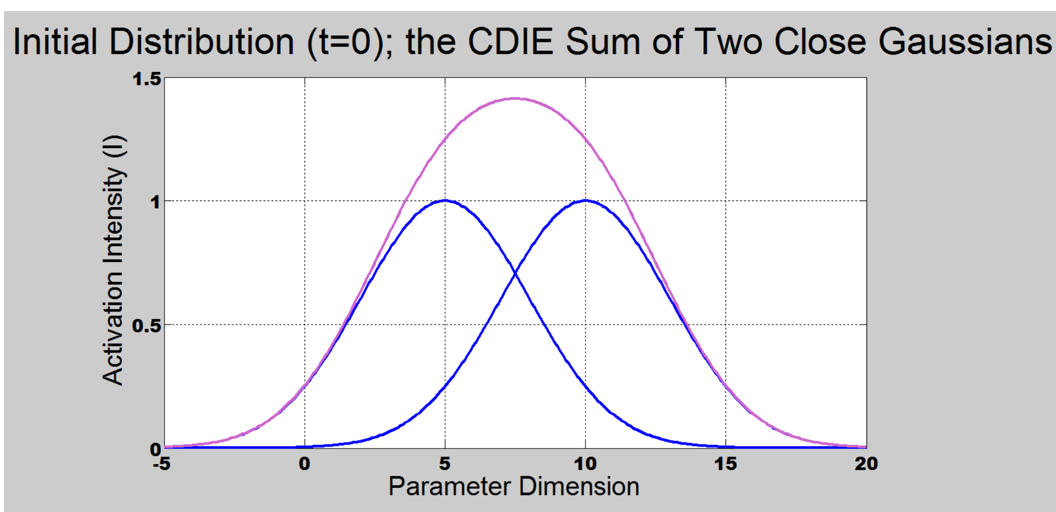

Figure 45: In BLUE: Two close Gaussians. Unlike the situation in Fig. 10, the two concepts activated here are very similar, as depicted by their positions. After they are introduced to the interaction field, they are integrated with each other and other activations and will evolve according to the CDIE equations. In PURPLE: CDIE sum of the two close Gaussians. An emerging highly activated peak forms between the original activation peaks. Unlike in the case depicted in Fig. 43 and 44, this is not the result of attention. However, as explained in the text, the attention may now be used to suppress the emergent activation peak.

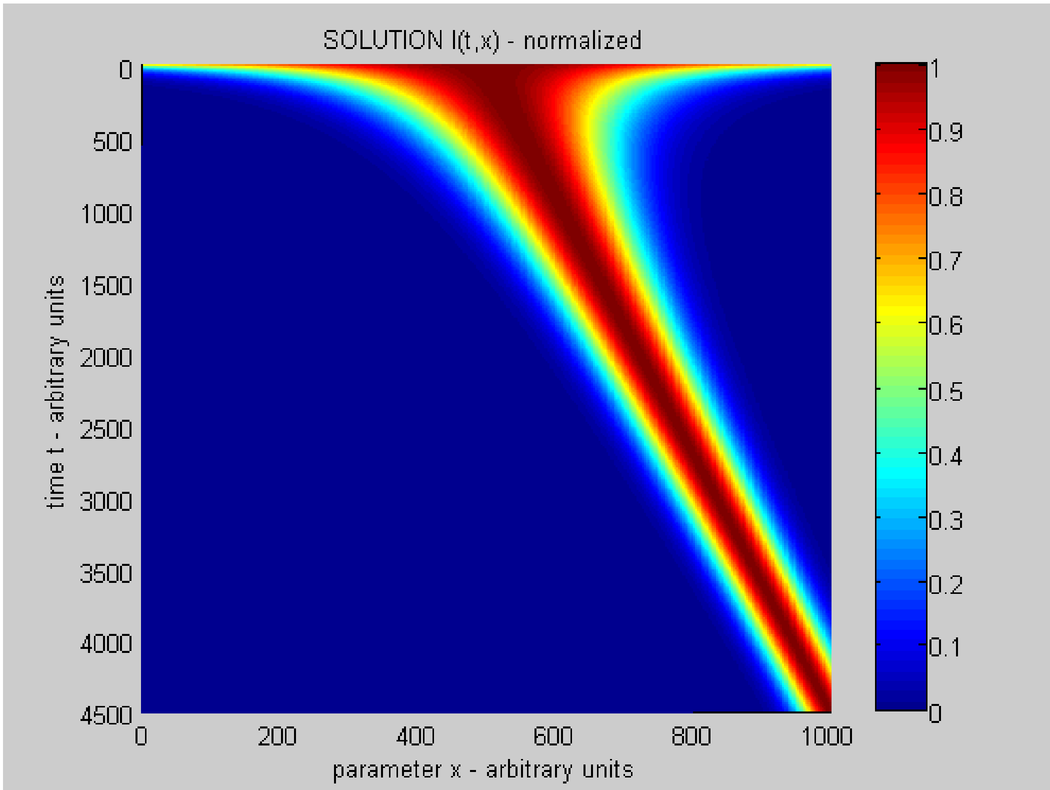

Figure 46: This image illustrates the normalized solution of the development of the CDIE sum of the two close Gaussians. This figure is analogous to the Fig. 43 in every way. The only difference is that the two initial Gaussians are closer together. More discussion is given in the text.

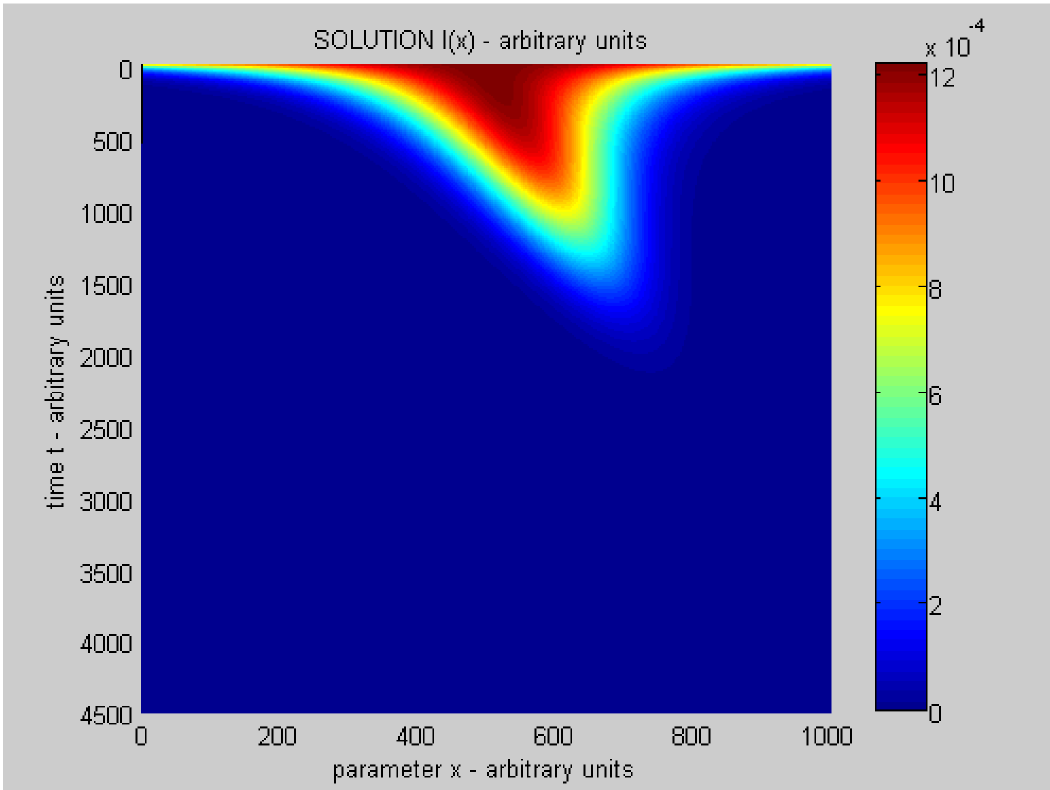

Figure 47: This is a non-normalized solution of the development of the CDIE sum of the two close Gaussians. This figure is analogous to the Fig. 44 in every way. The only difference is that the two initial Gaussians are closer together. More discussion is given in the text.

Plots in these figures show two true memories and a false memory forming between them as soon as both original peaks are activated. Again, the time evolution is depicted as the change of activations of points from the top to the bottom of the figures. During the evolution, the time and space dependent attention modifies decay and thus the evolution of the system. As the evolution progresses, highly active positions change, as before. At each point in time, as described, the most activated positions may be registered in another field, and remain there possibly, but not necessarily, along with previous highly activated positions. In the interactions fields shown here, traces of the episode that remain, again, may show interesting patterns of change, for example, continuous "drifting away" from the original activation peaks, but may also seem to change in a less "organized" way if the parameters are set differently. As mentioned, the experimental data constrain parameters. For example, the early version of CDIE was able to nicely simulate attentional suppression of false memories when instructions about possible false memories were given (Čadež, Čadež & Heit, 2010).

In sum, the attenuation function Alpha, a part of the spontaneous decay term of the model, if conceptualized as one kind of attention, may be responsible for eliciting or reducing false memories. Importantly, the role of spontaneous decay in short term memory is fairly controversial, as mentioned above, so it might be very useful to connect the two complementary areas of research via the CDIE, and this may allow for completely new insights which in turn may give new ideas for experimental work in episodic memory.

At this point it should be noted that the figures in this section show one parameter dimension as it evolves in continuous time. In principle, this parameter dimension may be either Conceptual Space (CS) or Time-As-Space (TAS), in this model, with two different decay terms meaning two different equations for attention. In addition, the nature of attention in a field that results when the CS and TAS field are combined is another theoretical issue. Keeping in mind that this model, as well as OSCAR (Brown et al., 2000), for example, postulate possibly more than one of these combined fields on different time scales, this approach seems to be a promising research program to continue in the future.

Besides using the modification of the spontaneous decay to model possible effects of attention, as mentioned earlier, one might imagine a different modeling of attention. Namely, a more direct top-down influence of attention mentioned in literature might be modeled as a part of the external source/sink, "S", in the general CDIE equation. Although this is still a part of complex integrated process, it is somewhat more straightforward the attention simply adds intensity to specific information. This model, however, now suggests that the attention, a different kind perhaps, or, again, is a more complex phenomenon than previously theorized. It does not simply change information intensity directly but changes the decay - the rate of change of intensity. Both influences completely change the entire dynamics of the complex system. The simpler version of attention was not simulated here as, if it is conceptualized as another input "S", it may be, for now, thought about using the rest of simulations.

Another important point of this demonstration is that it is hard to speculate about the effects of decay and attention (that can likely be modeled by the decay term) in a dynamical system without having simulations like these. This is for two reasons. First, it is impossible to devise experiments to measure this situation, and therefore it is hard to form intuitions about behavior of the system in this case. Second, there are too many interactions to follow only by theorizing about them. Therefore, any arguments about what to expect in experimental data when any part of the dynamical process is present or absent may greatly benefit from these simulations. Note here that the dissipation of attention (i.e., spontaneous decay) does not include any external inputs or the integration part from the CDIE equation. In other words, the temporal context as a whole is not taken into account yet in the above simulations. Taking it into account further complicates predictions but also sheds more light on possible attentional processes and makes simulations even more valuable.

After presenting these basic possibilities of obtaining false memories, a distinctive feature of the CDIE model should be revisited. Namely, as was mentioned earlier, the CS and TAS dimensions are combined in a fundamentally different way than in the DFT. Here, each concept at its position is a sum and not product of contributing inputs/information (Gaussians). The following two figures illustrate the difference.

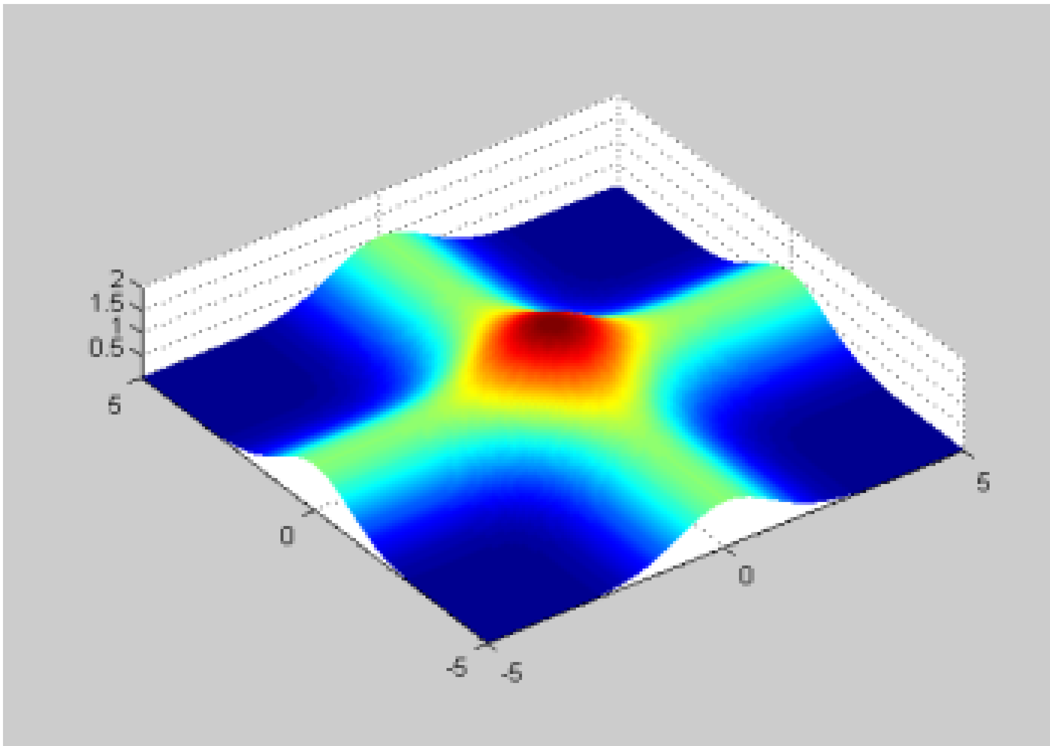

Figure 48: The sum of two one-dimensional Gaussians.

The formula of an one-dimensional Gaussian is:

$$f(x) = N\,\mathrm{e}^{-\frac{(x-x_0)^2}{2s^2}} - g$$

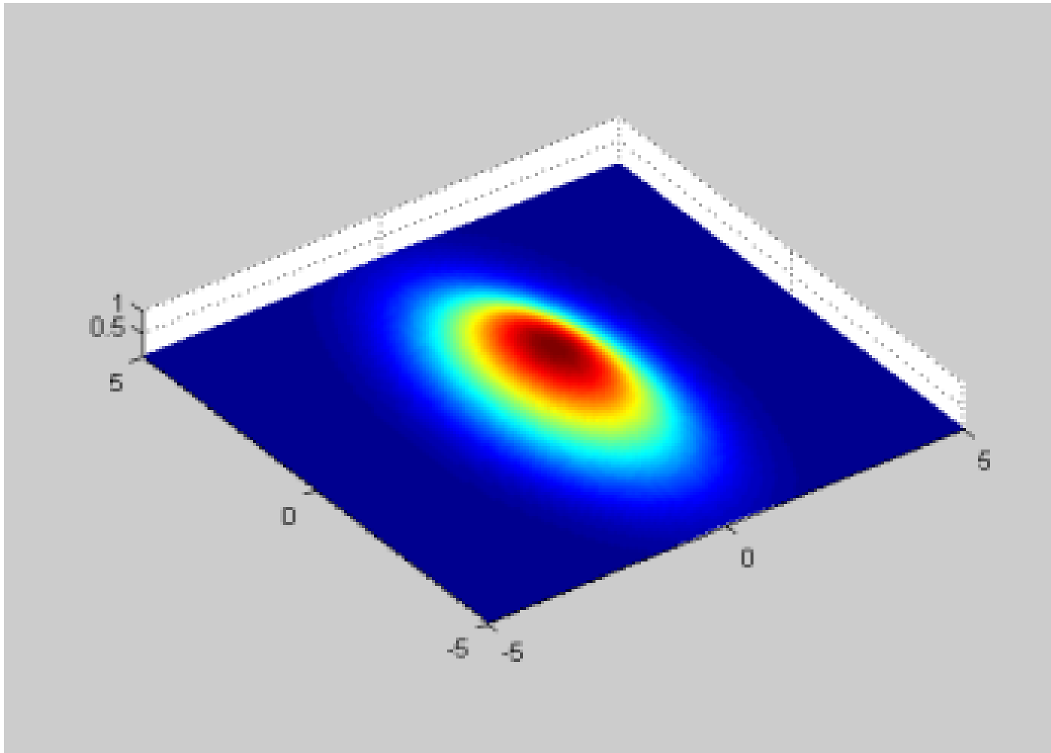

Figure 49: The two-dimensional Gaussian.

The formula of a two dimensional Gaussian is

$$f(x,y) = A\,\mathrm{e}^{-[a(x-x_0)^2+2b(x-x_0)(y-y_0)+c(y-y_0)^2]}$$

The profound consequence of this difference in CDIE model and DFT model is that when two or more activations are present in a "layer", if they are sufficiently close, as described earlier, they produce fundamentally different patterns of activation across the "layer." In the CDIE case, it only makes sense to obtain possible false memories/ false information at, as well as distinctions or confusability between, the specific order-concept points and dimensions (the horizontal and vertical color coded lines in Figure 3, pg 32 of this dissertation.

6.1 False Memory experimental results foreshadowed

Consider a phenomenon of false memory. The following pair of CDIE simulations show simplified Deese-Roediger-McDermott (DRM) false memory paradigm. The typical finding in the DRM experiments is that if participants are presented with a list of related concepts, they often falsely remember yet another related concept as being presented in the list.

The CDIE model nicely captures this by presenting two closely related concepts as overlapping Gaussians. When the overlap is significant, the middle point between the two original maxima reaches higher activation than any of the two originals. This may represent the very intense false memory.

I illustrate the use of the CDIE model in foreshadowing the results of an experiment used to examine the time evolution of a false memory by using recognition tasks. Heit et al. (2004) presented subjects with DRM lists to learn and recognize them. They measured the accuracy of recognition as a function of time. False memories were manipulated by instructions to be aware of them. In one of the conditions treating false memories as old items was discouraged (forewarning condition) while in another it was encouraged (inclusion condition). CDIE, including the decay modification function alpha (and only when it is not constant) is able to simulate the experimental results obtained by Heit et al., (2004). In these simulations, Alpha has the shape shown if Fig36 and it was only dependent on "x", and it may possibly be interpreted as attention:

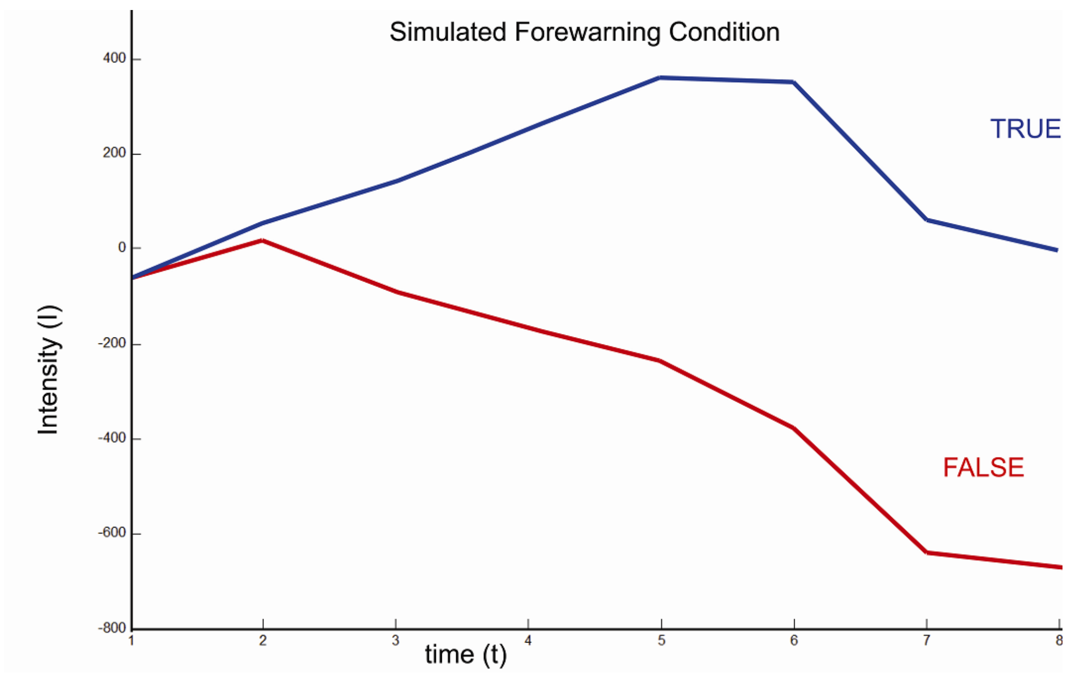

Figure 50: CDIE simulation of attention suppressing false memory in the Forewarning condition from Heit et al., (2004); simulation captures the experimental results very well.

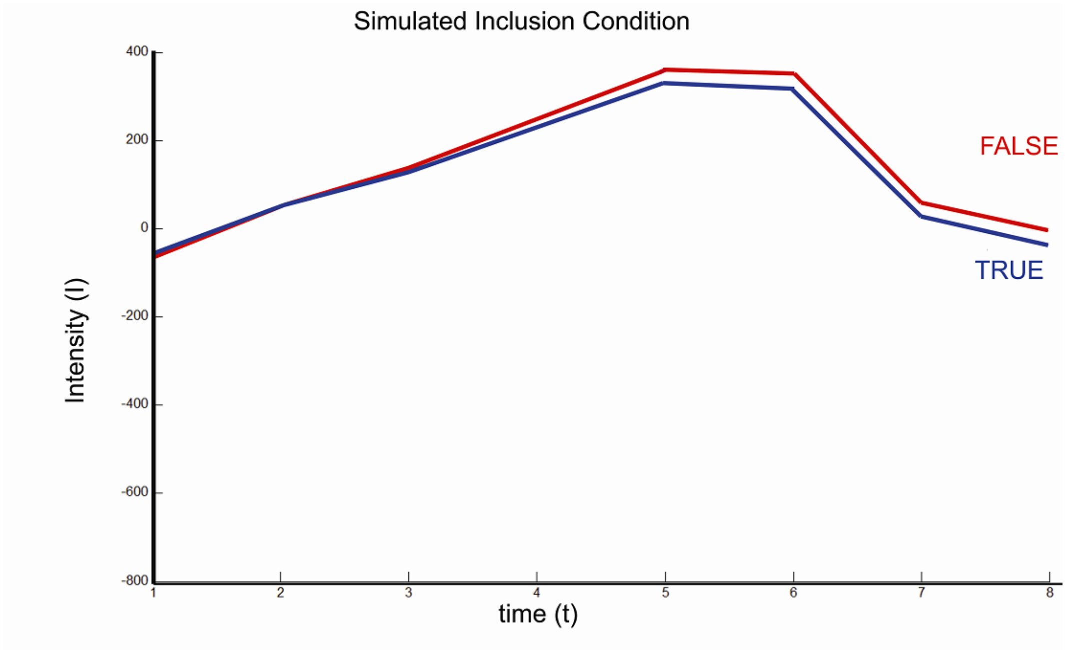

Figure 51: CDIE simulation of the Inclusion condition from Heit et al., (2004); simulation captures the experimental results very well.

In modeling this experiment, the x-position dependent Alpha was used to model the effects of attention drawn to false memories by warning subjects about them. Many researchers notice the role of attention in forgetting but they only offer intuitive explanations about how it influences memory (but see Grossberg, 2012). Specifically, the attention and decay often seem to be seen by researchers as closely connected. For example, the attentional refreshing has been argued to counteract decay (e.g., Cowan, 1992; Johnson et al., 2002; Raye et al.,2007; Waugh, 1961). In CDIE, the alpha term may be positive or negative to account for the attentional suppression of false memories and the attentional potentiation in refreshing, for example (section 1.2.b.ii). In the Lewandowsky, Duncan, & Brown (2004) SIMPLE model, attention plays a role as a weight given to the time dimension. Lewandowsky, Neemo, and Brown (2008) appeals to Browns use of the idea of primacy gradient: “Brown et al. (2000) justified the primacy gradient by appealing to the “intuition that each successive item ... is progressively less “surprising” or attention-demanding than the previous one” (p. 151),...”. Suggested here is a more rigorous mathematical definition for effects that may be due to attention. This might help in further examination of current models that would additionally specifically focus on the relation of decay and attention.

7 Forgetting Curves - Emergence

Forgetting trends, the initial fast loss of learned material and its slower decrease later, can be found in almost any memory task on any time scale (see Wixted & Ebbesen, 1991). Several mathematical functions can accommodate this general trend and therefore there is no consensus on whether an exponential, power, or some other non-linear function or combination of functions fits the data better than any other (e.g., Wixted & Ebbesen, 1991; Brown & Lewandowsky, 2010). Based on these shape similarities, Brown, Neath, & Chater (2007) suggested that the same kind of memory mechanisms may underlie episodic memory over various time scales and tasks: probed serial recall, free recall, immediate recognition, forward serial recall, etc. After reviewing decades of research devoted to finding the exact shape of the forgetting curve, Brown and Lewandowsky (2010) reported varying parameters in the Scale Invariant Memory, Perception, and Learning (SIMPLE; Brown, Neath, Chater, 2007) and obtaining several generally similar but not identical forgetting functions. They showed that parts of this function can be well described with one kind of curve while other parts of the same forgetting curve are better described by some other function. The authors concluded that finding the shape of The Forgetting Function might be an unobtainable goal.

Consider an evolution of several memory sets. Fig. 52 shows a simple case of evolution of several different and separate Gaussians. Each Gaussian is the activation of a set of entities in memory, which here may be thought of in different ways, depending on applications of the model: a set of features, a set of words, or some other complex knowledge. One may also think about these sets as representing the results of processing at different time scales. The use of a Gaussian function and continuous memory space is for mathematical convenience, not from a cognitive theory of how memories are distributed. Spontaneous decay of memory traces, interference from other memories and other possible sources of activation all integrated together guide the evolution of intensity of any activation, regardless of how a single Gaussian is interpreted, within this simulation. In following simulation (Fig. 52), all terms of the main CDIE equation are non-zero.

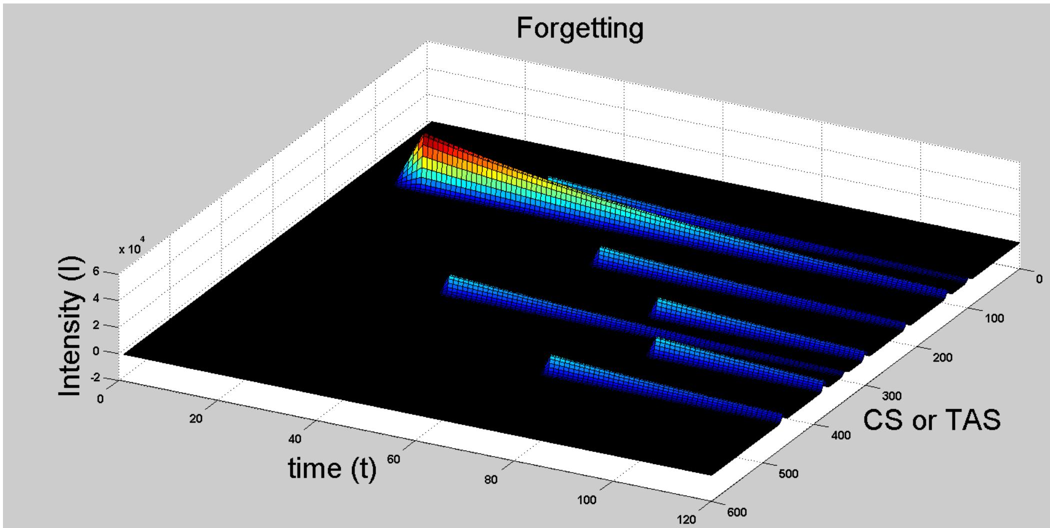

Figure 52: A possible set of learned material is being slowly forgotten over time.

The activations reach their peaks and if there is no new activation in their positions, they gradually loose intensity. Again, the intensity of a trace is calculated so that it includes all the influences of all other points along the x dimension, in accord with the main general equation.

Greater forgetting means lower magnitudes of remaining intensities, $I(x)$, after evolution. In Fig. 52, at the time t = 15 more x points have the activation above some level (which may be interpreted as a threshold for recall) than at the time t = 90. Relatively vague terms are used here to allow for variations in interpretation in different forgetting research paradigms while still pointing out important similarities in dynamics of the field. The amount of forgotten items, therefore, increases in time as a result of complex dynamics but the amount of forgotten items per unit of time decreases – this is a non-linear trend characteristic for forgetting curves.

To summarize, on the one hand, decay in an integrated system, without new inputs is enough to produce changes in the distribution of memories because different intensities change differently. Therefore, the effect of decay, even if it is the only mechanism of forgetting, is non-trivial. Importantly, the shape of decay, say an exponential function, is not necessarily the same as the forgetting curve. On the other hand, it may be possible to argue that the interference from other events may mimic decay both in this and the previous case. This would be very hard to do achieve in this model (previous simulations tried and failed to do so), due to its dynamics However, even without decay processes, the evolution is still guided by changes in integrated influences of the total context, along with new inputs, on the initial distribution. In both cases, therefore, the forgetting curve is an emergent property of a dynamical system. As mentioned, the shape of decay, say an exponential function, is not necessarily the same as the forgetting curve.

7.1 The shape of forgetting

It is clear that there is a non-linear decline in the amount of remembered material but it is not clear what function exactly may model this decline. This trend seems to be found in many forgetting situations: probed serial recall, free recall, immediate recognition, forward serial recall, etc. However, it is explained here, fitting a single function to the forgetting data has been and it may be misleading in our quest to understand the mechanisms of memory and forgetting.

Traditionally, a large discussion on forgetting was (and still is) about the shape of the forgetting function. In particular, it is about Jost' s second law of forgetting: If two memory traces are of equal strength but different ages, the older one will decay less rapidly in a given period of time than the younger one. (Youtz, 1941)

"If you can remember 100 French vocabulary words from your school education 20 years ago, and I can remember 100 French vocabulary words from 200 I learnt yesterday, it seems likely that you will remember more words than I do in a weeks time (assuming that neither of us engages in any further learning in the meantime)." (Brown and Lewandowsky, 2010)

Two well known functions are usually mentioned as potential fits for models of memory, the decreasing exponential and power functions. The amount of forgotten material decreases in time according to both of these. It has been argued that the exponential function is not actually capable of dealing with the subtleties evident it the French words example. This does not seem to be the case.

First, I comment on the properties of the exponential curve that are traditionally used to argue about its inadequacy for fitting the forgetting data and on a property of the power function which makes it a more attractive choice.

Herbert Simon put the first argument forward in 1966. If two exponential functions have the same half life (mathematically related to Alpha) the graph of the "older" function will always be under the graph of the "recent" one. It is never going to be the case that a person with older knowledge will remember more words at the same time point as the person with the newer knowledge.

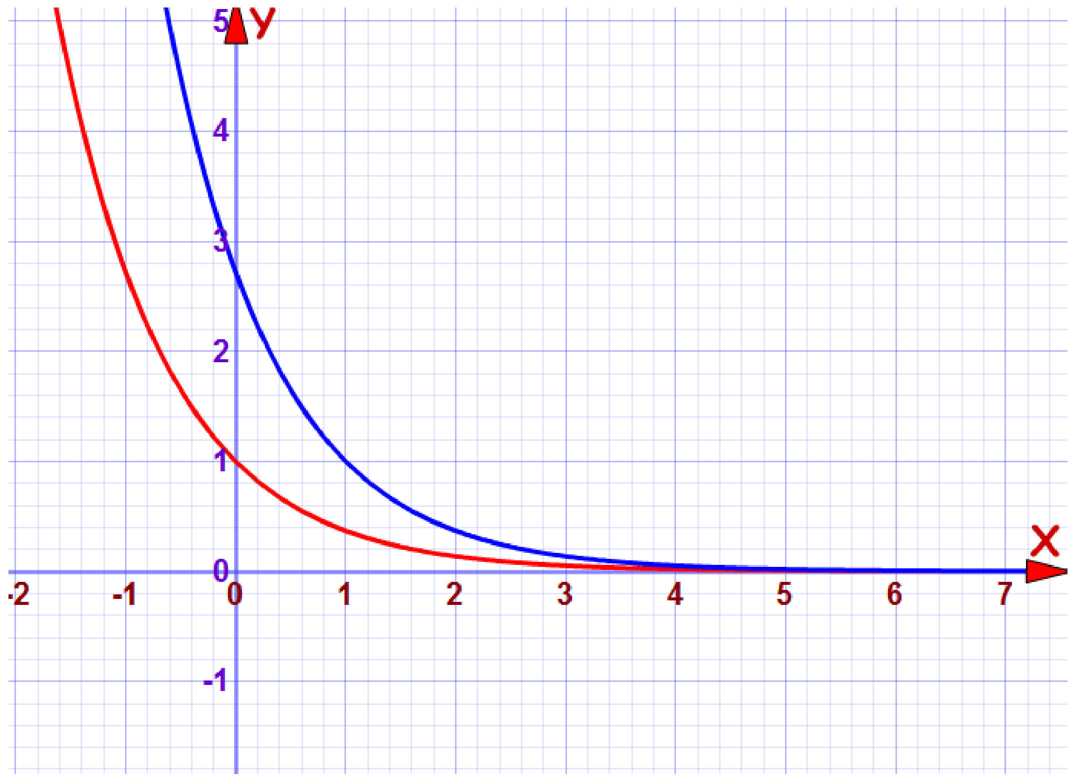

Figure 53: Two exponential functions with the same half-life. The red one begins its decay earlier than the blue one. Importantly, they never intersect.

This cannot, however, accommodate processes described by the French words example. It is important to note here, though, that it is not clear why the forgetting functions for different initial amount of words would have the same half-life. The following figure (Fig. 48) shows two exponential curves with different half-lives.

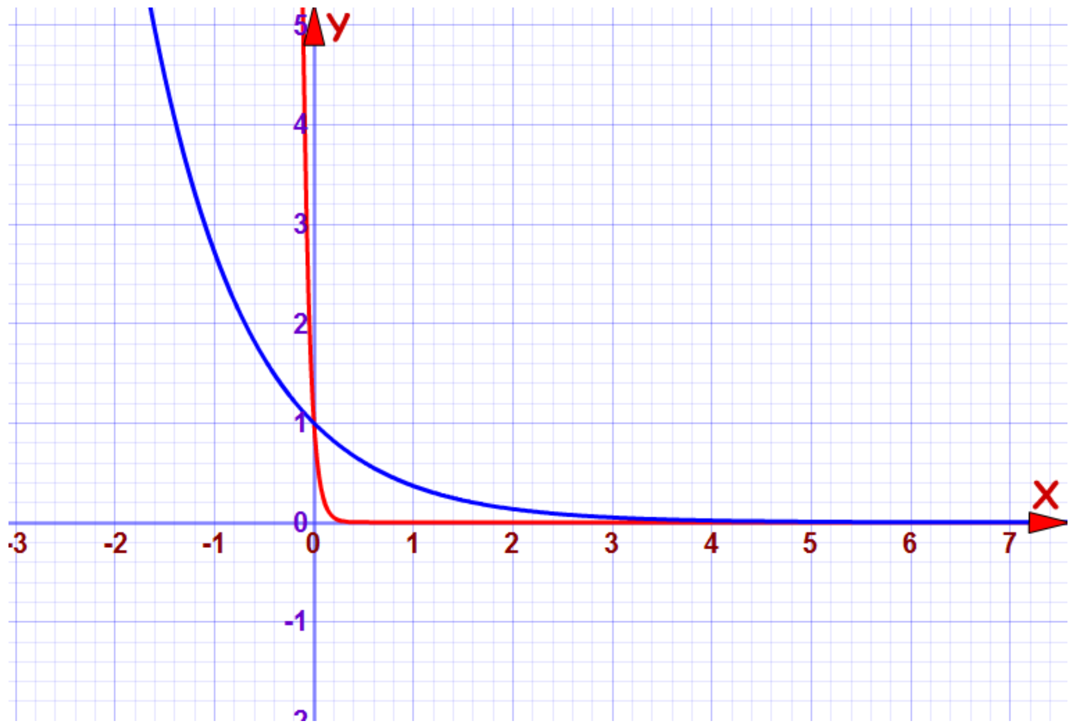

Figure 54: Two exponential functions with different half-life. The red one begins its decay after the blue one. Importantly, they intersect.

The intersection of lines in Figure48 means that the person that has learned the same amount of words at the point in time when another person has only that amount passed from previous learning, as in the French vocabulary example, may soon remember fewer of those words that the other person. This scenario, which incorporates manipulation of the Alpha parameter of the CDIE equation, is in accord The intersection of lines in Fig. 48 means that the person that has learned the same amount of words at the point in time when another person has only that amount left from previous learning, as in the French vocabulary example, may soon remember fewer of those words that the other person. This scenario, which incorporates manipulation of the Alpha parameter of the CDIE equation, is in accord with the main intuition behind the French vocabulary example. This seems to suggest that the exponential function in fact can be used for the fitting of forgetting data just as the power function can.

The second assumption made by those who argue against the exponential function is that the rate of change in the exponential function cannot be a function of time, that it is a constant. As noted above, this is mathematically not necessary. The rate of change in principle, as we have seen, may be another function dependent not only on time but also on, say, position in space. Figures illustrating this case are presented in the previous section. (Fig. 42)

The third assumption made by those who argue against the exponential function is that, unlike the exponential function, the power function has a property of self-similarity that may account for the similar shapes of forgetting curves across many tasks that are on different time scales. To account for similarities across time scales, however, one may simply suggest there are several interacting mechanisms that use the same general dynamics and differ only in the time scaling parameters with which they operate.

I now return to the main focus - the idea that there is a single forgetting curve across the entire time period of forgetting. Why would this be such an attractive idea despite the fact that so many results seem to conflict it? In addition to the value of fitting a line to the data to make predictions, it is tempting to think about a single curve being the result of a single feature of a forgetting model that produces it and it is theoretically interesting to take a look at this feature. If there was no implicit assumption about the single feature, the search for THE forgetting curve would have been probably abandoned a long time ago. Instead, only relatively recently this idea has become more popular among researchers, as mentioned.

Indeed, Sikstrom (1999), after reviewing a large amount of data and simulating forgetting in a modified Hopfield network, notes that the power function emerges from the dynamics of this network and therefore it is not an artifact of averaging of results in experimental paradigms on long-term memory with many subjects, as is sometimes suggested. In order to achieve this emergence, the Hopfield network was modified to include an exponential decay of weights (bounded weights) and the learning rate was not constant across the weights. The CDIE simulations are in line with these finding, with an additional suggestion that the learning rate of a connection perhaps does not have to be constant in time either. Sikstrom draws the attention to the fact that the review of empirical data does not clearly favor the power function as the best fit for the forgetting curves and that the value of his simulations is in showing that the shape of long term memory forgetting curve may emerge from the dynamics of Hopfield network. This is in line with the idea that perhaps there is not one specific shape of the forgetting curve for the entire duration of forgetting and that if the dynamics of the weights is even more variable within a network, this may result in complicated shapes of forgetting curves which in turn results in variability of suggested shapes in the literature.

Therefore, the following needs to be clearly stated: fitting a single function to the forgetting data has been and may be misleading in our quest to understand the mechanisms of memory and forgetting.

The CDIE simulations suggest that a clear distinction between terms spontaneous decay and decline needs to be made. The decline of memory, the forgetting in this case, is only partially resulting from spontaneous decay. The interference and other phenomena may also play a role.

The traditional forgetting curves describe this general decline but it seems that researchers have in mind something like spontaneous decay when they talk about the decline of a single trace.

In CDIE, the spontaneous decay is an exponential function, with a non-constant parameter alpha. However, the total shape of forgetting does not depend only on this term of the general equation and therefore is not necessarily an exponential function on any of its intervals.

8 Concluding remarks

Episode processing is a complex task and many models of some of its constituent tasks exist, although most of them are not dynamical process models. This is particularly true for the models of abstract hypothetical constructs such as a false memory, for example. The Complex Dynamical Integrated Episode processing hypothesis assumes combined ordered information from several sources – an existing memory of a person, new sources or sinks of information from the environment, both perceptual (such as a presented list of items) and more internal such as a form of relatively intuitive top – down attention. In addition, integrated with the rest of the information is the entire spatiotemporal context of the cognitive system involved in an episode processing. In the model it is assumed that order in time, an important information of episode processing, is treated by cognition as an order in a less abstract real space. A range of evidence from cognitive linguistics support this idea. The temporal change of this complex system is in this dissertation modeled by an integro – differential functional equation. Research on episode processing that is the theme of the CDIE hypothesis, is currently not a unified research area. Despite the continuity in cognitive processing, classical research areas are relatively arbitrary and separate categories, which are not always helpful in cognitive research. I have tried to use the most technical description, the most research-area-neutral name for the entire integrated set of processes: the "episode processing."

Novel contributions of this model, insights from simulations using it were listed and then illustrated, in order to make them as clear as possible. They include the following.

In the current literature there are (at least) two lines of research in some ways similar to the CDIE ideas presented here. These two research lines were compared and contrasted with the CDIE equations, simulations, and applications, to show significant novelties the CDIE hypothesis introduces. These lines of research all share the radiative transfer equation and have at least one additional similarity, besides differences from the CDIE. Grossberg's work from the 1960's models list learning at the same level as the CDIE model but simulates processes differently, using connectionist networks, while the Dynamical Field Theory of Kopecz, Engels, Schöner, Erlhagen, and their colleagues, has multiple differences, in level of analysis and thus model application, in a number of differential equations it uses and how they are combined, etc, while the kind of simulations it uses is the same.

Namely, on the one hand, in 1969, Stephen Grossberg published the paper "On the serial learning of lists." Here he uses very similar differential equation to model one aspect of processing of episodes - learning of a list of items. The work presented here has a broader scope, beyond learning, but it also has at least four other important novelties. The first is the non-constant time based decay. The decay term is in general (if present) in related models always assumed to be a constant. The second, arguably more important difference is that that in the present model the conceptual space - the space of items that might be learned in order - is a continuous dimension while it is a relatively small set of discrete items in Grossberg's work. Third, in the present model the order in time is treated as order in space (the idea that comes from research in cognitive linguistics) while in Grossberg's work only the real continuous time exists in the model and the order in time is dealt with via the Primacy Gradient, as we shall see later. Finally, the list learning is studied via connectionist networks in Grossberg's work but not here, which produces fundamentally different, but equally important insight in cognitive processes involved.

On the other hand, recently, a similar class of DS models has been used by Kopecz, Schöner, Erlhagen and their colleagues (Kopecz, Engels, and Schöner, 1993). Since the CDIE and DFT simulations are visually similar several fundamental distinctions between these models were specifically noted.

First, the level of cognition modeled buy the DFT is different than the level of the CDIE model. The DFT "is in a class of bi-stable neural networks first developed by Amari" (Johnson, Spencer, & Schöner, 2008). While the DFT uses at least two layers a layer of "excitatory neurons" and another one, the layer of "inhibitory interneurons," the CDIE does not deal with these concepts of neural or connectionist networks at all and it contains one "layer." In the CDIE model the terms "field" and "layer"are fundamentally mathematical concepts of an information integration "place". Therefore, these terms the CDIE are not a neural tissues or their analogs.

Second, a dimension that in the CDIE represents an abstract conceptual space, in DFT application represents the real space. Specifically, it may be "the retinal position of a point of light along a horizontal axis" or "the direction of a goal-directed movement." When the authors add new dimensions in their applications, they again simulate (two dimensional) real space as seen by subjects. The CDIE does not deal with this real perceived space at all. On the other hand, the DFT model does not use the time-as-space dimension and its dynamics are not present in any of its applications or theory.

Third, the details about the time-based decay term, which is a function of time and position along thew two dimensions in CDIE model, is a constant in DFT.

Fourth, the "resting level" term from the DFT equation is not contributing activation in each time step as is the case in the DFT but it is taken into account as a part of an initial state. Moreover, the details of how the context influences current information in CDIE differ from the DFT model and therefore in its applications. This is at the level of details of the integral terms of DEs used in these models as well as at the significantly more important level of how the dimensions in the two models are combined. The CDIE uses a sum of one-dimensional Gaussians while the DFT uses a two dimensional Gaussian. All of these substantially change the dynamics of these two systems. This has profound implications for modeling and theorizing about several cognitive phenomena attention and false memories (which are not explored in DFT modeling).

The model presented in this dissertation includes time based decay. The model mathematically defines time-based decay and using this definition, in a completely novel way it examines its the role in Serial Position Effects, Forgetting Curves, Item Distinctiveness and Confusability (in Serial Recall, Attention, False Memory, Word Length Effect, and Chunking in memory.) These analyses provide novel insights in complexity of decay's influences on cognitive phenomena. Related to decay but in the broader context of mathematical modeling, it was discussed how differential equations, along with currently more widely used methods of cognitive science, may be useful in producing strict mathematical definitions of theoretical concepts which in turn may help organize experiments and clarify their findings.

Decay seems to play a complex role in forgetting, may be interpreted as the part of the dynamics changed by attention, which also plays a role in false memory formation and control. The time based decay has a very important role in item distinctiveness or confusability, which in turn plays a complex and non intuitive role in the Word Length Effect, errors in recall, advantages of chunking in memory and so on. Using illustrations of time ordered inputs, it was suggested that some spontaneous decay may be needed for the cognitive mechanisms to function efficiently. IF the decay is too slow, the information processed may require or generate too much energy and no discriminability of items processed. If the decay is to fast, it was suggested, every stimulus is like the first one– they are the same and may be the basis for "perfect" memory, such as a "photographic" memory. Further, it was suggested that the the first stimulus in an episode arrives to an intact field. This means that it does not suffer interference from the previous stimuli but it also somewhat inhibits the entire field producing a possible disadvantage for the following ones. These phenomena are a partial basis for the Primacy Effect in immediate recall literature. On the other hand, the last stimulus as well is unique because it is not followed by others while at the same time it is the one with the most predecessors. This is probably part of the basis for the Recency Effect in Serial Position Curves. Notably, the even though the first and the last stimulus are both "special" they are not special in the same way which seems to be evident in literature. Advantages of chunking found in experiments seem to stem from a recovery of the field due to to breaks between chunks, which essentially "transforms" some of the mid-list items in long episodes into the "special" first and last ones in chunks. Based on these numerous insights and on survey of current models and literature, it was concluded that the time based decay may be a fruitful research area in the future despite some strong claims against this idea.

Lastly, a complex systems differential equations approach to time-as-space notion of cognitive linguistics presented here have not been done before in cognitive science and it might, perhaps, open an entirely new area of research in cognitive linguistics. Based on simulations of episode processing such as processing lists of words, it was explained why this was adopted to deal with order information that needs to be remembered in episodes, markedly with the notorious problem of recalling what was the first item or even in an episode. This was the first use of a DE model to this and related issues in cognitive linguistics. It worked well in situations where stimuli were highly variable in several ways among each other, which was not the case when existing ideas about, say, the Primacy Gradient, were used.

The presented dynamical process model is an innovatory application of an infinite – dimensional dynamical system to the ordered cognitive events in episode processing. In addition, numerous aspects of ordered events of episode processing, which are parts of many areas of cognitive science, are here treated as an integrated complex processes. The hypothesis presented in this dissertation unites in a new way perception, attention, memory, and more.

This dissertation discussed complex systems and made a call for wider use of differential equations, as one possible tool for analyzing dynamical systems, in cognitive science particularly at high levels of modeling cognitive phenomena that evolve in time. It was pointed out that mathematical and scientific uses of the notion of dynamical systems are not identical, that there are other tools for qualitative analyses of natural dynamical systems when DEs do not have simple enough analytical solutions. These kinds of analyzes are, however, relatively often seen in high level analyses of cognition. Finally, it was pointed out that same *mathematical* dynamical system equations may be used in relation to many different natural complex systems of nature in chemistry, physics, economics, biology, and so on and a particularly useful because they not only take into account measured variables but also the changes in these quantities. Cognitive systems are, arguably, complex dynamical systems and may be described by tools of mathematical dynamical systems theory at many levels – from neurons and neural networks to very abstract levels such as a conceptual level. It seems that physiological and biological level models readily use DEs but their use becomes less and less popular as the models approach high level cognition. This might be holding back cognitive science research, both "vertical" integration of different levels of analyses it employs (including its integration into the rest of science) and "horizontal" integration: Simulations, simply because they use a different perspective, may give insights on connections of abstract constructs of cognitive science that were previously not studied together. In this dissertation, using computer simulations, it was illustrated how this kinds of models might shed more light on cognitive processes by systematically changing details of the model, experimenting with them, which is not always possible in laboratory experiments.

Finally, several issues related to this kind of dynamical models were noted. The parameter fitting as a part of model development might be very hard. Notably, for the DCIE model presented it is especially hard because of the non-linear parameter Alpha. Nonetheless, simulations and behaviors of systems visible in them seem too offer a great deal of useful ideas about cognitive phenomena. Further, because of the abstract level of modeling, hypothetical constructs used have to be carefully thought of, and many possible measurement issues have to be kept in mind in interpretations. This is a trade-of situation where in turn relatively complex or abstract models are capable of accounting for a variety of data in many areas of cognitive science. In addition, it was argued that the variety of both static and dynamical models at different levels of application constrain and inform each other, and this wider perspective seems to be one of the driving forces of the development of science.

9 References

Amari, S. (1975). Homogeneous nets of neuron-like elements. *Biological Cybernetics, 17,* 211220.

Amari, S. (1977). Dynamics of pattern formation in lateral inhibition type neural fields. *Biological Cybernetics, 27,* 77 - 87.

Anderson, J. (1993). *Rules of the Mind.* Lawrence Erlbaum Associates, New Jersey: Hillsdale.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In: Spence, KW.; Spence, JT., (Eds). *The Psychology of Learning and Motivation: Advances in Research and Theory. 2,* (pp. 89-195). New York: Academic Press.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press, Clarendon Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory 8,* (pp. 47-90). New York: Academic.

Baddley, A. D., Thompson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, (57), 5-589.

Bechtel, W. (2012). Referring to localized cognitive operations in parts of dynamically active brains. In A. Raftopoulos and P. Machamer (Eds.), *Perception, Realism and the Problem of Reference.* Cambridge: Cambridge University Press.

Beurle, R. L. (1956). Properties of a mass of cells capable of regenerating pulses. *Philosophical Transactions of the Royal Society London B, 240,* 5594.

Bjork, E. J. & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior, 13,* 8097.

Bromberg-Martin, E. S., Matsumoto. M., Nakahara, H., & Hikosaka. O. (2010). Multiple timescales of memory in lateral habenula and dopamine neurons. *Neuron, 67,* (3), 499-510.

Brown, G. D. A., & Lewandowsky, S. (2010). Forgetting in memory models: Arguments against trace decay and consolidation failure. In S. Della Sala (Ed.), *Forgetting* (pp. 49-75). Psychology Press.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A ratio model of scale-invariant memory and identification. *Psychological Review, 114,* 539-576.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107,* 127-181.

Brown, J., (1958) . Some Tests of the Decay Theory of Immediate Memory. *Quarterly Journal of Experimental Psychology, 10,* 12-21.

Busemeyer, J. R., Pothos, E. M., Franco, R., & Trueblood, J. S. (2011). A Quantum Theoretical Explanation for Probability Judgment Errors. *Psychological Review, 118,* 193-218.

Čadež, E., Čadež, V. M., & Heit, E., (2010, August) Mathematical modeling of memory evolution pure decay. Poster presented at the 43rd Annual Meeting of the Society for Mathematical Psychology. Portland, OR.

Čadež, E., & Heit, E. (2011, July) Forgetting Curves Emerge from Dynamics of Integrated Memory. Paper presented at the 33rd Annual Conference of the Cognitive Science Society. Boston, MA.

Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., & Rosen, D.B., (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks, 3,* 698-713.

Casasanto, D., & Boroditsky, L. (2008). Time in the Mind: Using space to think about time. *Cognition, 106,* 579-593.

Casasanto, D., Fotakopoulou, O., & Boroditsky, L. (2010). Space and time in the childs mind: Evidence for a cross-dimensional asymmetry. *Cognitive Science, 34,* 387-405.

Chater, N., & Vitanyi, P. (2003). Simplicity: a unifying principle in cognitive science. *Trends in Cognitive Science, 7,* 1922.

Clark, H. H. (1973). Space, time, semantics and the child. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language.* New York: Academic Press.

Coombes, S., (2005). Waves and bumps in neural field theories. *Biological Cybernetics, 93*, 91-108.

Kopecz, K., Engels, C., & Schöner, G. (1993). Dynamic field approach to target selection in gaze control. In S Gielen and B Kappen, editors, International Conference on Artificial Neural Networks, Amsterdam, pages 96-101. Springer Verlag, Berlin.

Cowan N. (1992). Verbal memory span and the timing of spoken recall. J*ournal of Memory and Language, 31,* 668-684.

Cowan, N. (1995). *Attention and memory: An integrated framework.* New York: Oxford University Press.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 1722.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108,* (2), 452-477.

Erlhagen, W., Bastian, A., Jancke, D., Riehle, A., & Schöner, G., (1999). The distribution of neuronal population activation as a tool to study interaction and integration in cortical representations. *Journal of Neuroscience Methods 94,* 53-6.

Erlhagen W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review, 109, 3,* 545-572.

Eysenck M. W., & Eysenck M. C. (1980). Effects of processing depth, distinctiveness, and word frequency on retention. *British Journal of Psychology. 71,* 263274. doi: 10.1111/j.2044-8295.1980.tb01737.x.

Farlow, J., Hall, J. E., McDill, M. J., & West, B. H. (2002). *Differential Equations and Linear Algebra.* New Jersay: Prentice Hall.

Griffith, J. S., (1963). A field theory of neural nets: I: Derivation of field equations. *Bulletin of Mathematical Biophysics, 25* 111120.

Grossberg, S. (1964). The theory of embedding fields with applications to psychology and neurophysiology. Rockefeller Institute for Medical Research.

Grossberg, S. (1969). On the serial learning of lists. *Mathematical Biosciences, 4,* 201-253.

Grossberg, S. and Pepe, J. (1971). Spiking threshold and overarousal effects in serial learning. *Journal of Statistical Physics, 3,* 95-125.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors & II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics, 23,* 121-134 & 187-202.

Grossberg, S. (2013). Adaptive resonance theory. *Scholarpedia.*

Hebb, D. O. (1949). *Organization of Behavior.* New York: Wiley.

Heit, E., Brockdorff, N., & Lamberts, K. (2004). Strategic processes in false recognition memory. *Psychonomic Bulletin & Review,* 11, 380-386.

Henson, R. N. A. (1998). Short-term memory for serial order: the start-end model. *Cognitive Psychology, 36,* (2), 73-137.

Henson, R. N. A., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing signal hypothesis. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56(A),* 13071334.

Hitch, G., Fastame, M., & Flude, B. (2005). How is the serial order of a verbal sequence coded? Some comparisons between models. *Memory, 13.*

Jalbert, A., Neath, I., & Surprenant, A. M. (2011). Does length or neighborhood size cause the word length effect? Memory & Cognition, 39, 1198-1210.

Johnson, M. K., Reeder, J. A., Raye, C.L ., & Mitchell, K. J. (2002). Second thoughts versus second looks: Age-related deficit in reflectively refreshing just-activated information. *Psychological Science, 13,* 64-67.

Johnson, J. S., Spencer, J. P., & Schöner, G. (2008). Moving to higher ground: The dynamic field theory and the dynamics of visual cognition. In F. Garzn, A. Laakso, T. Gomila (Eds.) Dynamics and Psychology [special issue]. *New Ideas in Psychology, 26,* 227-251.

Kamp, S., Forester, G.R., Murphy, A.R., Brumback, T., & Donchin, E. (2012). Testing a distinctiveness explanation of the primacy effect in free recall using event-related potentials. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), Proceedings of the 34th Annual Conference of the Cognitive Science Society (pp.539-544). Austin, TX: Cognitive Science Society.

Keppel G, & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1,* 153-161.

Kim, A. V. (1999). Functional Differential Equations. *Mathematics and Its Applications, 479,* pp 41-59 .

Kondo, S., & Arai, R. (2005). A reaction-diffusion wave on the skin of the marine angelfish Pomacanthus. *Nature, 376,* 765 768.

Kopecz, K., & Schöner, G. (1995). Saccadic motor planning by integrating visual information and preinformation on neural, dynamic fields. *Biological Cybernetics, 73,* 49-60.

Ladd, G. T., & Woodworth, R. S. (1911). *Elements of physiological psychology: a treatise of the activities and nature of the mind, from the physical and experimental points of view.* [New ed.] New York: C. Scribner's Sons.

Larsen, J.D., Baddeley, A., & Andrade, J. (2000). Phonological similarity and the irrelevant speech effect: Implications for models of short-term memory. *Memory, 8,* 145157.

Lee, C. L., & Estes, W. K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 149-169.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review, 96,* 25-57.

Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin Review, 11,* 771-790.

Matlock, T., Ramscar, M., & Boroditsky, L. (2005). *Cognitive Science, 29,* 655-664.

Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory & Cognition, 21,* 689-698.

Newton, Sir Isaac (1968). *The Mathematical Principles of Natural Philosophy.* (Motte, A., Trans.). London, UK: William Dawson and Sons. (Original work published 1729).

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761-781.

Pavlov, I. P. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex.* Translated and Edited by G. V. Anrep. London: Oxford University Press.

Piaget, J. (1927/1969). *The childs conception of time.* New York: Ballantine Books.

Poznanski, R.R. & Cacha, L. (2011). Associable Representations as Field of Influence. *Journal of Integrative Neuroscience 10,* 423-437.

Raye, C. L., Johnson, M. K., Mitchell, K. J., Reeder, J. A., & Greene, E. J. (2002). Neuroimaging a single thought: Dorsolateral PFC activity associated with refreshing just-activated information. *NeuroImage, 15,* 447-453.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7,* 1-75.

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention, *Neuron, 61,* 168-185.

Riley, M. A., & Holden, J. G. (2012). Dynamics of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science.* doi: 10.1002/wcs.1200.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803 814.

Rumelhart, D. E., Smolensky, P., McClelland, J. L. & Hinton, G. E. (1986). Schemata and sequential thought processes in parallel distributed processing. J. L. McClelland, D. E. Rumelhart & the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models.* Cambridge, MA: MIT Press/Bradford Books. 757.

Schöner, G., Kopecz, K., & Erlhagen, W. (1997). The dynamic neural field theory of motor programming: Arm and eye movements. In P. G. Morasso and V Sanguineti, editors, *Self-Organization, Computational Maps and Motor Control, Psychology Series, 119,* 271-310. Elsevier-North Holland.

Shankar, K. H., & Howard, M. W. (2010). Timing using temporal context, *Brain Research.*

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317-1323.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart, J. L. McClelland & the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations.* Cambridge, MA: MIT Press/Bradford Books. 194281.

Smolensky, P. (1988). On the proper treatment of connectionism. *The Behavioral and Brain Sciences, 11,* 123.

Spencer, J. P., Barich, K., Goldberg, J., & Perone, S. (2012). Behavioral dynamics and neural grounding of a dynamic field theory of multi-object tracking. *Journal of Integrative Neuroscience, 11,* 339-362.

Spivey, M. J. (2007). *The continuity of mind.* New York: Oxford University Press.

Sun, R., & Bookman, L. (eds.). (1994). *Computational Architectures Integrating Neural and Symbolic Processes.* Dordrecht, the Netherlands: Kluwer Academic Publishers.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*, 49-100.

Tenenbaum, J. B., De Silva, V. & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science, 290,* (5500), 2319-2323.

Thelen, E., Schöner, G., Scheier, C., & Smith, L. (2000). The dynamics of embodiment: A field theory of infant perseverative reaching. *Brain and Behavioral Sciences, 24,* 133.

Van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review, 98,* 3-53.

Van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy, 91,* pp. 345-381.

Thorndike, E. L. (1898). Animal Intelligence: An Experimental Study of the Associative Processes in Animals. *Psychological Review, Monograph Supplements,* (8). New York: Macmillan.

Trope, Y., & Liberman, N. (2010). Construal level theory of psychological distance. *Psychological Review, 117,* 440463.

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology, 23,* 515-557.

Ulanovsky, N., Las, L., Farkas, D,. & Nelken, I. (2004). Multiple time scales of adaptation in auditory cortex neurons, *Journal of Neuroscience, 24,* 10440-10453.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Science, 7,* 483-488.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature, 393,* 440-442.

Watson, J. B. (1919). *Psychology from the standpoint of a behaviorist.* Philadelphia: Lippincott.

Waugh, N. C. (1961). Free versus serial recall. *Journal of Experimental Psychology, 62,* 496-502.

Weisstein, E. W. (2013). Dynamical System. From *MathWorld–A Wolfram Web Resource.* Retrieved July 10, 2013 from the World Wide Web: http://mathworld.wolfram.com/Dyn

Whitesides, G. M., & Ismagilov, R. F. (1999). Complexity in chemistry. *Science, 284,* 89-92.

Wilson, H. R. & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysics Journal, 12,* 1-24.

Wilson, H. R. & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik, 13,* 5580.

Yang, X. S. (2002). Fractals in small-world networks with time delay. *Chaos, Solitons & Fractals, 13,* (2), 215-219.