A Connectionist Extension to Kintsch's Construction-Integration Model

#Vicente Sanjosé, #Eduardo Vidal-Abarca and *Olga M. Padilla

#University of Valencia (Spain)

*University of Zacatecas (Mexico)

Date: December 31, 2005

Mail Address

Vicente Sanjosé: Science Education, Escuela de Magisterio "Ausias March", Universitat de Valencia; PO Box 22045; 46071- Valencia - Spain

E-mail: vicente.sanjose@uv.

Abstract

This paper proposes an extension to Kintsch's Construction-Integration mode of text comprehension, which changes its mathematical implementation and emphasizes the connectionist features of the model. Specifically, the extension proposed here a) simulates the learning process in a connectionist manner by making explicit changes in the connecting values among processing units; b) takes into account individual differences in prior background knowledge activation and individual ability to make logical inferences, using these parameters to adjust the results of the simulation; and c) implements an algorithm that constructs the connectivity matrix W from text processing. The proposed extension is tested on existing recall and contradiction-detection data from readers of science texts, and its predictions fit the empirical data better than the CI model's prediction

The Construction-Integration (CI) model (Kintsch, 1988, 1998; Kintsch & Welsch, 1991) is one of the functional models most widely used to simulate text comprehension (Britton & Graesser, 1996; van Oostendorp & Goldman, 1999). The model is based on a proposal by Kintsch and van Dijk (1978), later revised and extended by van Dijk and Kintsch (1983), to include two different levels of representation which the reader constructs when reading a text: the reader's prior background knowledge activation while reading the text and the distinction between the textbase and the situation model. As Kintsch (1988) has pointed out, Van Dijk and Kintsch's revised formulation (1983), which was influenced by the Schema Theory prevailing during the 70s and 80s, has at least two problems. The first is the existence of psychological data which question the idea that the direction of comprehension is top-down. The second problem is that human comprehension is highly flexible and sensitive to context; therefore, it is difficult to model it with rigid structures such as schemata. These weaknesses led Kintsch (1988, 1998) to formulate the Construction-Integration Model (CI).

The CI model incorporates the reader's prior knowledge into the simulation of comprehension processes with a bottom-up, connectionist approach in the integration phase. The model has been widely accepted because it is easy to implement, because it can model a vast array of comprehension phenomena, and because it is psychologically grounded. Moreover, its efficient simplicity allows a variety of developments and extensions. Our purpose in this paper is to show that the model can be developed while maintaining its basic features but extending the most relevant characteristic of connectionist models: their ability to model explicitly how people learn from the environment. Our extension moves the CI model closer to other connectionist models which simulate learning processes, and it models individual differences in these processes.

First, then, we will summarize the characteristics of the CI model by illustrating the construction and integration phases with an example. Second, we will use the same example to explain our extension in detail; i.e., its components, parameters, and implementation, so that the similarities and differences between our extension and the CI model will be apparent. Finally, our extension will be used to simulate two phenomena: a) students' free recall after studying a long scientific passage, and b) failure to detect contradictions in a text, as previously simulated with the CI model (Otero & Kintsch, 1992)

The Construction-Integration Model

According to the CI model (Kintsch, 1988, 1998), comprehension is a bottom-up process highly influenced by context. Comprehension involves an initial phase of construction, which is chaotic, but which attains coherence in the integration phase. During the construction process, the reader constructs a network of propositions both from the text and from their prior background knowledge. Construction processes are regulated by weak production rules in a bottom-up fashion; this results in disorderly, redundant, and even contradictory output. Construction is followed by an integration process in which the nodes of the network spread their activation until the network stabilizes in a way that takes account of the pattern of mutual constraints among the nodes. As a result, some irrelevant propositions formed in the construction phase are deactivated, whereas others receive higher activation according to the constraints that the text itself and the reader's prior background knowledge impose on them. The final activation values of the nodes reflect the constraining properties of the network as a whole.

In the CI model, the processing of the text is performed in cycles. In each cycle, a fragment of the text, usually a phrase, is processed. When the reader has processed the information from the first cycle and proceeds to the next one, a small number of propositions, generally one or two, are retained in working memory in order to be processed together with the material from the next cycle. These retained propositions help to connect the previously read text to the new input, and to assess the coherence of the information processed in the previous cycle with the information in the new cycle. Memory is understood as an associative network of nodes which influence one another, each node being activated or inhibited with respect to the other nodes of the network. What follows is a detailed account of how these two processing phases work in the CI model.

Construction Processes

In every cycle, the reader constructs the text propositions, as well as relevant and irrelevant prior background knowledge propositions, by incorporating all this information as nodes in a network. The mechanism to construct the propositions is not specified in the model; the model only specifies how the propositions are connected and integrated in a network. This network is represented mathematically in the model as a symmetric matrix W, called the connectivity matrix, whose elements represent the strength of the connection between propositions; positive values are used to link propositions which are mutually reinforced, and negative values are used for contradictory or mutually exclusive propositions. The second mathematical element of the CI model is vector A, which represents the activation level of each proposition in the network. Usually it is assumed that in the construction phase the propositions of the text are fully activated, while prior background knowledge is deactivated or activated to a very low degree.

It should be noted that the networks built on the CI model involve local representations, as Read and Miller (1998) affirm. In a local representation, a concept, or an entire proposition, is represented by a single node, which is different from a distributed representation, where a concept is represented by a pattern of activation over a number of nodes. Local models are more suited to simulating high-level cognitive phenomena such as comprehension. In comprehension, one is interested in the relationships among concepts or propositions, and the parallel satisfaction of mutual constraints among concepts or propositions is the key mechanism for attaining coherence in the network (Read & Miller, 1998). In other words, constraint-satisfaction is the main mechanism for the integration of propositions in order to reach coherence in the CI model (Kintsch, 1998). An additional advantage of local representations is that they can be better interpreted than distributed representations, as each concept or proposition corresponds to a single node.

To exemplify how the CI model works, consider a short fragment of a text used in our experiments:

(1) Heat is energy transferred from one body to another because of a difference in temperature. The unit of heat is the calorie.

According to Bovair and Kieras' guidelines (1985), the reader should extract the following textual propositions:

1.-T1: IS-A (heat, T2)

2.-T2: IN (energy, transference)

3.-T3: BETWEEN (transference, body1, body2)

4.-T4: OF (difference, temperature)

5.-T5: BECAUSE (T1, T4)

6.-T6: OF (T7, heat)

7.-T7: REF (unit, calorie)

Here, we will assume that a reader also activates the ideas that heat is a physiological sensation, and that energy is not a physiological sensation from their background knowledge. The corresponding propositions to both ideas are the following:

8.-K1: IS-A (heat, sensation)

9.-K2: NEG-IS-A (energy, sensation)

Provided that the reader had good reasoning strategies, from the textual idea that heat is energy and from their prior background knowledge that energy is not a sensation, the logical inference that heat is not a sensation could be generated. This would result in the following proposition:

10.-INF1: NEG-IS-A (heat, sensation)

It should be noted that this proposition contradicts the reader's prior background knowledge proposition, K1.

There are no unique rules to assign values to the elements of matrix W, which are left as free parameters to adjust to the empirical results. Among other possibilities, the CI model can assign the following values: (a) value 1 to the connections between the propositions which share arguments; e.g., the connection of each node to itself and the connection between T1 and T6, because they share the argument <u>heat</u>; (b) value -1 to the connections between contradictory or incompatible propositions; e.g., K1 and INF1; and (c) value 0 to the remaining connections. Accordingly, matrix W could be formed like this (with propositions T1, T2, T3, T4, T5, T6, T7, K1, K2, INF1 in this order):

	T1	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**K1**	**K2**	**INF1**	
	1	1	1	0	1	1	0	1	1	1	**T1**
	1	1	1	0	1	0	0	0	1	0	**T2**
	1	1	1	0	1	0	0	0	0	0	**T3**
	0	0	0	1	1	0	0	0	0	0	**T4**
W =	1	1	1	1	1	1	0	1	1	1	**T5**
	1	0	0	0	1	1	1	1	0	1	**T6**
	0	0	0	0	0	1	1	0	0	0	**T7**
	1	0	0	0	1	1	0	1	1	-1	**K1**
	1	1	0	0	1	0	0	1	1	1	**K2**
	1	0	0	0	1	1	0	-1	1	1	**INF1**

The initial activation vector A is defined with value 1 for the components associated to the textual propositions of the sentence being read, and value 0 for the rest.

A (0) = (1, 1, 1, 1, 1, 1, 1, 0, 0, 0)

Here the bracket ‘(0)’ stands for "initial activation" or iteration-0.

<u>Integration processes</u>

The integration phase is modeled through a spreading connectionist activation process which takes place in the whole network until it becomes stable. At the end of the phase, each proposition gains a level of activation which is generally different from the activation it had at the beginning of the process. High activation values indicate that the proposition is relevant for the current comprehension of the text, whereas low values indicate that it is barely relevant or even irrelevant.

The spreading of activation through the whole network is simulated by repeatedly multiplying matrix W by vector A, in order to obtain another activation vector A'. This new vector should be normalized so that the values are kept within a psychologically significant range. For the first iteration:

$A'(1) = W \cdot A(0);$ (1)

$A'(1) = (a'_{11}, a'_{21}, a'_{31}, ...)$ (2)

Now, normalization is achieved by setting the negative components to 0, and by dividing the positive ones by the highest component in A'(1):

$a_{j1} = a'_{j1} / \max(a'_{j1})$ if $a'_{j1} > 0$ (3a)

$a'_{j1} = 0$ if $a'_{j1} < 0$ (3b)

After the first iteration, a new vector A (1) is obtained:

$A(1) = (a_{11}, a_{21}, a_{31}, ...)$ (3c).

This process, multiplication and normalization, is repeated with vector A(1) in a new iteration, and so on. The process finishes when the value of the resulting activation vector does not change significantly after successive multiplications by W. This stabilization process corresponds psychologically to the achievement of a coherent representation of the text in the reader's mind.

In our case, the final vector for the activations is:

A (final) = (0.97 , 0.66 , 0.53 , 0.20 , 1.00 , 0.66 , 0.13 , 0.57 , 0.76 , 0.57)

According to this result, the fifth proposition T5 (heat is the transference of energy because of a difference in temperature) is the most highly activated. As can be seen in the matrix, it is the most interconnected proposition. The second most highly activated proposition is T1 (heat is energy in transference). In psychological terms, these two propositions are the most highly activated in the reader's working memory. According to the CI model, the first proposition, or maybe both, would be retained in working memory to be processed together with the propositions for the next cycle.

It should be noted that the two contradictory propositions, K1 heat is a sensation, and INF1 heat is not a sensation, gain identical final activation. In psychological terms, this means that at this stage of processing neither of them is discarded as inadequate or false. In subsequent cycles, both propositions might be reactivated from memory and one of them might receive more activation than the other.

In the CI model, the episodic memory of a text is composed of the accumulated results of each cycle. For each cycle, there is a connectivity matrix W, but the result of comprehension and storing in long-term memory in the CI model is a new matrix M, whose elements m_{ij} represent the strength of connection between propositions i and j stored:

$$m_{ij} = \Sigma_k \, p_{ij} \, a_i \, a_j \qquad (4)$$

where p_{ij} is an element of matrix W, a_i is the final activation of node i, and k is the number of cycles in which nodes i and j have been processed together. The equation includes the sum of all the cycles in which both nodes were activated, following the Hebbian principle at the same time: the more frequently these nodes are activated simultaneously, the higher the value of m_{ij} and the more intense the connection between propositions i and j in the memory of the reader.

At the moment shown in our example – only one processing cycle – the calculation for M would yield the following values:

	T1	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**K1**	**K2**	**INF1**	
	1.93	1.64	1.51	0	1.97	1.64	0	0.55	0.74	0.55	**T1**
	1.64	1.44	1.35	0	1.66	0	0	0	0.51	0	**T2**
	1.51	1.35	1.28	0	1.53	0	0	0	0	0	**T3**
	0	0	0	1.04	1.20	0	0	0	0	0	**T4**
M=	1.97	1.66	1.53	1.20	2.00	1.66	0	0.57	0.76	0.57	**T5**
	1.64	0	0	0	1.66	1.43	1.09	0.38	0	0.38	**T6**
	0	0	0	0	0	1.09	1.02	0	0	0	**T7**
	0.55	0	0	0	0.57	0.38	0	0.33	0.44	-0.33	**K1**
	0.74	0.51	0	0	0.76	0	0	0.44	0.58	0.44	**K2**
	0.55	0	0	0	0.57	0.38	0	-0.33	0.44	0.33	**INF1**

The values in the main diagonal of M are related to the probability of recall of each proposition. The absolute values are not relevant, but they are interpreted in relation to one another. According to this, proposition T5 should be best recalled, followed by T1 and T2[1].

Notes on the CI model

The CI model is simple, easy to use, and is based on a sound theory of comprehension (Kintsch, 1998), but its connectionist features and its mathematical implementation can be extended. The extension we propose is aimed at improving three specific issues: (a) to make explicit the change in the strength of the connections between pairs of propositions in the connectivity matrix, which is the core of learning in the connectionist models; (b) to relate this change to the process of reading, and (c) to improve the power of the model to simulate some individual differences. We will explain these three issues briefly.

First, it should be noted that, despite its connectionist nature, the CI model does not explicitly simulate the process of learning from a text. In connectionist models, learning is a dynamic process represented by the change in the weight or strength of the connections between nodes, i.e., the propositions change simultaneously with the activation of the nodes, and this simulates the learning process. Changes in the weights have an effect on unit activation and vice versa. The process stops when the system becomes stable. However, in the CI model, matrix W is static, that is, its values do not change during the processing cycle. What the reader learns is simulated in a matrix different from W, matrix M, which does not take part in the processing of each cycle. Thus, the simulation of learning in the CI model deviates from the usual procedure for connectionist models.

Second, in the CI model, the values for matrix W are the free parameters of the model (Singer & Kintsch, 2001). It is the researcher who decides upon those values following various criteria, such as argument overlap or causality, among others. Usually, after a propositional analysis of the text, those values are chosen so that the final results of the simulation fit the empirical data.

Therefore, the values of W do not derive from the process of reading, but from the analysis of the text itself.

The third issue has to do with how CI deals with individual differences. Generally speaking, individual differences have been modeled in the CI model by modifying matrix W to accommodate the results of the simulation to individual differences (see Otero & Kintsch 1992; Kintsch et al. 1990). However, due to its mathematical implementation, some problems related to the simulation of individual differences arise in the CI model. In an interesting and rigorous study, Guha and Rossi (2001) carried out a mathematical analysis of the CI model. They proved that, when comprehension and learning is simulated with the CI model, matrix W often results in a single point of equilibrium for the space of activation vectors, which means that there is a single solution for the final activation after integration. Therefore, it is not possible to simulate individual differences in those cases. Furthermore, if none of the values of W are negative, defining W means predetermining the result of the final activation vector, independently of the initial activation vector selected (Rodenhausen 1992, Kintsch 1998, page 99). In this case, the stabilized result of the integration will be the autovector that corresponds to the highest autovalue in W (Padilla, 2000; Guha and Rossi, 2001, Proposition 2, p. 363). Therefore, if there is a lack of inhibitory connections, the researcher's decision to define W determines the result of the final activation and the stored value M. In fact, the only way to get more than one final activation vector for the same matrix W is that W degenerates for the dominant eigenvector. It is not easy to find the family of W matrices able to fulfill this condition for a specific text. Guha and Rossi propose a backwards mechanism to construct W from the different final activation vectors compatible with states which can be psychologically interpreted. In most cases, this mechanism is difficult to implement and the result is difficult to interpret in psychological terms.

Extending the connectionist features of the CI model

We have implemented some modifications that extend the connectionist characteristics of the CI model to deal with the three issues just mentioned. Our proposal, the CECI model – or Connectionist Extension of the CI model – takes the basic structural features of the CI model: the processing in cycles due to the limitations of working memory, the propositions as processing units and the two phases of processing, construction and integration. The main innovation of the CECI model is that the learning process, that is, the dynamics of change in the connection weights between propositions during text processing, is made explicit. Thus, learning consists in modifying the internal connections so that the internal input to each proposition will be the same as the external input from the environment, that is, the text. Ideally, after processing and connections have been stored in memory, when some pieces of information from the text (e.g., some important textual words) feed the net as input, the system should be activated in the same way as it was with the text. In other words, the ideal reader will be able to remember the text after reading it in the same way as when it was read. We will explain in detail how the CECI model proceeds and for the simulation we will use the fragment of text about heat which was used to exemplify the CI model. This way, the differences between CI and CECI will be apparent.

CECI incorporates a mathematical implementation used in other connectionist models, that is, the delta rule of learning (McClelland & Rumelhart, 1985). Another innovation is that CECI incorporates new parameters that may account for individual differences in text processing in three aspects: (a) activation of prior background knowledge, (b) ability to perform logical inferences and (c) time of text processing. We will explain all these innovations in the following sections.

Construction Phase

As with the CI model, the mechanism that specifies the construction of propositions is outside the CECI model. Once the propositions are constructed, they will be activated and interconnected. However, unlike the CI model, CECI incorporates a general mechanism to specify the way in which each constructed proposition will be activated by textual input and will be interconnected with others.

For CECI, each proposition receives input from the text being read in a cycle. In every cycle, text input is represented by vector t, whose components are the number of times each word processed in the cycle is repeated in the text processed in that cycle (see Table 1, vector t for the text about heat). In order to avoid undesirable effects due to differences in length of the text processed in a cycle, we will divide vector t by the total number of words in that cycle.

Each proposition i has another specific vector associated with it, called w_i, in which each component $w_{i\alpha}$ is the external weight of the connection between word α and proposition i. This weight takes as initial constant value 0, when the α-th word is not present as an argument of proposition i, and a positive value $w_{i\alpha}$, when the α-th word is present as an argument of proposition i. In the matrix that represents the processing of the sentence about heat (see Table 1), we assigned 1 as initial value when the α-th word is present as an argument of the proposition; otherwise, the value is 0. The external weights regulate the intensity of the external connection between each proposition and the environment, i.e., the text. In every cycle, the value of the external weights can vary from $w_{i,\alpha}(0)$ to a maximun of $w_{i,\alpha,max} = 1$.This depends on the processing; higher values indicate stronger connections between the text and the constructed propositions, which means deeper processing in psychological terms.

The value of the external input to each proposition i is mathematically constructed by multiplying the components of vectors t and $w_{i,\alpha}$ (scale product):

$$e_i = \Sigma_\alpha t_\alpha . w_{i\alpha} \tag{5}$$

PLEASE, INSERT TABLE 1 ABOUT HERE

A number of consequences of our procedure of activating propositions from external input should be noted. First, the value of external input e varies from 0 to 1. If a proposition included all the words of the textual fragment read, the scale product (5) would take maximum value 1. Second, simple propositions that contain few textual words, (e.g., T4: OF (difference, temperature)) reach lower activation than complex propositions whose arguments are other propositions (e.g., T5: BECAUSE (T1, T4)). However, complex propositions are less likely to be constructed by a reader than simple propositions. Third, textual propositions reach higher activation than propositions coming from the reader's prior background knowledge when the arguments of the prior background knowledge propositions are not explicit in the text. Finally, though the CI model has simulated predictive and bridging inference generation phenomena (Schmalhofer, McDaniel & Keefe, 2002), in the CECI model we only consider simple logical inference propositions constructed during the processing of the text input. These propositions are activated by a combination of some antecedent propositions through a procedure that will be explained later on.

In order to exemplify how our model works, and to compare it with the CI model, we shall take the same fragment of the text we analyzed before:

(1) Heat is energy transferred from one body to another because of a difference in temperature. The unit of heat is the calorie.

It can be shown that the maximum value of e for propositions T1, T5 and INF1 is 4/12 = 0.333, 7/12 = 0.583, and 2/12 = 0.167, respectively.

Apart from receiving external input e from the text, each proposition may also receive excitatory or inhibitory signals from other propositions. Those connections between pairs of propositions are internal connections. There are two types of internal connections; namely, binary connections and Y-connections. Binary connections can be established between: (a) two T-propositions, (b) two K-propositions, and (c) a T-proposition and a K-proposition. Y-connections are inference pi-sigma connections (Rumelhart & McClelland, 1986) established among three propositions; one of these is an INF-proposition, while the other two are one of the three possible combinations mentioned before. In our example, INF1: NEG-IS-A (heat, sensation) is connected to textual proposition T1: IS-A (heat, energy-transferred) and the proposition from prior background knowledge K2: NEG-IS-A (energy, sensation). It should be noted that, unlike the CI model, self-connections are not allowed in our model.

Connections have an internal weight associated with them. The internal weight for binary connections can be either positive, which corresponds to mutual excitations, or negative, i.e., a weight between propositions which are explicitly and mutually contradictory. Therefore, each proposition may receive an excitatory or inhibitory signal from the neighboring propositions in a cycle. The initial value of the internal weight between two propositions in every cycle is 0, except for propositions from the previous cycle remaining in the reader's working memory. It is also possible that two K-propositions have an initial weight higher than 0, when they are already connected in the reader's mind. As a consequence, it is only external input e that feeds the processing system. At the beginning of each cycle, all activations are null.

Internal binary weights are represented by p_{ij}, where i and j are the number of the interconnected propositions. It should be noted that p_{ij}, as in the CI model, can be negative when the two propositions are contradictory. For the binary connections, the internal input to unit i, in_i, is the sum of all the individual weighted signals sent to this unit:

$$in_i = \Sigma_{i\,j}\, a_j\, p_{ij} \qquad (6a)$$

where j refers to all the units in the cycle, except i; a_j is the current activation of unit j, and p_{ij} is the current value of the connection weight from unit j to i.

Inference weights corresponding to Y-connections are represented by p_{ijk}, where i represents the inference, and j and k are the respective numbers of the interconnected antecedent propositions (either T- or K-propositions). Thus, this weight involves three propositions. For an inference concerning a Y-connection, the internal input to a i unit is:

$$in_i = 2.(a_j \,.\, a_k)^{1/2}.p_{ijk} \qquad (6b)$$

where j and k refer to the two antecedent propositions from which proposition i is inferred. The square root of the product of both activations is coherent with the dimensionality of the equation (6a) and with logic: only when both antecedent units are active can inference be activated.

As can be seen, internal input depends on both the activation of neighboring propositions and on the strength of the connections. It should be noted that in our model, internal weights are not symmetric (i.e., p_{ij} may be different from p_{ji}). This asymmetry is also a characteristic of van den Broek and his colleagues' Landscape Model of Reading (van den Broek, Risden, Fletcher, & Thurlow, 1996; van den Broek, Young, Tzeng, & Linderholm, 1999), though it is not included in the CI model. Mathematically, asymmetry implies that the weight connecting a simpler proposition to a complex one can be higher than the weight connecting a complex proposition to a simpler one. Asymmetry is very important for simulating the reader's recall, as it implies that complex propositions can be more highly activated than simpler ones, and thus the probability that they will be recalled is also higher.

To summarize, each proposition constructed by the reader in a cycle receives two types of signals, one coming from the text (external input e, equation 5), and another one coming from its neighboring propositions (internal input in, equations 6a and 6b). The total input to a unit i, which is called the net input, is the sum of internal and external inputs:

$$net_i = e_i + in_i, = \Sigma_\alpha t_\alpha . w_{i,\alpha}. + \Sigma_j a_j p_{ij} \qquad (7a)$$

for binary connections to unit i

$$net_i = 2 (a_j . a_k)^{1/2} . p_{ijk} \qquad (7b)$$

for Y-connections when i is an inference from units j and k. Note that inferences are activated by other propositions, but not by the text.

Thus, every proposition is activated or inhibited to a higher or lower degree, according to the value of the total input net_i.

Integration Phase

As in the CI model, in our extension, the result of the construction phase may be a network with incoherent or even contradictory nodes. It is at the end of the integration phase that coherence, which is simulated by the stabilizations of the activations and connections, is attained. However, in the CECI model, both the activation of propositions and the connection weights between nodes change, and this means learning in connectionist terms. Moreover, changes in the weights have an influence on the activation of units and vice versa, which simulates the changes occurring during text processing in the reader's mind. In the following sections, we will explain how all these changes are mathematically implemented.

Activation rule. The activation value of a proposition is a real number that indicates the level of response of the proposition to the net input. The activation rule is non-linear (Rumelhart & McClelland, 1986) and produces activation values that range between 0 and 1, though the limits –1 and +1 could be adopted without significant structural changes. In every iteration, CECI calculates the increase or decrease in activation, keeping the value gained or lost by the previous repetition in memory. For unit i, with an activation value a_i at a given moment, the equations for the change of activation are:

$$\Delta a_i = E\, net_i\,(1-a_i) - D a_i \quad \text{if } net_i > 0 \qquad (8a)$$

$$\Delta a_i = E\, net_i\, a_i - D a_i \quad \text{if } net_i < 0 \qquad (8b)$$

where E is a parameter which regulates the rate of increase or decrease of activation; D is a decline rate; a_i is the activation value in the iteration just before the one being calculated; and net_i is the sum of all internal and external input to unit i. That is, activation will increase if net input net_i is positive and the value of a_i is not the maximum possible, 1. Activation will decrease if the value of net_i is negative. Additionally, for an excitatory signal (equation 8a), the increase is larger when the current activation is far from the maximum, but it is smaller when the activation is near the maximum. In the same way, for an inhibitory signal (equation 8b), the decrease becomes smaller as the value of the activation is lower, i.e., as it is closer to the complete deactivation of the unit.

The new activation value in each iteration a_i' will be:

$$a_i' = a_i + \Delta a_i \qquad (9)$$

It should be noted that, initially, all the activation values are null, except for the propositions carried over from previous cycles. According to Kintsch (1998), the pattern of activation in the network indicates the role that each node plays within the network, and may be considered a measure of the importance of each node in the reader's mind at a certain time.

The learning rule. One of the most important characteristics of a connectionist network is the capacity to learn from the environment. In practice, the rules of learning must specify how weights change during processing. This change is performed through a continuous interaction between the network and the environment. In the CECI model the reader's learning is represented by the capacity of the network to reproduce the same or a similar pattern of activation of the network, only with the activation of some pieces of information from the text. If the network behaves in a similar way as when the textual input was complete, it means that the reader has stored both the text propositions and their connections during reading. Initially, interconnection weights are 0, but they change during processing, and in connectionist terms this represents the process of learning.

In general terms, the process of change is as follows. Every time the network interacts with the environment, i.e., with the text, every node in the network receives external input. Then, the internal connections of the propositions are adjusted in order to reduce the distance between the signal received from the neighboring propositions (internal input) and the signal received from the external input. At the same time, the external connections of the propositions with the text input rise by increasing the weight of the connection between every proposition and the textual words. As our model is synchronic, the adjustment of the weights takes place in each iteration when activation changes, and occurs simultaneously in all the connections in the cycle.

The way external weights change is implemented by a modification of the rule used by Grossberg (1980; Rumelhart & Zipser 1986):

$\Delta w_{i,\alpha} = \phi\, a_i\, (w_{i,\alpha,max} - w_{i,\alpha})$ (10)

where a_i is the activation of unit i; $w_{i,\alpha}$ is the value of the connection from word α to unit i; $w_{i,\alpha,max}$ is the maximum possible value of this connection (see example in Table 1); and ϕ is the parameter that controls the rate of weight increase in every iteration. If the network is exposed to the environment long enough (enough number of iterations), which means deep processing in psychological terms, the external weights would reach values very close to the maximum. However, if the network is not exposed long enough, i.e., there are few iterations or little processing in psychological terms, the external weights would be close to the initial values.

The internal connection weights are also modified, but the way this is mathematically implemented depends on the type of connection considered, that is, excitatory, inhibitory or inferential.

(a) Excitatory connections. As we noted in the previous section, the connection weights between T-units, or between T- and K- units are initially 0. They change according to McClelland and Rumelhart's (1986) delta rule , i.e., the weights must be adjusted in such a way that the internal connections reproduce the pattern of activation produced from external input to that unit. In other words, the effect produced in the network by the text will be reproduced without the text. The delta value for the (i) unit is defined as follows:

$Dta_i = e_i - in_i$ (11)

If Dta_i is positive, this means that the units neighboring (i) are not producing enough activation to this unit. The system should react by increasing internal input in_i. In contrast, if Dta_i is negative, internal input is too high, and it should decrease. If Dta_i is zero, no change is needed. Accordingly, internal weights p_{ij} between two T- units or between a T-unit and a K-unit, must be modified in the following way:

$\Delta p_{ij} = \alpha\, dta_i\, a_j$ (12a) which constitutes the delta rule of learning, where p_{ij} is the connection weight from unit (j) to unit (i). In each iteration, a_j is the activation of unit (j) in the previous iteration; dta_i is the difference between internal and external inputs to unit (i) in the previous iteration; α is a parameter which controls the rate of the changes in the internal weights during each iteration. In equation (12), if dta_i is positive, weight p_{ij} increases, and the activation coming from unit (j) will increase, thus raising the value of the next internal input in_i. If dta_i is negative, the weight decreases and reduces the subsequent value of in_i. The value of all internal weights is stabilized when internal input equals external input, ($dta_i = 0$, according to equation 11).

It should be noted that equations 8, 9, 11 and 12a determine the weight value p_{ij}, even though the initial value is 0. At the beginning, a positive external input e will mean positive net (net > 0), which will produce an activation value higher than 0, although initially a_i and p_{ij} have null values (equations 8 and 9). On the other hand, with a non-null activation value, the value dta_i (equation 11) may be calculated, as well as the change for the weight value (equation 12a), which will produce a non-null value for p_{ij}. Thus, from the first iteration, the activation and weight values are non-null, and the system uses those values in the next iterations until they are stabilized.

During the first reading of the text, the pre-existing connections with non-null value, that is, the connections between the two K-units of prior background knowledge, are also modified as a consequence of the process described above. In this case, and also when the network is activated in further readings, all the internal weights are modified according to Hebb's rule for the excitatory connections:

$$\Delta p_{ij} = \gamma a_i a_j \qquad (12b)$$

where γ is a parameter that controls the change in the weights in each iteration. It should be noted that equation 11 is used to learn the connection between a T-proposition and another T-proposition or K-proposition in the presence of the text. Nevertheless, equation 12 is used to update the connection value between two units with a connection that had been learnt before (the connection between two propositions of prior background knowledge).

(b) Inhibitory connections. In the case of inhibitory connections for any pair of contradictory propositions, the anti-Hebb rule is used in such a way that the weight value decreases:

$$\Delta p_{ij} = -\eta a_i a_j \qquad (13)$$

where η is a parameter that controls the weight change in each iteration. Equation (13), together with the activation rule (equations 8a and 8b) and the fact that weights p_{ij} are not symmetrical, tends to nullify the activation value of one of the two contradictory propositions.

(c) Y-connections to process inferences. Finally, for connections among propositions that process inferences, it is necessary to define a relationship between three propositions, the two premises and the inference. To simulate this three-fold situation in our model, we use a modification of Hebb's rule:

$$\Delta p_{ijk} = \beta a_i (a_j+a_k)/2 \qquad (14)$$

where β is the parameter that controls the weight change, (i) is the unit associated with the inferred proposition, and (j) and (k) are the previous proposition units which give rise to the inference.

In the mathematical implementation of each cycle, the process of calculating activations and weights, which involves equations (7) to (11-14), is repeated iteratively until the system stabilizes in fixed values for the activations of all the units involved. The values of the activations and of the weights change in each iteration, which means that the system learns.

Once the integration phase is finished, the values of the internal weights between propositions are stored in long-term memory. When a cycle has finished, and processed propositions do not receive external input any longer, activations decrease until they reach null value, and the weights are weakened due to inactivity. In the CECI model, the decrease of the weight values is a percentage of the last increase. Our model does not include a progressive decrease of activation as the cycles begin again, like van den Broek et al.'s Landscape Model (van den Broek, Young, Tzeng, & Linderholm, 1999). This could be easily incorporated, and it would produce a sort of landscape of activations of the processed propositions.

As in other connectionist models, what the reader has stored in long-term memory after processing is the set of weights, but not the activation values. That is, the representation of the text that remains in long-term memory is represented by the set of weight values between propositions. Then, when some pieces of information stored in memory are activated, the complete pattern of connections stored during processing may be wholly reactivated. In our model, those pieces of information could be significant textual words (e.g., text title or main concepts). Therefore, our extension eliminates the CI model's distinction between matrix W, where processing takes place, and matrix M, which simulates what the reader recalls.

Parameters of the model and the simulation of individual differences in recall

Our model includes a number of parameters, some of which are common to all readers, while others must be adjusted for each reader. The parameters which are common to all readers are E and D, which control the rate of change in the activation values. The parameters that should be adjusted to different readers are α, γ, η, β, which regulate the rate of change of the internal weights with the delta rule, Hebb's rule, the anti-Hebb rule and the modified rule for inferences, respectively.

Other parameters that could be adjusted for each reader are $w_{i,\alpha}(0)$, $p_c(0)$, $p_{ijk}(0)$, and $in(0)$. Parameter $w_{i,\alpha}(0)$ is the initial value of the internal connection between word α and proposition i. Parameter $p_c(0)$ is the initial, residual and stored value of the connection between any two contradictory propositions; it indicates the reader's skill at detecting two contradictory pieces of information processed in a cycle. Parameter $p_{ijk}(0)$ represents the initial, residual and stored value to make Y-connections; that is, readers may have developed different skills to make logical inferences. Parameter $in(0)$ represents the action of an unknown part of the network of the reader's prior background knowledge to every K-proposition, incorporated into the system (K1: IS A (heat sensation) and K2: NEG-IS A (energy sensation) in our example). That is, every time a proposition $\underline{K_x}$ is activated, an unknown part of the reader's prior background knowledge will also be activated due to the pre-existing connections stored in the reader's LTM, and it will return an input to $\underline{K_x}$, called $\underline{in_x(0)}$. The value of this parameter will depend on the specific configuration of the reader's prior background knowledge network, which is unknown in the simulations.

Again, we use the same fragment of the text about heat we have already analyzed in order to examine how CECI simulates individual differences. We focus on one of the parameters just mentioned: initial prior background knowledge about some ideas related to the information of the text. A reader (Reader-1) could have a clear prior idea that heat is a sensation, i.e., proposition K1, which is easily activated when they read the text, and a vague prior idea that energy is not a sensation, proposition K2, which means low initial activation value. We assign initial values to these two propositions, $\underline{in_{K1}(0)} = 0.35$ and $\underline{in_{K2}(0)} = 0.05$. Another reader (Reader-2) has vague prior knowledge of both propositions, which would correspond to values $\underline{in_{K1}(0)} = 0.1$ and $\underline{in_{K2}(0)} = 0.1$.

The values given to the rest of parameters to both readers are the following: $\underline{E} = 0.50$; $\underline{D} = 0.25$; $\underline{\alpha} = 0.002$; $\underline{\beta} = 0.010$; $\underline{\gamma} = 0.002$; $\underline{\eta} = 0.005$; $\underline{P_{INF,T1,K2}(0)} = 0.50$; $\underline{\phi} = 0.02$; $\underline{p_c}(0) = -0.50$; $w_{i,\alpha}(0) = 0.20$; $w_{i,\alpha,\max} = 1$. As can be seen, in order to avoid the instability inherent in non-linear equations and abrupt changes in the activation and weight values, the values of the parameters for the equations involved in weight changes are very low, between 10^{-2} and 10^{-3}. Another possibility is to substitute the corresponding parameters of Hebb's rule, γ, η, β, for the parameters γ^n, η^n, β^n, where $\underline{n}$ is the iteration number. As γ, η, β are lower than 1, when we raise them to the natural power of $\underline{n}$, their values decrease in each iteration and change decreases in each step, thus facilitating the convergence to stable values. Finally, as to the decline in connection weights, the percentage of decrease has been set to 75% of the last increment.

For the two readers, the initial activation vector for propositions T1, T2, T3, T4, T5, T6, T7, K1, K2, INF1, is:

A(0)=(0,0,0,0,0,0,0,0,0,0),

For Reader-1, the final activation vector is:

A (final) = (0.48, 0.31, 0.40, 0.31, 0.61, 0.48, 0.31, 0.35, 0.22, 0.42)

As can be seen in the final vector, the most highly activated proposition for this reader is the complex causal proposition T5: BECAUSE (T1, T4). The final values for the internal weights are

Sending units

	T1	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**K1**	**K2**	**INF1**	**Receiving units**
W=	0	0.02	0.03	0.02	0.06	0.04	0.02	0.05	0.02	0.02	**T1**
	0.02	0	0.01	0.01	0.02	0.02	0.01	0.02	0.01	0.01	**T2**
	0.03	0.02	0	0.02	0.04	0.03	0.02	0.03	0.01	0.02	**T3**
	0.02	0.01	0.01	0	0.02	0.02	0.01	0.02	0.01	0.01	**T4**
	0.08	0.04	0.06	0.04	0	0.08	0.04	0.09	0.04	0.04	**T5**
	0.04	0.02	0.03	0.02	0.06	0	0.02	0.05	0.02	0.02	**T6**
	0.02	0.01	0.01	0.01	0.02	0.02	0	0.02	0.01	0.01	**T7**
	-0.01	0	-0.01	0	-0.02	-0.01	0	0	0	-0.52	**K1**
	0	0	0	0	0	0	0	0	-0.01	0	**K2**
	0	0	0	0	0	0	0	-0.52	0	0	**INF1**

The final weight for the inference connection from T1 and K1 to INF1 $\underline{p_{INF1,T1,K2}}$ is 0.612.

If now we simulate Reader-1's recall by re-activating the network with the word heat, the result is the following:

A (recall) = (0.24, 0.05, 0.08, 0.05, 0.28, 0.24, 0.05, 0.44, 0.10, 0.09)

This indicates that the most highly activated proposition is K1, followed by T5, T1 and T6, but that INF1 proposition gets much lower activation. In other words, CECI predicts that the prior knowledge idea that heat is a sensation will continue being very active in the reader's Long-Term Memory, whereas the inference heat is not a sensation will not appear in the recall protocol. It should be noted that for the CECI model, the word chosen to re-activate the network has an impact on the final activation. That is, if another word had been chosen to reactivate the network, the final activation vector would have been different, which is psychologically plausible.

The results of the simulation for Reader-2 are quite different. This reader had much less prior knowledge of proposition K1 and a slightly better prior knowledge of proposition K2 (energy is not a sensation). In this case, the final activation vector is the following:

A (final) = (0.46, 0.28, 0.38, 0.28, 0.59, 0.46, 0.28, 0, 0.25, 0.54)

This indicates that INF1 (heat is not a sensation), which contradicts K1, gets highly activated, whereas K1 reaches the lowest activation. This means that the inference has reduced the initial low activation of K1. The internal weights are very similar to those mentioned above with the following exception: $p_{INF1,T1,K2}$ = 0.701 and p_c = -0.50. If now we re-activate the network with the word heat to simulate the reader's recall, the final activation vector is:

A (recall) = (0.22, 0.03, 0.06, 0.03, 0.25, 0.22, 0.03, 0.07, 0.16, 0.38)

The activation of INF1 continues to be very high, whereas it is very weak for K1. In other words, CECI predicts that the idea that heat is not a sensation would be included in this reader's recall protocol, along with propositions T5 ,T1 and T6, whereas the vague contradictory prior knowledge would not be apparent.

To summarize, CECI can easily deal with individual differences in recall by introducing small differences in some parameters, such as prior background knowledge. Other possibilities will be explained in the following section in which CECI predictions are tested against empirical data.

Testing the CECI model

The CECI model was tested against two sets of data. The first was an empirical study on conceptual change in science in which CECI predictions are tested against the recall scores of two students with different prior background knowledge. The second set of empirical data comes from

Otero and Kintsch's (1992) study on failure to detect contradictions in a text. There we show that CECI predictions fit empirical data better than the CI model.

Prediction of students' recall

In order to test the model, we compared the results of an empirical study on conceptual change in science to the predictions made by the model. The empirical study was conducted with a group of undergraduate students at the School for Teachers at the University of Valencia, who differed in their background knowledge about heat, temperature and heat transfer (Padilla, 2000). Some of them had many misconceptions about heat and temperature, while others had well-elaborated scientific knowledge acquired at school.

In the empirical study, the information regarding the previous background knowledge of the students was obtained through a number of tests: an open-ended questionnaire, a multiple choice test, a concept map, and some observation and explanation from an experiment in the lab. Then, the participants read a scientific text where the notions of heat, temperature and heat transfer were explained. Part of the text was the same for all the participants, whereas another part was especially designed to refute the specific misconceptions detected for every student. After reading the text, the students were asked to answer literal and inference questions in order to promote conceptual change. Students were allowed to re-read the text to answer the questions. Finally, the students' comprehension was tested with a recall test.

To test the CECI model we compared the recall data of two students to CECI predictions. Both subjects showed good reading skills but differed considerably in their prior background knowledge. One of the students had very little scientific knowledge and many misconceptions about heat, temperature and heat transfer, whereas the other student had a great deal of scientific knowledge and very few misconceptions. Both students showed significant learning after reading the text (Padilla 2000).

The recall protocol of both students was compared to the most highly activated propositions in the connectionist network produced with the CECI model for each of the students.

The propositions forming the network for each student (T-propositions, K-propositions and INF-propositions) were obtained in the following way. T-propositions were obtained by dividing the text into propositions. K-propositions; that is, propositions coming from the reader's background knowledge, were detected from the answers of every student to the prior background knowledge tests. INF-propositions were obtained by analyzing the recall protocol of the two students. All the inferred ideas included in every recall protocol were selected. Then, those inferred ideas which could be obtained by combining either two T-propositions, a T-proposition and a K-proposition, or two K-propositions, by applying simple syllogisms such as modus tollens or modus ponens, were considered. It should be noted that these inference propositions were the minimum necessary to account for the information recalled by the student. The network was processed according to the CECI model previously explained. The final result is the stored value of the connection weights within each cycle and between cycles.

The values of the parameters used in the simulation were as follows: $E = 0.50$; $D = 0.25$; $\alpha = 0.25$; $\gamma = 0.35$; $p_c(0) = -0.20$, $w_{i,\alpha}(0) = 0.75$, $w_{i,\alpha,max} = 1$. The parameters that should be adjusted for each student were chosen from a narrow range. For the reader with deeper prior background knowledge and few misconceptions, the definite values were chosen from the following range:

$in(0) = 0.05 - 0.5$; $p_{ijk}(0) = 0.30 - 0.80$; $\eta = 0.35 - 0.50$; $\beta = 0.35 - 0.45$.

For the reader with less prior background knowledge and many misconceptions, the values were chosen from the following range:

$in(0) = 0.05 - 0.35$; $p_{ijk}(0) = 0.75 - 0.85$; $\eta = 0.80 - 0.90$; $\beta = 0.75 - 0.85$.

In this simulation, we used parameters γ^n, η^n, β^n, where n is the iteration number, instead of γ, η, β, in order to avoid the instability inherent in non-linear equations and the abrupt changes in the activation and weight values that were mentioned in the previous section. In each case, the value that best fitted the empirical results was chosen. The weakening of the internal weights was established at 75% of the value of the last increment, and the decline of external weights was 50% of the last increment.

Once the processing was simulated, the following five textual words were activated: heat, temperature, energy, conductor, and insulator. The same five words were given to the students to activate their recall. It should be noted that the words were important concepts in the text and that they were the arguments of many propositions. Therefore, those words spread their activation throughout the network, activating many propositions, though none of them reached maximum activation as none of the terms corresponded to a complete proposition. Finally, the system stabilized, showing the final activation of each proposition. It was assumed that the most highly activated propositions were those that the students would include in their recalls.

Figure 1 shows the distribution of the activation of propositions for the simulation of the student with little prior background knowledge. The distribution for the student with deep prior background knowledge was very similar.

As can be seen in Figure 1, the number of propositions as a function of the activation is distributed with a bi-modal shape. The group of propositions with the highest level of activation values is located on the right. It was predicted that they would be the most frequently recalled propositions.

PLEASE, INSERT FIGURE 1 ABOUT HERE

To test CECI prediction against these empirical data, a threshold value t was established. It was assumed that all the propositions with an activation value equal to or higher than t would be included in the recall protocol. However, other propositions could be recalled, but their activation value would be below t. Finally, other propositions would not be recalled at all, some of which would reach a value equal to or above t, whereas some others would have a value below t. These four possibilities are represented in Table 2.

PLEASE, INSERT TABLE 2 ABOUT HERE

From Table 2, three characteristics of the simulation can be established: (1) Sensitivity, or the capacity for the simulation to identify propositions recalled by the student, which are propositions with an activation value equal to or above t, that is, A/(A+B); (2) Specificity, or the capacity to identify non-recalled propositions, which are propositions having an activation value below the threshold value t, that is, D/(C+D); and finally, (3) diagnostic effectiveness, or the capacity to identify both recalled and non-recalled propositions, that is, (A+D)/(A+B+C+D). These three indices take values between 0 and 1, and they should be high enough simultaneously for the simulation to be considered acceptable.

PLEASE, INSERT FIGURES 2a AND 2b ABOUT HERE

Figures 2a and 2b show the three indices, i.e., sensitivity, specificity and diagnostic effectiveness, as a function of the threshold value t, for the simulation of the recall of the students with high and low prior background knowledge, respectively. As can be seen, there is a range for the threshold value t for which the three indices are simultaneously high. For example, for the student with high prior background knowledge (Figure 2a), the three indices are higher than .75 when t reaches a value between .63 and .67. More specifically, when t is equal to .66 the values are .92; .84 and .94 for sensitivity, specificity and diagnostic effectiveness, respectively. For the student with low prior background knowledge (Figure 2b), the three indices are higher than .75 when t reaches a value between .60 and .64. When t is equal to .63, sensitivity, specificity and diagnostic effectiveness reach the value of .92; .84 and .94, respectively. This means that for the optimal threshold activation values -.66 and .63 -, more than 9 out of 10 propositions are correctly identified as both recalled, and recalled and non-recalled simultaneously. It should also be noted that all the information actually recalled by the two students was included within the subgroup of the most activated propositions represented in Figure 1.

We have just showed that the simulation for a low-knowledge reader has enough sensitivity, specificity and diagnostic effectiveness to capture the behavior of a low-knowledge reader, and that the simulation for a high-knowledge reader also has sensitivity, specificity and diagnostic effectiveness to capture the high-knowledge reader's behavior . However, to better test the model we computed the sensitivity, specificity and diagnostic effectiveness corresponding to the high-knowledge simulation for the low-knowledge reader and vice versa, i.e., to compute the three indices corresponding to the low-knowledge simulation for the high-knowledge reader. If our simulation is adequate to capture individual differences, then the three indices should be worse now than before. The results of both simulations are represented in Figures 3a and 3b, respectively.

PLEASE, INSERT FIGURE 3a AND 3b ABOUT HERE

As can be seen in both Figures 3a and 3b, there is no threshold value $\underline{t}$ for which the three indices, i. e., sensitivity, specificity and diagnostic effectiveness, are simultaneously high enough. Moreover, when diagnostic effectiveness reaches high values, for example, higher than .75, predictions of recall, i. e., sensitivity, are very low, around .25 for the two simulations (see Figures 3a and 3b). It should also be noted that those simulations can not give an account of the entire recall, even when all the activated units are taken into account (i. e., sensitivity max $\approx$ 0,70 when $\underline{t}$ = 0). Therefore, the simulation now is worse than it was when the parameters corresponding to the low-knowledge reader were applied to the low-knowledge reader's recall, and vice versa. To conclude, it can be stated that the CECI model fits the individual differences in recall obtained by Padilla (2000) quite well.

<u>Failures to detect contradictions in a text</u>

Otero and Kintsch (1992) showed that the CI model could simulate empirical data on students' failure to detect contradictions in a text. Otero and Kintsch asked a group of 10th and 12th graders to read a short text of six sentences on <u>superconductivity,</u> where the second and the sixth sentences were explicitly contradictory (see the <u>underlined</u> sentences below).

> Superconductivity is the disappearance of resistance to the flow of electric current. Until now it has only been obtained by cooling certain materials to low temperatures near absolute zero. That made its technical applications very difficult. Many laboratories are now trying to produce superconducting alloys. Many materials with this property, with immediate technical applicability, have recently been discovered. Until now superconductivity has been achieved by considerably increasing the temperature of certain materials.

The readers were instructed to read the text and indicate any difficulty understanding it, though they were not told about the contradiction between the two sentences. Then, the students wrote everything they could recall from the text. Otero and Kintsch found that only 40.3% of participants were able to detect the contradiction. After analysing the recall protocols, Otero and Kintsch classified the non-detectors students into different groups; the ones relevant for our purposes are the following two:

Type II: Participants that recalled that superconductivity had been obtained by cooling certain materials, but not the contradictory idea.

Type IV: Participants that recalled that superconductivity had been obtained by increasing the temperature of certain materials, but not the contradictory idea.

Then, Otero and Kintsch simulated empirical data of normal readers (detectors) and superficial readers (non-detectors of Types II and IV) with the CI model. According to the model, the contradiction is detected only when, in the last cycle, the two contradictory propositions, (UNTILNOW{OBTAIN (supercon, cool (material, temperature))} from cycle 2 and UNTILNOW{OBTAIN (supercon, increase (material, temperature))} from cycle 6, are simultaneously processed, and they both reach a non-null activation. Otherwise, the detection is not possible.

To simulate the reader’s recall, Otero and Kintsch adjusted some parameters in the CI model. First, the authors identified important propositions, i. e., the macropropositions, and they hypothesized that macropropropositions should be processed differently than the rest. Then, all link strengths were set to 1 in matrix W, except for the self-strength of macropropositions, which was set to 2 for detectors and 10 for non-detectors. It should be noted that neither the selection of a proposition as a macroproposition, nor the strength value for a link, derive directly from the CI model.

The CECI model assumes that the activation of the propositional network in a cycle depends on the text input. Accordingly, we simulated the differences between normal (detector) and superficial (non-detector) readers with the CECI model, by varying the exposition of the propositional network to the text, i.e., by varying the number of iterations in a cycle. It should be noted that for CECI, similarly to most connectionist models, the number of iterations has psychological meaning as the number of iterations is closely related to the time the system is exposed to the environment, which causes learning. Thus, the longer the system is exposed to the environment, the more learning is produced, though this relationship is not linear. Consequently, using the number of iterations to simulate the differences between detectors and non-detectors sounds logical for CECI, though it makes no sense for the CI model, as it does not incorporate this connectionist feature. In fact, a major goal of CECI is to incorporate some of the features of the usual connectionist models to the CI model.

We assumed that the number of iterations in the model is tightly related to reading times, as the more times the network iterates, the longer the system is exposed to the text. On the other hand, increased reading have empirically been found to be associated with the detection of inconsistencies in a text. For example, Albrecht and O’Brien (1993) and Poynor and Norris (2003) have found that readers show longer reading times when reading sentences inconsistent with a portion of a text previously read than when reading consistent sentences. Moreover, Long and Chong (2001, Experiment 1) found that good comprehenders read inconsistent sentences more slowly than consistent ones, whereas this difference was not apparent in the case of poor comprehenders. Therefore, we could assume that the poor readers’ (i.e., non-detector) behavior could be simulated by allowing the propositional network to iterate fewer times, i.e., exposing it to the environment for a shorter time, than the propositional network of good readers (i. e., detectors).

Consequently, for normal readers, the network was allowed to iterate until the change of activation was smaller than 0.001. This simulates a reader who processes the text carefully and stores all the links among propositions in their LTM, which allows them to recover all the textual information after receiving any fragment of the text. To simulate the processing of superficial readers, the network was exposed to less textual input by stopping iterations when the change of activation was smaller than 0.05 . Thus, the links among propositions are not so high as to allow recovery of all the information from the text after receiving any fragment. Therefore, the key difference between normal and superficial readers is the depth of processing in terms of the iterations needed to establish all the links among the propositions in every cycle.

It should be noted that in the CECI model, any proposition already processed in a cycle can be re-activated during processing. Thus, it is possible that, when the reader processes the sixth sentence, some propositions from the second sentence are re-activated because they share some arguments. This possibility should be greater for normal readers than for superficial readers, as they processed the text more. It should also be noted that neither prior knowledge propositions nor inference propositions were included in our simulation.

The parameters were adjusted to the following values:

$\underline{E} = 0.50$; $\underline{D} = 0.25$; $\underline{\alpha} = 0.004$; $\underline{\gamma} = 0.002$; $\underline{\eta} = 0.002$; $\underline{\phi} = 0.05$; $\underline{p}_c(0) = -0.6$; $\underline{w}_{i,\alpha}(0) = 0.2$; $\underline{w}_{i,\alpha,max} = 1$. The weakening of the internal weights was established at 75% of the value of the last increment and the decline of external weights was 50% of the last increment.

According to the CECI model, the processing of a normal reader is simulated by allowing the network to iterate until the change of activation is smaller than 0.001. This makes the external weights of connections between the text and the propositions reach the maximum value, and it also makes the internal weights between pairs of propositions adequate to reactivate the pattern of connections among the propositions stored during processing. When cycle 6 begins, proposition P64: UNTIL-NOW{OBTAIN (superconductivity, increase (material, temperature))} from cycle 6 reactivates proposition P29: UNTIL-NOW{OBTAIN (superconductivity, cool (material, temperature))} from cycle 2 as they overlap in many arguments. At the beginning, proposition P29 had higher activation than proposition P64 (see Figure 4) since it had already been processed, and thus, its external weights were very high. However, after about 800 iterations, proposition P64 increases its activation because of the increase in connections between the proposition and the text, until it reaches a similar level of activation as proposition P29. Then, the contradiction is detected.

PLEASE, INSERT FIGURE 4 ABOUT HERE

If the network iterates less than 800 times in cycle 6, which means that sentence 6 is processed superficially, proposition P29 inhibits proposition P64. Therefore, the contradiction is not detected, and the reader will recall proposition P29 but not P64. This simulates non-detector Type II according to Otero and Kintsch’s classification.

A non-detector Type IV would be a reader who processes the whole text superficially. According to the CECI model, the networks in every cycle stop iterating when the change of activation is smaller than 0.05, which makes main propositions reach very low activation values, about 0.20. This makes the reactivation of proposition P29 from cycle 2 when cycle 6 begins very low, and it decreases after a few iterations; whereas activation of proposition P64 rises (see Figure 5). Thus, proposition P64 inhibits proposition P29. Therefore, the contradiction is not detected, and in this case the reader would recall proposition P64 but not P29. This simulates a non-detector Type IV.

PLEASE, INSERT FIGURE 5 ABOUT HERE

It should be noted that the way superficiality is simulated is similar for detectors type II and type IV : in both cases the key element is that the superficial readers’ behavior is simulated by allowing the system to iterate fewer times than in the simulation of detectors. In summary, the

CECI model simulates the differences between detectors and non-detectors (i.e., normal and superficial readers) in terms of the time the system is exposed to the environment (text); i.e., the number of iterations needed to establish all the links among propositions in every cycle, assuming that the differences in the number of iterations when modelling text processing simulate the differences in reading times in empirical studies.

Apart from studies on the detection of inconsistencies in a text (Albrecht & O'Brien, 1993; Poynor & Norris, 2003; Long & Chong, 2001), reading times have also been found to be related to deep processing in other domains, including text recall (Coté, Goldman, & Saul, 1998, Experiment 2), answering questions to on a text (Vidal-Abarca, Martínez, Gilabert & Sellés, 2005), and strategic text processing (Magliano, Trabasso & Graesser, 1999). Therefore, modelling differences between detectors and non-detectors through the number of iterations when processing the text is consistent with empirical studies which show that reading times decrease when processing demands increase, and that good comprehenders are more sensitive to those demands than poor comprehenders.

Apart from simulating the processing of normal and superficial readers, the CECI model can also predict detectors' recall better than the CI model. As explained in the first section of this paper, the values in the main diagonal of matrix M obtained with the CI model are related to the probability of recall of each proposition. According to Otero and Kintsch (1992), the CI model underpredicted the recall of the two contradictory sentences, S2 and S6, but it overpredicted the recall of S3, S4 and S5 (see Figure 6). However, the predictions of the CECI model fit the empirical data of recall better, as explained below.

PLEASE, INSERT FIGURE 6 ABOUT HERE

To simulate CECI predictions for recall, we followed a procedure similar to that in the previous section. After simulation, the network was re-activated by activating the word superconductivity. It spread its activation throughout the network, activating all the propositions. It was assumed that the most activated propositions would be those that the students would include in their recall protocols. The final activation levels for the main propositions in each sentence were as follows [2] : S1: 0.50; S2: 0.56; S3: 0.17; S4: 0.12; S5: 0.14; S6: 0.39.

To test CECI prediction against the empirical data, we established a threshold value t, above which a proposition would be included in the recall protocol, as we did in the previous section. We could assume that value t was normally distributed. According to this, the area within the Gaussian curve, abscissa axis and tj represented the percentage of population with a t value lower than tj. Therefore, these people could recall the propositions with an activation value higher than tj. For one proposition with activation value a, the percentage of the population who could recall this proposition corresponded to the area in the t-distribution for t values lower than a (i.e., the activation was higher than the threshold). Accordingly, we could calculate the percentage of the population who could recall the main propositions of each sentence. As Figure 6 shows, a good fit to Otero and Kintsch's data was obtained with a normal distribution of t with mean value equal to 0.35 and standard deviation of 0.3. A similar procedure could be implemented to simulate non-detectors' performance.

Conclusions

In this paper, we have presented a model that extends the connectionist features of Kintsch's CI model (1988, 1998):the Connectionist Extension of the CI (CECI) model. One important innovation of the CECI model is that it simulates the process of learning from text. The model incorporates a connectionist algorithm to simulate the process of learning, i.e., the change in the connection weights between the nodes (propositions) in the network. The algorithm establishes a simple mechanism to simulate the process by which the connection values between propositions are created and changed during text processing. It is an important innovation because in the CI model the connections are established a priori by researchers and they do not change during processing. Furthermore, the CECI model simulates the processing of contradictory or mutually inhibiting propositions more satisfactorily than the CI model, due to the asymmetry resulting from the processing. By means of this, the reader discards one of the contradictory propositions and accepts the other one. The CECI model also simulates the process of storing and retrieving information from long-term memory in a different way than the CI model. The CI model considers two matrices with different weights, one of them playing a part in the processing of each cycle, and the other one simulating the storage of relationships among propositions in long term memory. However, for the CECI model, the same matrix is used to simulate both the processing of the text and the storage of the links among propositions in long term memory.

The second innovation of the CECI model is that it takes into account two of the individual differences readers may bring to their reading. Only two are considered here, although others could be incorporated easily. First, apart from the background knowledge propositions incorporated into the network for the simulation, an unknown part of that background knowledge is also activated during processing, which also influences the network. This unknown component depends on the specific background knowledge of each reader and as such, it is a source of individual differences. The CECI model incorporates a parameter to simulate the influence of the activation of the unknown part of the reader’s background knowledge on actual processing. Second, readers differ in their ability to make inferences, which is essential for learning from a text. The CECI model incorporates a parameter to adjust the simulation to individual differences in making simple logical inferences such as modus ponens or modus tollens. These two free parameters are adjusted a posteriori. The range of their values has not been established yet, although a statistical study, which will investigate their distributions in the population, has been planned.

The third innovation of the CECI model is that it simulates the processing of superficial and normal readers in terms of differences in reading times by varying the number of iterations allowed in every cycle during the integration process. This parameter corresponds to the time the network is exposed to the environment (the text). A short exposition produces poor learning of the internal weights between propositions, whereas an exposition that is long enough produces good learning; that is, the complete pattern of connections stored during processing may be wholly reactivated. This allows the model to simulate empirical results on text comprehension.

The comparison between the results of the simulation and the empirical results obtained with two participants in an empirical study is promising. Two students who differed in their previous background knowledge read and studied scientific texts and were then tested on inferences and recall. By establishing the appropriate threshold value for the propositions to be recovered from the network, the diagnostic effectiveness of the CECI model, i.e., the capacity to identify both recalled and non-recalled propositions, was above 0.90 for both participants. Moreover, most of the propositions actually recalled by both students reached the highest activation level in the simulation. We could also simulate the detection of contradictory sentences in a text, a phenomenon already simulated with the CI model (Otero and Kintsch, 1992). The CECI model not only simulates the pattern of empirical results found by Otero and Kintsch with normal and superficial readers, but it also fits results on recall better than the CI model.

Other researchers have also constructed connectionist implementations grounded on the psychological bases of Kintsch's CI model. A relevant example is van den Broek and his colleagues' Landscape Model (van den Broek, Risden, Fletcher, & Thurlow, 1996; van den Broek, Young, Tzeng, & Linderholm, 1999), where the processing units are concepts and propositions. The mathematical implementation of the activation and deactivation rule of each concept, cycle after cycle, is especially relevant in this model. If we consider how each concept is activated from process cycle to process cycle, we obtain a general picture of the relative importance of each piece of information in any given moment and we can make reasonable hypotheses about the reader's recall. Important aspects of this model, such as the dependency of subsequent activation on previous values and on existing connections, or the directionality of the connections (asymmetry), are shared by the CECI model. Another feature of the Landscape model, such as a rule for the gradual decrease of activation through the processing cycles, could be incorporated into the CECI model without major structural changes. This would reproduce the simulation in a similar way to the Landscape Model, but would use propositions instead of concepts. Other components of the Landscape Model, such as the limited pool of activation, or auto-connections, are not present in the CECI model.

Some features of the CECI model need further investigation. tone concerns one of the main goals of the model, that is, to deal with individual differences in text processing. For instance, the CECI model is very sensitive to small changes in the parameters of the model. For certain ranges, small changes in the values produce significant differences in the results, which is a common feature of many non-linear systems. Another feature related to individual differences is the distribution of the values of parameters in the general population. It is necessary to know the distribution of some parameters (e.g., the ability to make logical inferences) to simulate a wide range of phenomena.

Another area for further investigation is that the CECI model does not incorporate a procedure to simulate the reader’s recall. That simulation involves a complex process which is strongly influenced by the piece of information used as stimulus to reactivate the network. Another feature that needs to be investigated is how the textual input activates propositions. CECI’s general mechanism is extremely simple in psychological terms, but it is obvious that among textual words and propositions there are complex processes mediated by grammar, among other elements.

References

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. Journal of Experimental Psychology: Learning, Memory and Cognition, 19(5), 1061-1070.

Bovair, S., & Kieras, D. E. (1985). A Guide to Propositional Analysis for Research on Technical Prose. In B. K. Britton and J. B. Black (Eds.), Understanding Expository Text, pp. 315-362. Hillsdale, NJ: Lawrence Erlbaum Associates.

Britton, B. K., & Graesser, A. C. (1996). Models of Understanding Text. Mahwah, NJ: Lawrence Erlbaum Associates.

Coté, N., Goldman, S. R., & Saul, E. U. (1998). Students making sense of informational text: Relations between processing and representation. Discourse Processes, 25(1), 1-53.

Grossberg, S. (1980). How does a Brain Build a Cognitive Code? Psychological Review 87; 1-51.

Guha, A. & Rossi, J. P. 2001. Convergence of the Integration Dynamics of the Construction. Integration Model. Journal of Mathematical Psychology 45, 355-369.

Kintsch W. & van Dijk, T. A. (1978). Towards a Model of Text Comprehension and Production. Psychological Review 85, 363-394.

Kintsch, W. & Welsch, D. (1991). The Construction-Integration Model: A Framework for Studying Memory for Text. In W.E. Hockley & S. Lewandowsky (eds): Relating Theory and Data: essays on human memory in honor to Bennet B. Murdock. Hillsdale, NJ: Lawrence Erlbaum; pp 367-385.

Kintsch, W. (1988). The Use of Knowledge in Discourse Processing: A Construction-Integration Model. Psychological Review 95, 163-182.

Kintsch, W; Welsch, D; Schmalhofer, F; & Zimny, S. (1990). Sentence Memory: A Theoretical Analysis. Journal of Memory and Language 29, 133-159.

Kintsch,. W. (1998). Comprehension. A Paradigm for Cognition. Cambridge , UK: Cambridge University Press.

Long, D. L., & Chong, J. L. (2001). Comprehension skill and global coherencde: A paradoxical picture of poor comprehender's ability. Journal of Experimental Psychology: Learning, Memory and Cognition, 27(6), 1424-1429.

Magliano, J.P., Trabasso, T. & Graesser, A. C. (1999). Strategic Processing during comprehension. Journal of Educational Psychology, 91(4), 615-629.

McClelland, J.L., & Rumelhart, D.E. (1985). Distributed Memory and the Representation of General and Specific Information. Journal od Experimental Psychology: General 114, 159-188.

McClelland, J.L., Rumelhart, D.E. & the PDP Research Group. (1986). Parallel Distributed Processing: Vol I: Foundations. Cambridge, MA: MIT Press.

McClelland, J.L., Rumelhart, D.E. & the PDP Research Group. (1986). Parallel Distributed Processing: Vol II: Psychological and Biological Nodels. Cambridge, MA: MIT Press.

Otero, J. & Kintsch, W. 1992. Failures to Detect Contradictions in a Text: what readers believe versus what they read. Psychological Review 3 (4); 229-235.

Padilla, O. M. (2000). Conceptual Change in Heat and Temperature by Reading Text: a Connectionist Model. Unpublished doctoral dissertation,. University of Valencia, Spain.

Poynor, D. V., & Morris, R. K. (2003). Inferred goals in narratives: evidence from self-paced reading, recall and eye movements. Journal of Experimental Psychology: Learning, Memory and Cognition, 29(1), 3-9.

Read, S. J. & Miller, L. C. (1998). Connectionist Models of Social Reasoning and Social Behavior. Mahwah, NJ: Lawrence Erlbaum Associates.

Rodenhausen, H. (1992). Mathematical Aspects of Kintsch's Model of Discourse Comprehension. Psychological Review 99, 547-549.

Rumelhart, D.E. & Zipser, D. 1986. Feature Discovery by Competitive Learning. In D.E. Rumelhart, J.L. McClelland & the PDP research group. PDP Explorations in the Microstructure of Cognition. Vol I: Foundations, pp 151-193. Cambridge, MA: MIT Press..

Schmalhofer, F; McDaniel, M.A; & Keefe, D. (2002). A Unified Model for Predictive and Bridging Inferences. Discourse Processes 33 (2), 132.

Singer, M; & Kintsch, W. (2001). Text Retrieval: A Theoretical Exploration. Discourse Processes 31 (1), 27-59.

Van den Broek, P; Risden, K.; Fletcher, C.R. & Thurlow, R. (1996). A "Landscape" View of Reading: Fluctuating Patterns of Activation and the Construction of a Stable Meaning. In B.K. Britton and A.C. Graesser, (Eds), Models of Understanding Text, pp 165-188. Mahwah, NJ: Lawrence Erlbaum Associates.

Van den Broek, P; Young, M. , Tzeng, Y. & Linderholm, T. (1999). The Landscape Model of Reading. Inferences and the Online Construction of a Memory Representation. In H. Van Oostendorp and S. R. Goldman (Eds.), The Construction of Mental Representations during Reading, pp. 71-98. Mahwah, NJ: Lawrence Erlbaum Associates.

Van Dijk, T. A., & Kintsch, W. (1983). Strategies of Discourse Comprehension. New York: Academic Press.

Van Oostendorp, H., & Goldman, S. R. (1999). The Construction of Mental Representations during Reading. Mahwah, NJ: Lawrence Erlbaum Associates.

Vidal-Abarca, E., Martínez, T., Gilabert, R., & Sellés, P. (2005). Behavioral on-line measurements to assess reading comprehension skills and strategies: From research to practice. Paper presented at the 11th Biennial Conference of the European Association for Research on Learning and Instruction, Cyprus.

AUTHORS NOTE

We thank Walter Kintsch, two anonymous reviewers and the editor for their stimulating and helpful comments. All of them have contributed to improve the paper substantially.

Correspondence concerning this article should be addressed to Vicente Sanjosé, Department of Science Education, Escuela de Magisterio "Ausias March", Universitat de Valencia; PO Box 22045; 46071- Valencia – Spain. E-mail: Vicente.Sanjose@uv.es. A copy of the software (in QBASIC) to use CECI can be requested from this author.

Footnotes

[1]Note that it is inconsistent that CI could predict the recall of a complex proposition (e.g., T5), but not of the simpler propositions embedded in it (e.g., T1 and T4). Here, the values in the main diagonal for T5 and T1 are high, which indicates a high probability of recall for the two propositions, but it is not the case for T4, which is logically inconsistent. This is a problem not just to the CI model, but also to all proposition-based models of text comprehension, such as the CECI model.

[2]We suppose that the readers keep special atention to the first sentence. Then the net in the first cicle stops when the change in the activations is smaller than 0.0002. For the other sentences the net stops when the change in activation is smaller than 0.001.

Table 1. Propositions, text vector (t) and maximum possible strength of the external connection of each proposition to the words in the text ($w_{i,\alpha,max}$) for the sentence about heat

Cycle 1: Heat is energy transferred from one body to another because of a difference in temperature. The unit of heat is the calorie											
Propositions		Heat	Energy	Transference	Body	Because	Difference	Temperature	Unit	Measure	Calorie
	t	2/12	1/12	1/12	2/12	1/12	1/12	1/12	1/12	1/12	1/12
	w_{i,α_max} : maximum value for the external weights										
T1: IS-A (heat, T2)	$w_{T1\alpha}$	1	1	1	0	0	0	0	0	0	0
T2: IN (energy, transference)	$w_{T2\alpha}$	0	1	1	0	0	0	0	0	0	0
T3: BETWEEN (transference, body1, body2)	$w_{T3\alpha}$	0	0	1	1	0	0	0	0	0	0
T4: OF (difference, temperature)	$w_{T4\alpha}$	0	0	0	0	0	1	1	0	0	0
T5: BECAUSE (T1, T4)	$w_{T5\alpha}$	1	1	1	0	1	1	1	0	0	0
T6: OF (T7, heat)	$w_{T6\alpha}$	1	0	0	0	0	0	0	1	1	0
T7: REF (unit, calorie)	$w_{T7\alpha}$	0	0	0	0	0	0	0	1	0	1
K1: IS-A (heat, sensation)	$w_{K1\alpha}$	1	0	0	0	0	0	0	0	0	0
K2: NEG-IS-A (energy, sensation)	$w_{K2\alpha}$	0	1	0	0	0	0	0	0	0	0
INF1: NEG-IS-A (heat, sensation)	$w_{inf\alpha}$	1	0	0	0	0	0	0	0	0	0

Table 2. Classification of propositions depending on empirical data (i.e., recall) and the simulation (i.e., equal to, above or below the threshold value t)

Threshold t	Propositions actually recalled	Non-recalled propositions
Propositions with activation equal to or above **t**	**A**	**C**
Propositions with activation below **t**	**B**	**D**

Figure 1. Distribution of the activation of propositions for the simulation of the student with little prior background knowledge.

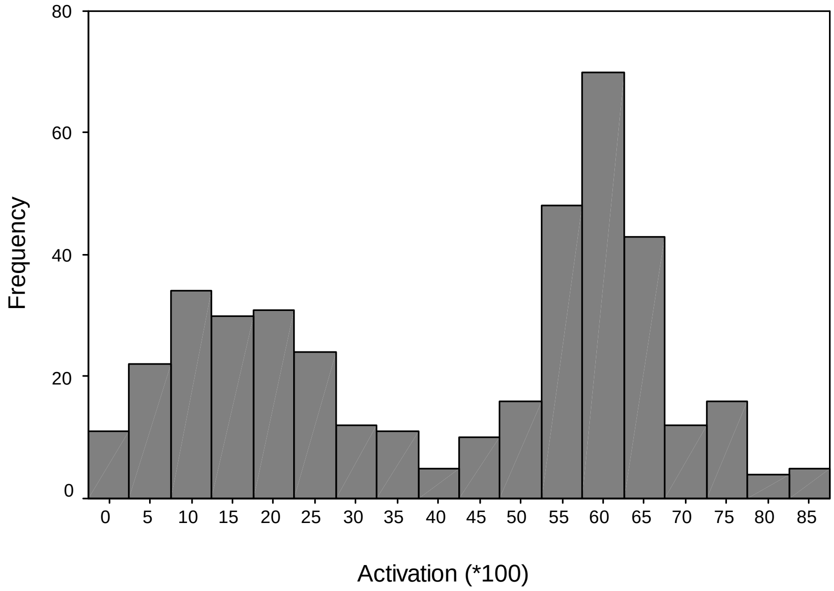

Figure 2a. Values of diagnostic effectiveness, sensitivity and specificity as a function of the threshold value t, for the simulation of the recall of the student with high prior background knowledge.

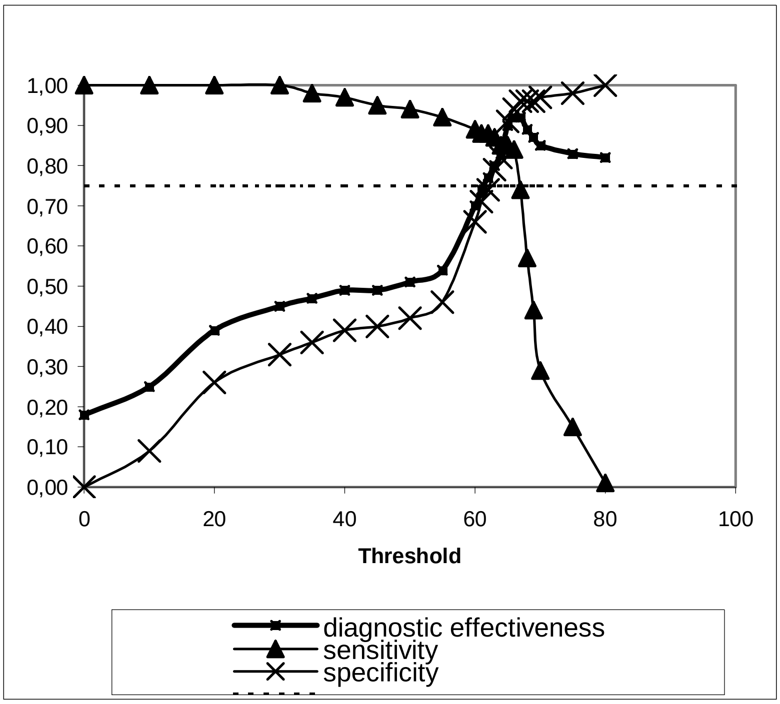

Figure 2b. Values of diagnostic effectiveness, sensitivity and specificity as a function of the threshold value t, for the simulation of the recall of the student with low prior background knowledge.

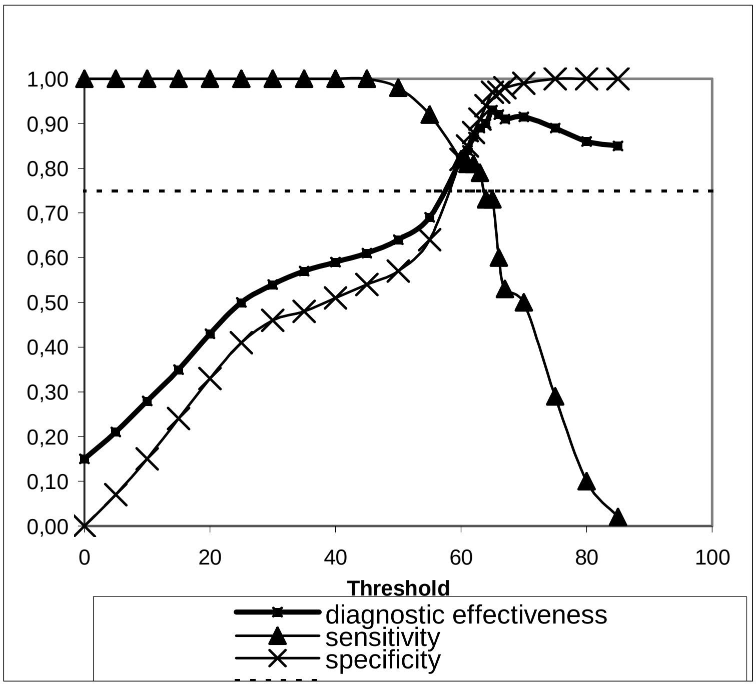

Figure 3a. Values of diagnostic effectiveness, sensitivity and specificity as a function of the threshold value t, for the high-knowledge simulation applied to the low prior background knowledge student's recall

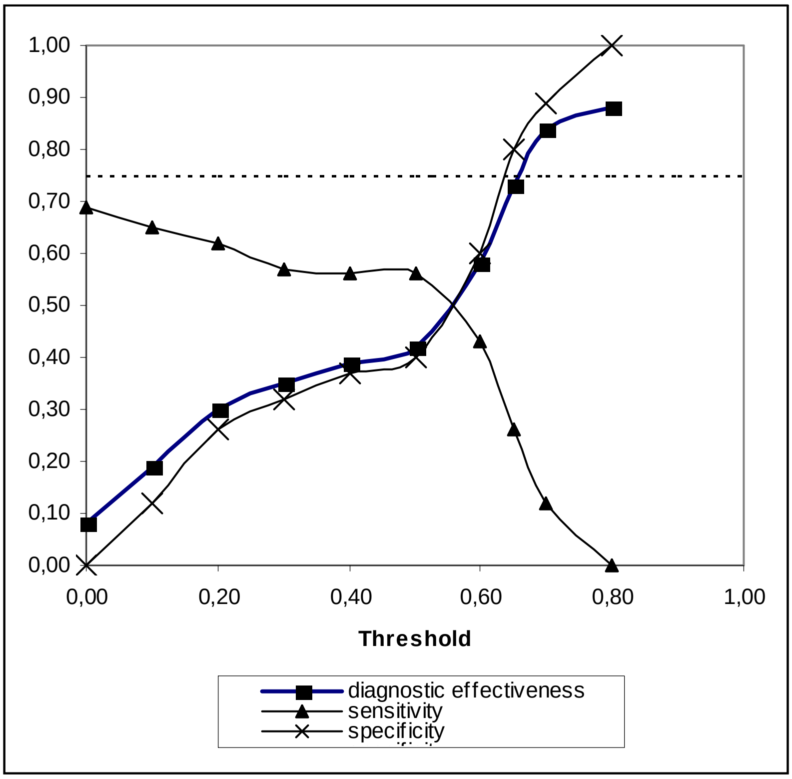

Figure 3b. Values of diagnostic effectiveness, sensitivity and specificity as a function of the threshold value t, for the low-knowledge simulation applied to the high prior background knowledge student's recall

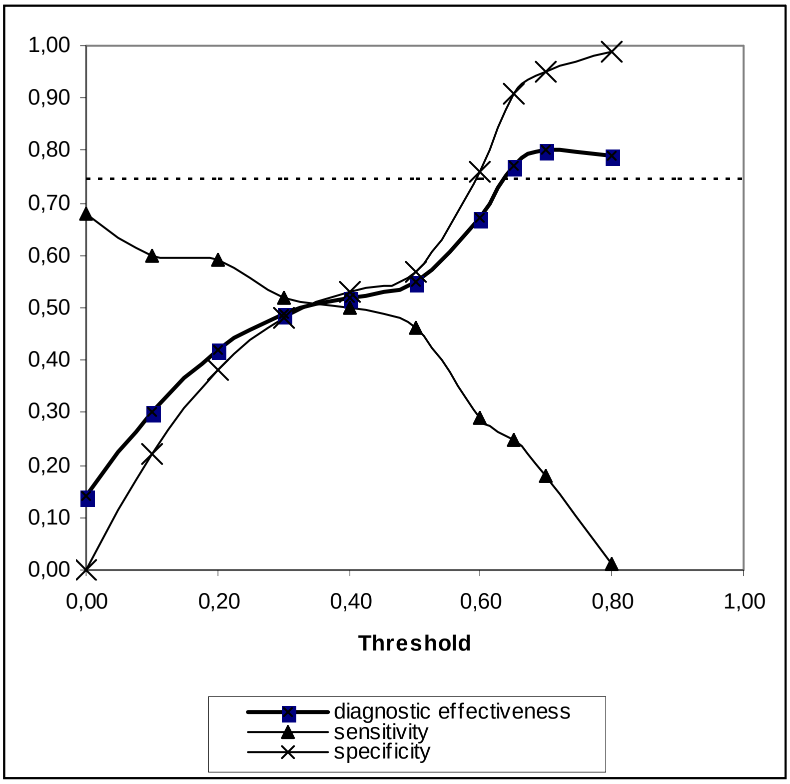

Figure 4: Evolution of the two contradictory propositions in the last cycle for normal readers.

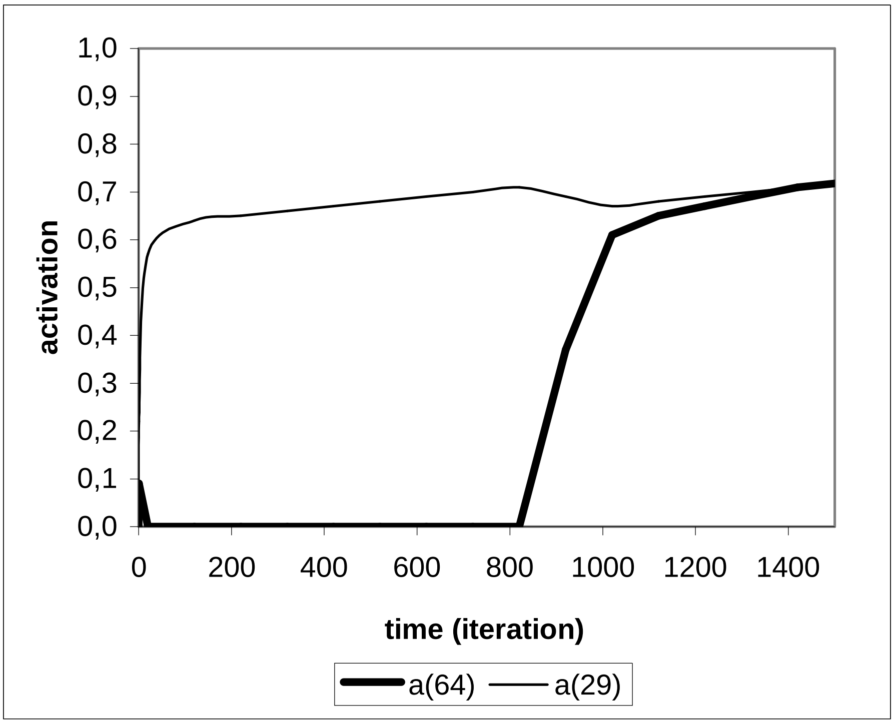

Figure 5: Evolution of the two contradictory propositions in the last cycle for superficial readers.

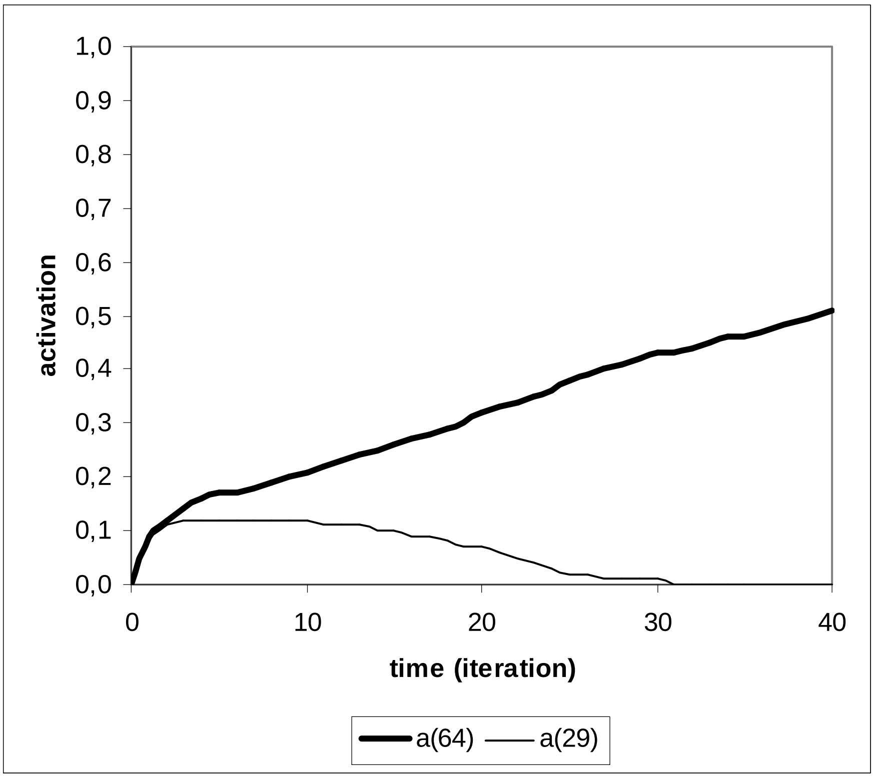

Figure 6. Recall of text sentences, and CI and CECI predictions of recall for normal readers (detectors)