

A computational modeling of semantic knowledge in reading comprehension: Integrating the landscape model with latent semantic analysis

Menahem Yeari^{1,2} · Paul van den Broek²

Published online: 6 July 2016
© Psychonomic Society, Inc. 2016

Abstract It is a well-accepted view that the prior semantic (general) knowledge that readers possess plays a central role in reading comprehension. Nevertheless, computational models of reading comprehension have not integrated the simulation of semantic knowledge and online comprehension processes under a unified mathematical algorithm. The present article introduces a computational model that integrates the landscape model of comprehension processes with latent semantic analysis representation of semantic knowledge. In three sets of simulations of previous behavioral findings, the integrated model successfully simulated the activation and attenuation of predictive and bridging inferences during reading, as well as centrality estimations and recall of textual information after reading. Analyses of the computational results revealed new theoretical insights regarding the underlying mechanisms of the various comprehension phenomena.

Keywords Reading comprehension · Computational modeling · Landscape model · Latent semantic analysis · Semantic knowledge · Inference generation · Information centrality · Text recall

Computational modeling of the cognitive mechanisms underlying discourse and reading comprehension has become popular in the last three decades. Various algorithms and mathematical equations have been formulated to simulate the operation of the cognitive structures and processes that support this

high-order, complex human skill (e.g., Frank, Koppen, Noordman, & Vonk, 2003; Golden & Rumelhart, 1993; Goldman & Varma, 1995; Kintsch, 1988, 1998; Langston & Trabasso, 1999; Myers & O'Brien, 1998; Schmalhofer, McDaniel, & Keefe, 2002; Thibadeau, Just, & Carpenter, 1982; van den Broek, Ridsen, Fletcher, & Thurlow, 1996). These computational models, which were evaluated against findings obtained with human readers, have been generally successful in simulating various phenomena in discourse and reading comprehension (e.g., Frank et al., 2003; Kintsch, 1998; Kintsch & Welsch, 1991; Linderholm, Virtue, Tzeng, & Van den Broek, 2004; Myers & O'Brien, 1998; Sanjosé, Vidal-Abarca, & Padilla, 2006; Singer & Kintsch, 2001; Trabasso & Wiley, 2005). However, the cognitive complexity of the comprehension skill has limited the extent of computational modeling. Computational models have simulated some of the underlying comprehension mechanisms by using detailed algorithms and mathematical equations, whereas other mechanisms have either been ignored or only been realized using hand-coded human approximations. Perhaps the most apparent challenge in devising a complete computational model of discourse and reading comprehension has been the integration of semantic knowledge representations (also referred as *world* or *general knowledge*) with dynamic comprehension processes (see Frank, Koppen, Noordman, & Vonk, 2008, for a review). On the basis of a large body of empirical evidence, the central role of semantic knowledge in discourse and reading comprehension has been widely accepted and well defined within diverse theoretical models (e.g., Bower, Black & Turner, 1979; Bransford & Johnson, 1972; Cain, Oakhill, Barnes, & Bryant, 2001; Chiesi, Spilich, & Voss, 1979; Cook & Guéraud, 2005; Graesser, Millis, & Zwaan, 1997; Graesser, Singer, & Trabasso, 1994; Kintsch, 1988, 1998; McNamara & Kintsch, 1996; McNamara & O'Reilly, 2009; Shapiro, 2004; van den Broek, 2010). Nevertheless,

✉ Menahem Yeari
Menahem.Yeari@biu.ac.il

¹ Bar Ilan University, Ramat-Gan, Israel

² Leiden University, Leiden, Netherlands

most computational models have not implemented both semantic knowledge and comprehension processes into one unified mathematical algorithm (e.g., Kintsch, 1988, Myers & O'Brien, 1998; van den Broek, 2010), whereas the remaining models have implemented a limited version of one of them that could not be applied to an actual set of behavioral data (e.g., Frank et al., 2003). The present article introduces a computational model that integrates the simulation of semantic knowledge and comprehension processes. First, however, we will briefly elaborate on the limitations characterizing the current computational models of discourse and reading comprehension.

Manual simulation of semantic knowledge

Several well-known computational models, such as the construction–integration model (Kintsch, 1988, 1998), the resonance model (Myers & O'Brien, 1998), the causal-network model (Langston & Trabasso, 1999), and the landscape model (Van den Broek, Young, Tzeng, & Linderholm, 1999), have specified in detail, using mathematical algorithms, the dynamic alterations in the activations of and connections between text units (words or propositions) throughout reading. However, these models have simulated semantic knowledge *manually* by establishing specific semantic connections between a *selected subset* of text units based on the researcher's knowledge, intuition, and/or research goals (Gaddy, van den Broek, & Sung, 2001; Goldman & Varma, 1995; Kintsch, 1988; Kintsch & Welsch, 1991; Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Langston, Trabasso, & Magliano, 1998; Linderholm et al., 2004; Myers & O'Brien, 1998; Otero & Kintsch, 1992; Radvansky, Zwaan, Curiel, & Copeland, 2001; Schmalhofer et al., 2002; Singer & Halldorson, 1996; Tapiero & Denhière, 1995; Trabasso & Bartolone, 2003; Trabasso & Wiley, 2005; van den Broek & Kendeou, 2008; van den Broek et al., 1996; van den Broek, Young, Tzeng, & Linderholm, 1999). This kind of hand-coded simulation is vulnerable to theory-confirmation biases. Moreover, this procedure has narrowed the variation of semantic connection strengths to a small number (sometimes only one; e.g., Singer, 1996) of strength values.

A semimanual approach has been adopted by the distributed situation space (DSS) model (Frank et al., 2003, 2008). This model uses a microworld representation of semantic knowledge. This microworld is initially created by formulating several propositions that describe simple situations occurring in that world (e.g., Bob and Jilly play soccer; Bob and Jilly play computer; Bob is tired; Jilly is tired; Bob wins; Jilly wins). In addition, semantic and logic relations are formulated between a subset of these propositions based on the researcher's knowledge (e.g., *if* Bob is tired *then* Jilly wins; *if* Bob and Jilly play soccer, *then* Bob and Jilly do not play computer). Then, a

computational algorithm, based on fuzzy logic equations, expands this microworld to include all possible combinations of, and relations between, the micro-world situations. This microknowledge supports the operation of dynamic computational models of discourse comprehension, such as the construction–integration model (Kintsch, 1998) and the connectionist model of Golden and Rumelhart (1993). These compute the semantic connections between the discourse units of a narrative that describes situations occurring in the same microworld. The DSS model has been successful in simulating comprehension phenomena, such as inference generation and reading times, in a conceptual manner, using artificial text-like scenarios occurring in the designed microworld (e.g., Frank, Koppen, Noordman, & Vonk, 2003, 2007, 2008). However, it could not be applied to simulate actual behavioral results obtained with natural complex texts that describe situations that occur in diverse microworlds.

A computational simulation of semantic knowledge

In a recent study, Yeari and van den Broek (2015) simulated knowledge-based inference activation during reading comprehension using the latent semantic analysis (LSA) model. LSA is a statistical algorithm that assesses the semantic relatedness of words or larger units of text (e.g., sentences, paragraphs) on the basis of the co-occurrence of words in the same text segments (e.g., paragraphs) across a large corpus of texts (e.g., Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). Words that appear together more frequently in these text segments (or appear together with the same set of other words) are considered to have higher semantic relatedness than words that appear together less frequently. The basic idea behind this statistical computation is that restricted contexts use semantically related ideas that come from the same conceptual topic. For example, *read* and *book*, which are semantically related concepts, tend to appear together more frequently within the same contexts than *walk* and *book*, which are semantically less related. The mathematical procedure behind LSA reproduced a multi-dimensional matrix that represents a “semantic space.” The “meaning” of each word is represented as a vector in this semantic space. The semantic relatedness between two words is determined by computing the cosine (i.e., distance in high dimensional space) between the vectors of these words. Cosines normally range from 0 (*no similarity*) to 1 (*high similarity*). For example, the cosine between *read* and *book* is .50, whereas the cosine between *read* and *ball* is .05.

Yeari and van den Broek (2015) applied LSA to compute the strength of semantic relatedness between texts and probe items (words or sentences) presented to human readers to assess inference activation (Beeman, Bowden, & Gernsbacher, 2000; Cook, Limber, & O'Brien, 2001; Klin, Guzman, & Levine, 1999; Klin, Murray, Levine, & Guzmán, 1999; Linderholm, 2002; Shears & Chiarello, 2004; Shears et al., 2007; Singer,

Halldorson, Lear, & Andrusiak, 1992; Singer & Halldorson, 1996). In these studies, inference activation was indicated by a faster response (e.g., naming) to probe items (e.g., *cut*) that represented a possible inference from the preceding texts (the inference condition; e.g., *stepping on a glass*) than to probe items that did not (the control condition; e.g., *noticing a glass*). Yeari and van den Broek examined whether the semantic relatedness between texts and probe items in the inference condition is stronger (i.e., higher mean of LSA cosines) than the semantic relatedness in the control condition.

Yeari and van den Broek (2015) found that LSA was quite successful in predicting inference activation. In most simulations, higher text-probe cosines were observed in the inference condition than in the control condition. However, LSA failed to predict alterations in inference activation (i.e., augmentation or attenuation of activation) when subtle manipulations were applied to the same set of texts (Klin, Guzman, et al., 1999; Singer & Halldorson, 1996). Yeari and van den Broek argued that these failures are grounded in the basic limitations of LSA (Kintsch, 2001; Landauer & Dumais, 1997; Landauer et al., 1998; Perfetti, 1998). Specifically, LSA simulations treat the texts as “bags of words” in which word order and inter-relations between textual words are ignored (Jones & Mewhort, 2007). Moreover, all textual words had an equal weight in the computation of cosines between texts and probe items. However, according to reading comprehension models (e.g., Kintsch, 1998; van den Broek et al., 1999), textual words that appear closer to the probing point (i.e., still active in working memory) and distant words that are associated with probe-adjacent words (i.e., reactivated into working memory) should have a greater weight in the computation of text-probe semantic relations. Yeari and van den Broek suggested that integrating LSA computations within a dynamic comprehension model might overcome these shortcomings. The present study was designed to construct and test such an integrative model, exploring its theoretical and computational contribution to the field.

The present computational model: The landscape model–revised (LS-R)

The present computational model integrates the dynamic landscape model of reading comprehension (Tzeng, van den Broek, Kendeou, & Lee, 2005; van den Broek, 2010; van den Broek et al., 1996; van den Broek et al., 1999) with LSA representation of semantic knowledge (Landauer & Dumais, 1997; Landauer et al., 1998). This revised landscape model (LS-R model) computes fluctuations in the activation of text units and the interconnections established between them throughout reading (to download the model, see www.brainandeducationlab.nl/downloads). The text is initially segmented into text units (e.g., words or propositions) and reading cycles (e.g., clauses or sentences). The sizes of text units (e.g., words or propositions)

and reading cycles (e.g., clauses or sentences) are parameters determined by the researcher. Relevant criteria for such decisions can be length of the simulated text segments, desired resolution of simulation, experimental design of the simulated study, and so forth.

Activations and connections are updated with each reading cycle in the order that they appear within the text. *Activations* are updated as a function of three simulated mechanisms: (a) *attention*—text units of the current cycle are activated to the highest value; (b) *working memory*—text units from prior cycles carry residual activation (following a *decay rule*); and (c) *long-term memory*—text units from prior cycles are reactivated via connections with text units that are active in the current cycle. *Connections* between text units are established as a function of two simulated mechanisms: (a) *Episodic* connections are formed between text units that are coactivated (due to any activation mechanism) in the same reading cycle. These connections represent relations that readers create in specific texts or episodes (e.g., man and fly in superhero stories). The strength of episodic connections is a function of the activation levels of the interconnected text units, and it accumulates with each concurrent activation (following a logarithmic *learning rule*). (b) *Semantic* connections are formed between text units that are generally associated across various episodes. Semantic connections are computed by LSA between all text units before the beginning of the dynamic simulation. These connections represent the world knowledge readers possess before the reading of any specific text (e.g., man and walk). The strength of semantic connections (i.e., the cosine value) determines the baseline connection strength between each pair of text units, whereas episodic connections are added to and augment the baseline connection strength throughout the dynamic flow from one reading cycle to the next.

These simulated cognitive mechanisms are realized using the following mathematical equations (inspired by other computational models; Gluck & Bower, 1988; McClelland & Rumelhart, 1985; Myers & O’Brien, 1998). The activation of a text unit in a given cycle (A_{i_c}) is computed by the sum of the activations spread from m connected units. The activation spread from a connected unit j to a target unit i equals the multiplication of the activation of j from the previous cycle ($A_{j_{c-1}}$) with the connection strength between units i and j as computed in the previous cycle ($S_{ij_{c-1}}$) (see Eq. 1). The activations of the units in the current cycle are set to maximum, irrespective of the equation result. The *maximal activation value* is set by a parameter, and the *minimal activation value* is always 0. The parameter δ determines a *decay rate* of unit activation from one cycle to the next. The function σ ensures a positive logarithmic change in the connection strength (see Eq. 2).

$$A_{i_c} = \sum_{j=1}^m \delta \cdot A_{j_{c-1}} \cdot \sigma(S_{ij_{c-1}}) \quad (1)$$

$$\sigma(x) = \text{Tanh}[3(X-1)] + 1 \quad (2)$$

Note that units not yet mentioned in the text, but that have strong associations with units which have already been mentioned in the text, are activated via semantic connections and influence the activation of a current unit. This process resembles the generation of inferences and predictions during reading. Nonetheless, units of preceding cycles (i.e., already read) have a much stronger effect on the activation of a current unit than units in the following cycles (i.e., not yet read).

The strength of connections between each pair of units (S_{ij}) is first computed by LSA. Cosine computations are performed using the *term-to-term* procedure, available at the Latent Semantic Analysis website (<http://lsa.colorado.edu>). The text corpus used to construct the LSA Words \times Documents matrix was a corpus of “general reading up to the first year in college,” which was condensed to 300 dimensions with the singular value decomposition procedure. A *semantic strength coefficient*, which multiplies (or reduces, when it is smaller than 1) the strength of LSA connections, determines the relative contributions of the semantic connections to the computation of the overall connection strengths. The connection strength of each text unit in a given cycle (S_{ij_c}) is updated following the computation of the activation level of that text unit. Connection strength is accumulated from one cycle to the next (over the baseline LSA cosine) as a function of the activation levels of the connected units (see Eq. 3). The parameter λ determines the *learning rate*, with a high value of λ representing a higher rate of learning from previous textual information. Because λ or activation values cannot be smaller than 0, the connection strength necessarily is above 0, and changes are incremental.

$$S_{ij_c} = S_{ij_{c-1}} + \lambda \cdot A_{i_c} \cdot A_{j_c} \quad (3)$$

Finally, the LS-R model simulates a limited *working memory capacity* that defines the total amount of possible activation in each reading cycle. If the sum of activations exceeds this limit, activations are reduced proportionally to attain the limit (Just & Carpenter, 1992). This and the other parameters of the LS-R model are set to default values before the beginning of any simulation, but can be modified by researchers according to their research questions. The default values were chosen on the basis of theoretical grounds (e.g., working memory capacity is five times the maximal value of the text units; Kintsch, 1988), neutrality (e.g., λ is close to 1, the semantic strength coefficient equals 1), and computational constraints (e.g., δ is set to .1 to prevent text units from noncurrent cycles from reaching the maximal activation value).

The present study

In three sets of simulations, the LS-R model was applied to simulate the activation of (a) predictive inferences, (b) bridging inferences, and (c) centrality estimation and recall of

textual information. The first two sets of simulations of inference activation concerned the same data that had proven to be problematic in the simulations of the LSA-only model (Yeari & van den Broek, 2015). The third set of simulations was performed to extend the validity of the LS-R model to simulate offline measures of reading comprehension, such as centrality estimation and text recall.

To conduct the simulations, the original texts presented to human readers were segmented into words (Simulations I and II) or propositions (Simulation III) to determine the text units, and into sentences to determine the reading cycles. In some cases (fewer than 5 %), pairs of words were defined as single text units, when word pairs conveyed a different meaning than the one conveyed by each word separately (e.g., *break up, pass out, take off*).¹ Function words such as *to be* verbs (e.g., *is, was, been*), prepositions (e.g., *of, to, for*), and connectives (e.g., *because, therefore*) were excluded. Verbs were presented in their present simple form. All simulations used the same default values set by the model in the following parameters: (1) semantic strength coefficient = 1, (2) maximal activation = 1, (3) working memory capacity = 5, (4) learning rate = .9, and (5) activation decay rate = .1. The default values were used in the present simulations because the main goal of this research was to validate the LS-R model rather than to answer specific theoretical questions. At the end of the simulations of the entire set of texts of a given simulated behavioral experiment, we computed the average activation levels and/or connection strengths of all or specific units in the different conditions, depending on the theoretical question of the simulation. The means observed in the different conditions were compared statistically using within-items paired *t* tests. Then we compared the simulated data with the behavioral data. Successful simulation should capture the pattern of the behavioral results or correlate with the behavioral data. Minor differences between the methods of the different simulations are described before each simulation.

Simulation I: Activation of predictive inferences

The behavioral study—Klin, Guzman, et al. (1999)

The first set of simulations was conducted on data from a study by Klin, Guzman, et al. (1999). In three experiments, Klin, Guzman, et al. (1999) applied the probing procedure to examine the prevalence of activation of predictive inferences. Participants read 14 short narratives and immediately named

¹ Note that LSA is generally limited in capturing this type of integrated meaning (although, computationally, it treats single words and groups of words a bit differently). Yet, LSA captures to some extent the semantic relations between single words from word pairs and other single words that appear in the same text segments (e.g., the relation between “pass” and “fall” due to the use of “pass out” in texts).

(read aloud) single-word probes as quickly as possible. Two versions were prepared for each narrative: an “inference” version, in which the probe item represented a highly predictable consequence of the final action described in the preceding narrative, and a matched “control” version, which used many of the same words, but did not predict the consequence represented by the probe item (see Table 1). Predictability was verified by an independent group of readers. In the first experiment, Klin, Guzman, et al. found that naming latencies were significantly shorter for probes in the inference condition ($M = 441$ ms) than for probes in the control condition ($M = 455$ ms). This result was interpreted as an indication that predictive inferences are activated when they are highly predictable.

In a following experiment, Klin, Guzman, et al. (1999, Exp. 3) examined the activation of predictive inferences when additional, competing prediction could be derived from the events described in the inference-evoking text. For this purpose, Klin, Guzman, et al. added an introductory segment to the narratives used in the first experiment that suggested a second possible consequence of the final action. For example, in the sample narrative in Table 2, not only will the vase break when Steven throws it against the wall, but also his wife will probably leave him. In line with resonance theory (Myers O’Brien, 1998), Klin, Guzman, et al. predicted that the activation of the target inference (i.e., *break*) would be attenuated under this condition, because activation is divided between the two possible consequences of the final action. As expected, no evidence was found for the activation of the target predictive inferences, as indicated by the equal naming latencies found for probes in the inference ($M = 477$ ms) and control ($M = 477$ ms) conditions. Given this result, Klin, Guzman, et al. concluded that a predictive inference is more probable when it is the only possible consequence of the inference-evoking events (see also Fincher-Kiefer, 1996).

In a final experiment, Klin, Guzman, et al. (1999, Exp. 4) verified that the deactivation observed for predictive inferences in the previous experiment was not a result of lengthy

Table 1 A text-probe example from the first experiment in Klin, Guzman, et al. (1999)

Introduction

Today Steven was angry with his wife because she had left a mess in the kitchen.

Inference version

He tried to cool down, but felt his resentment building. No longer able to control his anger, he threw a delicate porcelain vase against the wall.

Control version

He reacted by acting cool toward her. He then apologized for getting angry, and offered to clean her delicate porcelain vase to make up for it.

Probe

BREAK

Table 2 A text-probe example from the third experiment in Klin, Guzman, et al. (1999)

Introduction

After years of abuse, Susan had enough. She joined a support group for battered women and told her husband, Steven, that she was going to leave him if there was even the mildest violent incident in the house. Steven was taking her seriously and had started counseling. He had managed to control his temper for the past month. He could not bear the thought of her leaving. He felt his life would be over if she and the children left. Today Steven was angry with Susan because she had left a mess in the kitchen.

Inference version

He tried to cool down, but felt his resentment building. No longer able to control his anger, he threw a delicate porcelain vase against the wall.

Control version

He tried to cool down using all of his skills. Controlling his anger, he offered to clean her delicate porcelain vase to make up for it.

Probe

BREAK

narratives, the bulk of which (i.e., the introduction) were semantically unrelated to the target inferences. Thus, the narratives in this experiment included an introductory segment that was unrelated to the target inferences, but that also did not suggest any alternative, competing prediction to the target inferences (see Table 3). Using this neutral introductory segment, the activation of the predictive inferences reappeared, as demonstrated by significantly faster naming of the probes in the inference condition ($M = 451$ ms) than in the control condition ($M = 461$ ms). Klin, Guzman, et al. concluded that the competing predictions, rather than the additional neutral events, were responsible for the elimination of predictive inference activation in the previous experiment.

Table 3 A text-probe example from the fourth experiment in Klin, Guzman, et al. (1999)

Introduction

Steven and Susan had been married for twenty years. They had met their senior year in high school and had gotten married when they were twenty one. Steven had just started a new job as the assistant manager of the accounting department at Sears. He was excited about the promotion. It meant a large raise and a lot of extra responsibilities. But it also meant long hours and a lot of extra stress. He and Susan were having a hard time adjusting their life to fit his schedule. Today Steven was angry at Susan because she had left a mess in the kitchen.

Inference version

He tried to cool down, but felt his resentment building. No longer able to control his anger, he threw a delicate porcelain vase against the wall.

Control version

He reacted by acting cool toward her. He then apologized for getting angry, and offered to clean her delicate porcelain vase to make up for it.

Probe

BREAK

The computational simulation

To simulate the findings of Klin, Guzman, et al. (1999), we ran the LS-R simulation on the inference and control versions of each of the 14 narratives in each of the three experiments. The narratives were segmented into words as text units, and sentences as reading cycles. The probe items were added as additional text units at the ends of the relevant narratives. A simulation was concluded following the last sentence of the text, before the “reading” of the probe item. The activation levels reached by the probe items at the end of each simulation functioned as the dependent measure.

The results of the simulations captured the pattern of the behavioral findings of Klin, Guzman, et al. (1999) (see Table 4). The simulations of the narratives used in the first experiment yielded a significantly higher activation level for probe items that followed the inference versions ($M = 0.056$) than for probes items that followed the control versions ($M = 0.049$), $t(13) = 2.6$, $SE = .003$, $p < .05$, Cohen’s $d = 1.44$. This difference in activation levels was attenuated ($M_{\text{inference}} = .071$, $M_{\text{control}} = .069$) and did not reach significance when simulations were conducted on the narratives used in the third experiment, $t(13) = 1.5$, $SE = .002$, $p = .15$, Cohen’s $d = 0.83$. In simulating the narratives used in the fourth experiment, the difference in activation levels between the probe items that followed the inference ($M = .070$) and control ($M = .066$) versions reappeared, $t(13) = 2.5$, $SE = .002$, $p < .05$, Cohen’s $d = 1.38$.

Although Klin, Guzman, et al. (1999) did not exhibit a significant interaction between the effects observed in the different experiments, we conducted a two-way analysis of variance (ANOVA) with Experiment and Text Version as within-items factors, to ensure that the changes found in the three effects were significant. One text was omitted from this and subsequent analyses that compared the effects in the different experiments. The omitted text was the only one that differed across the three experiments (e.g., the probe item was “sting” in the first experiment, “hit” in the third, and “burn” in the fourth). Because our analyses are based on within-items comparisons, we could not compare computational results yielded by items that differed in content across

the experiments. The ANOVA yielded a significant interaction, $F(2, 24) = 3.62$, $MSE = .0002$, $p < .05$, $\eta_p^2 = .23$. Post-hoc comparisons revealed that the effect found in the simulation of the third experiment was significantly smaller than the effect found in the first experiment, $F(2, 24) = 5.1$, $MSE = .0003$, $p < .05$, $\eta_p^2 = .30$, and also significantly smaller than the effect found in the simulation of the fourth experiment, $F(2, 24) = 4.6$, $MSE = .0003$, $p < .05$, $\eta_p^2 = .28$. These computational results simulate and support the pattern of results found with human readers.

These successful simulations bear two important implications—one theoretical and one computational. From a *theoretical* perspective, the present simulations suggest that the activation of predictive inferences was not necessarily attenuated due to the activation of competing predictions (Klin, Guzman, et al. 1999, Exp. 3). The simulations captured the pattern of the behavioral findings even though the activation of competing predictions was not simulated. Analyzing the computational data revealed an alternative explanation for the attenuation of predictive inferencing. We found that the reactivation of the “competing” introductory segments themselves competed with and attenuated the activation of the predictive inferences in the simulation of the third experiment. The closing inference-evoking text segments established stronger conceptual connections with the competing introductory segments in the simulation of the third experiment ($M = .215$) than with the neutral introductory segments in the simulation of the fourth experiment ($M = .198$), $t(12) = 3.1$, $SE = .005$, $p < .01$, Cohen’s $d = 1.79$. In fact, Klin, Guzman, et al. intentionally designed the texts that way (i.e., the competing introduction evoked a competing prediction via its relation to the closing sentences; see Table 2), and our simulations confirmed this design computationally. As a result of these connections, the competing introductory segments in the simulation of the third experiment were reactivated by the closing inference-evoking text segments ($M = .388$) in a marginally stronger manner than were the neutral introductory segments in the simulation of the fourth experiment ($M = .363$), $t(12) = 1.5$, $SE = .002$, $p = .15$, Cohen’s $d = 0.87$. This difference reached significance when one text, which exhibited a strong reverse effect and, correspondingly, did

Table 4 Means of the response times (RTs), activation levels, and text-probe cosines obtained in the inference and control conditions in each of the three behavioral experiments in Klin, Guzman, et al. (1999) and their LS-R and LSA-only (Yeari & van den Broek, 2015) computational simulations

	Behavioral RTs (ms)			LS-R (Activation)			LSA (Cosine)		
	Inference	Control	Difference	Inference	Control	Difference	Inference	Control	Difference
Exp. 1	441	455	−14*	.056	.049	.007*	.116	.069	.047*
Exp. 3	477	477	0	.071	.069	.002	.086	.074	.012
Exp. 4	451	461	−10*	.070	.066	.004*	.055	.044	.011

* Significant differences, $p < .05$

not exhibit activation of the predictive inference in any of the simulations, was excluded from the analysis, $t(11) = 2.5$, $SE = .001$, $p < .01$, Cohen's $d = 1.50$. Taken together, these results suggest that under a limited activation capacity, the stronger reactivation of the competing introductory segments competed with and attenuated the activation of the predictive inferences in a stronger manner than the neutral introductory segments, which were reactivated to a lesser extent.

Further support for this kind of competition between the activation of predictive inferences and the reactivation of the introductory segments was revealed in manipulating the total amount of activation allowed in each reading cycle (i.e., working memory capacity parameter). When doubling the total amount of activation allowed in the basic simulations, the difference between the activations of predictive probes following the inference and control texts used in the third experiment (comprised the competing introduction) increased ($M_{\text{inference}} = .138$, $M_{\text{control}} = .133$) and became marginally significant, $t(13) = 1.8$, $SE = .009$, $p = .09$, Cohen's $d = 1.02$. Moreover, this difference obtained in the simulation of the third experiment was not significantly different from the difference obtained between the inference and control texts used in the fourth experiment (including a neutral introduction), $F(2, 24) = 2.8$, $MSE = .008$, $p = .12$, $\eta_p^2 = .19$. When we tripled the total amount of activation allowed in the simulation of the third experiment, the activation of predictive probes became significantly higher following the inference texts ($M = .198$) than following the control texts ($M = .192$), $t(13) = 2.5$, $SE = .002$, $p < .05$, Cohen's $d = 1.39$. These findings suggest that the fact that predictive inferences were not activated following the texts comprising the competing introduction (Exp. 3) was caused by the competition of the inferential and textual information for a limited amount of activation. This explanation actually coincides with Klin, Guzman, et al. (1999) original concern that the addition of introductory segments that were conceptually unrelated to the target predictive inferences might be sufficient to attenuate their activation. The present simulations support this concern (i.e., inference activation was attenuated even without the simulation of competing inference activation) and clarify why these semantically less-related introductory segments attenuated the activation of predictive inferences in one experiment (Exp. 3) and did not in another experiment (Exp. 4). Further studies employing behavioral methods will be needed to confront the alternative explanations.

From a *computational* perspective, these simulations illustrate the value added by the dynamic integrated LS-R model to simulations conducted using the LSA algorithm alone (Yeari & van den Broek, 2015). LSA simulations successfully captured the activation of predictive inferences in the first experiment of Klin, Guzman, et al. (1999) and the attenuation of this activation in their third experiment (see Table 4). By computing the strength of the semantic relations (cosines)

between the probe items and the inference and control versions of the preceding texts, LSA yielded significantly higher cosines in the inference than in the control condition in the simulation of the first experiment's texts, but not in the simulation of the third experiment's texts. These LSA simulations demonstrated the role of textual semantic constraints in the activation and deactivation of predictive inferences. However, the failure to simulate the reappearance of predictive inference activation in the simulation of the fourth experiment (see Table 4) demonstrated that textual semantic constraints are not the whole story. LSA-only computations could capture the reduction in overall semantic relatedness between the predictive inferences and the texts when they comprised unrelated introductory segments in the third and fourth experiments. However, it could not capture the differences in the within-text "episodic" relations between the introductory segments and the closing sentences and the difference in the influences of the unrelated introductory segments on the activation of predictive inferences in the two experiments. The present LS-R model, on the contrary, captured these differences by simulating dynamic fluctuations in the activations of different parts of the texts due to within-text relations (i.e., reactivation) and limited activation capacity (i.e., activation decay and competition).

Simulation II: Activation of bridging inferences

The behavioral study—Singer and Halldorson (1996)

The second set of simulations was performed on data from a study by Singer and Halldorson (1996). This study examined the activation of bridging inferences—mediating information that specifies the conceptual relations between textual ideas. Participants read 16 two-sentence passages and then answered, as quickly as possible, a yes–no question that probed bridging information (see Table 5). In the inference version, the queried information specified the conceptual link between the two sentences, whereas in the control version, the queried information did not resolve the conceptual gap between the

Table 5 A text-probe example from the study of Singer and Halldorson (1996)

Inference version

Sharon decided to buy the sports car.

Control version

Sharon decided to sell the sports car.

Second sentence

She went straight to the bank.

Question probe

Do banks lend money?

two sentences. A faster answer to the question probe in the inference than in the control condition was taken as an indication for the activation of the bridging inference queried by the question.

In addition to the reading of the two-sentence passages, Singer and Halldorson (1996) asked a different group of participants to read only the first sentences of the passages and to answer the same questions. This control group was added to confirm the authors' hypothesis, according to which bridging inferences are not activated via autonomous semantic priming processes, but rather via controlled reasoning-like processes in which the readers resolve an incomplete syllogism (i.e., the first sentence serves as a minor premise, the second sentence serves as a conclusion, and the bridging idea is a missing major premise). In the case that inference activation is observed only in the two-sentence condition but not in the one-sentence condition, inference activation cannot be attributed to semantic priming, because the semantic differences between the inference and control versions are comparable in the one-sentence and two-sentence conditions (i.e., the differences are embedded only in the first sentences, whereas the second sentences are identical in the two versions; see Table 5).

As expected, Singer and Halldorson (1996, Exp. 1) found a significantly faster answering time to questions in the inference condition ($M = 1,981$ ms) than to questions in the control condition ($M = 2,251$ ms) for the two-sentence passages, but not for the one-sentence passages ($M_{\text{inference}} = 2,011$ ms, $M_{\text{control}} = 1,998$ ms). They concluded that bridging inferences are activated due to a strategic objective to resolve coherence gaps, rather than due to an autonomous semantic priming by textual words.

The computational simulation

The computational simulations of Singer and Halldorson's (1996) study were performed for all 16 passages in a manner similar to the simulations conducted for the study by Klin, Guzman, et al. (1999). In the present simulations, the words of the question probes were added as additional units, and the dependent measure was the mean activation level of the question words at the end of the simulation.

Consistent with the behavioral findings, the mean activation level of the question words that followed the two-sentence inference passages ($M = .111$) was higher than that of the question words that followed the two-sentence control passages ($M = .104$), $t(15) = 3.6$, $SE = .002$, $p < .01$, Cohen's $d = 1.86$ (see Table 6). This difference between the inference and control conditions was significantly attenuated in the one-sentence passages ($M_{\text{inference}} = .045$, $M_{\text{control}} = .040$), as indicated by the significant two-way interaction of Passage Type (inference, control) \times Passage Length (two-sentence, one-sentence), $F(1, 15) = 4.7$, $MSE = .00017$, $p < .05$, $\eta_p^2 = .24$.

Nevertheless, in contrast to the behavioral findings, the difference between the mean activation levels of the question words that followed the one-sentence inference and control passages also reached significance, $t(15) = 3.9$, $SE = .001$, $p < .01$, Cohen's $d = 2.01$.

As we noted previously, these successful simulations bear two important implications—one theoretical and one computational. From a *theoretical* perspective, these simulations suggest that autonomous spreading activation processes can account for the activation of bridging inferences (for a similar account, see the “convergence and constraints satisfaction” rule in Graesser et al., 1994). The present simulations captured the attenuation found for the activation of bridging inferences following the one-sentence passages, even though we did not simulate any strategic, controlled processes (e.g., by applying special production rules; see Graesser et al., 1994). Analyzing the activations and connections developed during the simulations shed some light on the underlying mechanisms that intensified the activation of bridging inferences in the two-sentence relative to the one-sentence passages, even though the additional second sentences were identical in the inference and control versions. We found that the activation of the second sentences in the inference passages ($M = .637$) that exhibited higher probe activation (i.e., those passages that yielded higher probe activation in the inference than in the control passages) was significantly higher than in the control passages ($M = .632$), $t(12) = 2.5$, $SE = .002$, $p < .05$, Cohen's $d = 1.44$. This significant difference in the activations of the second sentences was largely due to stronger connections between the two sentences in the inference passages ($M = .231$) than between those in the control passages ($M = .218$), $t(12) = 1.9$, $SE = .028$, $p = .09$, Cohen's $d = 1.10$. These results suggest that the activation spread between and from the two sentences in the inference passages, which were internally more coherent and externally more related to the target inferences than the control passages (e.g., “buying a car” and “going to a bank” are internally more coherent and externally more related to “lending money” than are “selling a car” and “going to a bank”), was sufficiently high to activate the bridging inferences in the two-sentence passages.

In addition, our simulations predicted a weaker but significant activation of predictive inferences following the one-sentence inference passages (i.e., “lending money” is a possible consequent action of a “buying a car” goal), in contrast to the behavioral findings. This result may suggest that strategic reasoning-like processes are also involved in the activation of predictive and bridging inferences, as was argued by Singer and Halldorson (1996). Alternatively, it may show that small differences in activations, below a certain level, cannot be captured by response time measures.

From a *computational* perspective, the present simulations stressed the contribution of the dynamic LS-R model to the computations of semantic relatedness made by LSA (Yeari &

Table 6 Means of the response times (RTs), activation levels, and text-probe cosines obtained in the inference and control conditions of two-sentence and one-sentence passages in the behavioral experiment by

Singer and Halldorson (1996) and the LS-R and LSA-only (Yeari & van den Broek, 2015) computational simulations

Passage Type	Behavioral RTs (ms)			LS-R (Activation)			LSA (Cosine)		
	Inference	Control	Difference	Inference	Control	Difference	Inference	Control	Difference
Two-sentence	1,981	2,251	−270*	.111	.104	.007*	.262	.234	.028*
One-sentence	2,011	1,998	13	.045	.040	.005*	.171	.122	.049*
Difference			−283*			.002*			−.021*

* Significant differences, $p < .05$

van den Broek, 2015). LSA-only was successful in predicting the activation of bridging inferences following the two-sentence passages. The cosine of the question probes with the two-sentence inference passages was significantly higher than that obtained with the two-sentence control passages (see Table 6). However, in contrast to the present computational findings and the behavioral findings of Singer and Halldorson (1996), the difference yielded by the LSA-only computations between the cosines of the one-sentence inference passages and the one-sentence control passages was even stronger than the difference observed between the two-sentence inference and control passages (see Table 6). Because semantic differences occurred only in the first sentences, the addition of second identical sentences attenuated the overall semantic difference (cosines) between the inference and control passages. The present LS-R simulations demonstrated the role played by within-text episodic relations and within-text spread of activation in the generation of extratext information from semantic knowledge. They showed that the same idea (i.e., the second sentence) could affect the activation of a semantically related idea (i.e., the bridging inference) to a different extent as a function of its semantic relation with preceding ideas (i.e., the first sentence) and the resulting level of its activation. Further computational investigations, which will examine the possible implementation of activation thresholds, will be required to minimize the gap between the present simulations and the behavioral findings.

In considering supplementary computational investigations, we were interested to examine whether the original, semimanual landscape model (Tzeng et al., 2005) would be more (or less) successful in simulating Singer and Halldorson's (1996) findings. In this version of the landscape model, researchers decide which text units are semantically connected, and by which type of semantic connections (e.g., referential, causal, enablement, associative, or any other types of connections that the researcher may add to the model). Each type of connection receives a connection strength value (predefined by the model's parameters) that is constant across all connections of the same type (see further details in Tzeng et al., 2005).

In simulating the passages from Singer and Halldorson (1996), we coded causal relations between the text units of the inference passages (e.g., “go to a bank” is a consequent action of the motivation “buy”; see Table 5), and associative relations between the corresponding text units of the control passages (“bank” is associated to some extent with “sell”). In addition, we defined enablement relations (i.e., X is a possible but not a necessary outcome of Y) between the probed question units and text units from the first sentence of the inference passages (e.g., “buy a car” enables the action “lend money”), and associative relations between the inference units and text units from the first sentence of the control passages (e.g., “sell” is associated with “money”). Finally, we defined enablement relations between the probed question units and text units from the (identical) second sentences of the inference and control passages (e.g., “go to a bank” enables the action “lend money”). On theoretical grounds (see Trabasso, van den Broek, & Suh, 1989; van den Broek, 1994; van den Broek et al., 1999), the landscape model assigns a higher value to activations resulting from causal connections ($A_i = .6$) than to those from enablement connections ($A_i = .3$), and a higher value to activations resulting from enablement than from associative connections ($A_i = .2$).

Similar to the LS-R simulations, the landscape simulations yielded significantly higher activation values for the probed question units following the inference passages than for those following the control passages, in both the two-sentence condition ($M_{\text{inference}} = .059$, $M_{\text{control}} = .054$), $t(15) = 4.5$, $SE = .001$, $p < .01$, Cohen's $d = 2.32$, and the one-sentence condition ($M_{\text{inference}} = .022$, $M_{\text{control}} = .018$), $t(15) = 4.0$, $SE = .001$, $p < .01$, Cohen's $d = 2.06$. However, in contrast to the LS-R simulations, the difference between the two-sentence and one-sentence conditions was not significant, $F(1, 15) = 2.5$, $MSE = .00004$, $p = .14$, $\eta_p^2 = .14$. Thus, the LS-R model was more successful than the original landscape model in capturing the attenuation observed by Singer and Halldorson (1996) in the activation of bridging inferences under the condition that only one sentence was presented, and no bridging was required. These findings illuminate the advantage of the LS-R coding, which estimates connection strengths over a continuous scale,

over the original landscape model coding, which estimates different types of connections in an all-or-none fashion (e.g., information units are causally connected or not connected, are associated or not associated). Moreover, the manual, researcher-based landscape coding was more recurrent across the various texts, and corresponded more closely to the textual semantic structure originally designed by Singer and Halldorson. The diverse, objective coding applied by the LS-R model eventually resulted in a more successful simulation of Singer and Halldorson's behavioral findings than did the original landscape model.

Simulation III: Centrality perception and text recall

The two previous sets of simulations simulated online, transient processes of reading comprehension (i.e., the target inferences were probed immediately after the reading of the text segments). The present set of simulations was designed to simulate the offline, enduring products of reading comprehension. Previous studies had demonstrated successful computational simulations of offline comprehension products, such as text recall, using hand-coded manual representations of semantic knowledge (e.g., the construction–integration model, Singer & Kintsch, 2001; or the landscape model, van den Broek et al., 1996). In the present simulations, we examined whether the LS-R model, which relies on LSA representations of semantic knowledge and computes the dynamics of online processes, can simulate the offline performances of human readers. Building on theories of reading comprehension (Trabasso & Sperry, 1985; van den Broek, 1988), we expected that the “memory” connections established between textual units throughout the online simulation of the LS-R model would be associated with offline products of reading comprehension.

The behavioral study—Yeari, Oudega, and van den Broek (in press)

For the present simulations, we used data from a recent study conducted in our lab (Yeari et al., in press). In this study, Yeari et al. examined (among others things) the recall of textual ideas as a function of their centrality. *Centrality* refers to the importance of textual ideas to the overall comprehension of the text. Nine expository texts were initially parsed into information units, each comprising a central predicate, its arguments (including time and place), and the adjectives or adverbs of these arguments (see the Appendix). Three trained judges evaluated the centrality level of each information unit on a scale from 1 (*least central*) to 5 (*most central*). Centrality was defined as a function of two criteria: (a) the extent to which an information unit is important for the overall understanding of the text, and (b) the extent to which

comprehension would be impaired if the information unit were to be missing (e.g., Albrecht & O'Brien, 1991; Miller & Keenan, 2009; van den Broek, 1988; see the Appendix). Adult participants read the nine texts in three blocks. Following the reading of each block of three texts, they freely recalled the content of the texts in the order they read them. Yeari et al. found that the text units that received the highest centrality scores ($M_{\text{centrality}} = 4.6$) were recalled better ($M_{\text{recall}} = .32$) than the text units that received the lowest centrality scores ($M_{\text{centrality}} = 1.9$, $M_{\text{recall}} = .22$). The authors explained that these findings are consistent with the causal network model (e.g., Trabasso & Sperry, 1985; van den Broek, 1988), which holds that central ideas are recalled better than peripheral ideas because they establish more conceptual connections with other textual ideas in long-term memory during reading. In this account, the connected ideas serve as retrieval cues after reading.

The computational simulation

To simulate these findings, we ran the LS-R simulation on each of the nine expository texts. The simulation procedure was similar to the previous simulations, except that the information units parsed in the behavioral study (Yeari et al., in press) served as the text units in the computational simulations. Sentences defined the reading cycles, as in the previous simulations. At the end of the simulations, we computed for each text unit in each text the sum of the connection strengths it established with the other text units in the same text, as well as the sum of the activation levels that the text unit reached across the reading cycles. We used a summation rather than an average of the simulated connection strengths and activation levels of the texts units, because the numbers of connections and of activated cycles, and not just their averages, play a role in comprehension outcomes such as text recall and centrality estimation (van den Broek, 1988).

Using a one-way ANOVA with text as a random effect, we compared the mean connection strengths and the mean activation levels of the central and peripheral units. We found that the connections established by central units with other units ($M = 17.0$) were significantly stronger than the connections established by peripheral units with other units ($M = 14.1$), $F(1, 103) = 15.8$, $MSE = 16.742$, $p < .01$, $\eta_p^2 = .65$ (see Table 7). Similarly, the activation levels that central units reached during the simulations ($M = 5.5$) were significantly higher than the activation levels the peripheral units reached ($M = 4.6$), $t(1, 103) = 24.5$, $MSE = 1.127$, $p < .01$, $\eta_p^2 = .74$ (see Table 7). A two-way ANOVA with Mechanism (connections, activations) as a within-items factor and Information Centrality (central, peripheral) and Text as between-items factors revealed that the difference between the connection strengths of the central and peripheral units was significantly greater than the difference between the activation levels of the

Table 7 Means of the human centrality scores and human recall proportions, along with computed sums of connection strengths and activation levels, obtained for the central and peripheral information units in the behavioral experiment by Yeari et al. (in press) and the LS-R computational simulations

	Behavioral		LS-R	
	Centrality Score (1–5)	Recall (Proportions)	Connections Sum	Activations Sum
Central units	4.6	.32	17.0	5.5
Peripheral units	1.9	.22	14.1	4.6

two types of units, $F(1, 103) = 9.5$, $MSE = 6.372$, $p < .01$, $\eta_p^2 = .08$. In all ANOVAs, the interactions with text were not significant ($F_s < 1$).

To confirm these computational findings using a stricter test, we computed correlations of the simulated connection strengths and activation levels of the text units with the recall proportions and centrality estimations obtained from human readers (Yeari et al., in press). Estimations of the centrality levels of information units were varied over 13 different scores obtained in the rating procedure (from 1 to 5, in increments of 0.33; see Fig. 1a). Therefore, we computed means of the connection strengths, activation levels, and recall proportions for each group of information units that obtained the same centrality score. We found that connection strength was positively associated with both centrality estimations ($r_s = .79$, $p < .01$; Fig. 1a) and recall proportions ($r_s = .70$, $p < .01$; Fig. 1b). The associations of activation levels with the centrality estimations ($r_s = .63$, $p < .05$; Fig. 1a) and recall proportions ($r_s = .46$, $p = .11$; Fig. 1b) were somewhat weaker, with only the former being significant. These findings further support the LS-R model by demonstrating its ability to predict offline products on the basis of the simulation of online processes, such as the activations and interconnections of text units.

These findings are consistent with the causal network model (e.g., Trabasso & Sperry, 1985; van den Broek, 1988). They demonstrate that textual information centrality is a function of the number and strength of connections that a textual information unit establishes with other units during reading. Moreover, the number and strength of interconnections predict the likelihood that a textual information unit will be retrieved from long-term memory after reading. We also found that the extent that a textual unit is active during reading also predicts its centrality estimation and retrieval, although to a lesser extent than its interconnection strength. These findings are somewhat consistent with the construction–integration model (Kintsch, 1988, 1998), which holds that central ideas have a higher probability of being stored in long-term memory, because they remain active during reading for a longer period than peripheral ideas. Finally, from a

computational perspective, the present simulations demonstrate that the LS-R model is capable of simulating offline comprehension products, such as centrality estimation and text recall, in addition to the simulation of online text processing, such as inference generation.

General discussion

In the present research, we introduced a new configuration of the computational modeling of reading comprehension, which integrates the landscape simulation of dynamic reading processes (van den Broek, 2010) with LSA representations of semantic knowledge (Landauer & Dumais, 1997; see Millis & Larson, 2008, for an integration of LSA with the construction–integration model to predict aesthetic responses, such as the enjoyment and comprehension of artworks). In three sets of simulations, the integrated LS-R model was found useful in simulating both the online processes and offline products of human reading comprehension. Specifically, the LS-R model successfully simulated the activation of predictive (Simulation I) and bridging inferences (Simulation II) during reading, and the centrality estimation and recall of textual information after reading (Simulation III), as they were observed with human readers in previous behavioral studies (Klin, Guzman, et al. 1999; Singer & Halldorson, 1996; Yeari et al., in press). The LS-R model appears to be a good candidate to extend existing computational models, which have used manual, researcher-based coding of a limited, selected subset of semantic knowledge (see Frank et al., 2008). The application of a computational representation of semantic knowledge allowed us to simulate the reading of complete sets of texts that were originally presented to human readers. Moreover, semantic connections could be simulated on a continuous scale of connection strengths. These benefits resulted in a more reliable simulation procedure of behavioral findings.

The first set of simulations captured the pattern of the findings obtained by Klin, Guzman, et al. (1999) regarding the activation and deactivation of predictive inferences under different textual constraints. Using a probe-naming procedure, Klin, Guzman, et al. demonstrated that predictive inferences are activated when they are strongly suggested by the preceding text (Exp. 1), attenuated when the majority of the text suggests a competing prediction (Exp. 3), and reappear when the majority of the text is neutral with regard to the target predictive inferences (Exp. 4). Consistent with these results, the present computational simulations yielded higher activations for probe items that followed the inference (high-predictability) texts (simulation of first experiment), no significant difference in the activations of probe items that followed the inference and control texts when the texts suggested competing predictions (simulation of third experiment), and higher activations for probe items that followed the inference texts

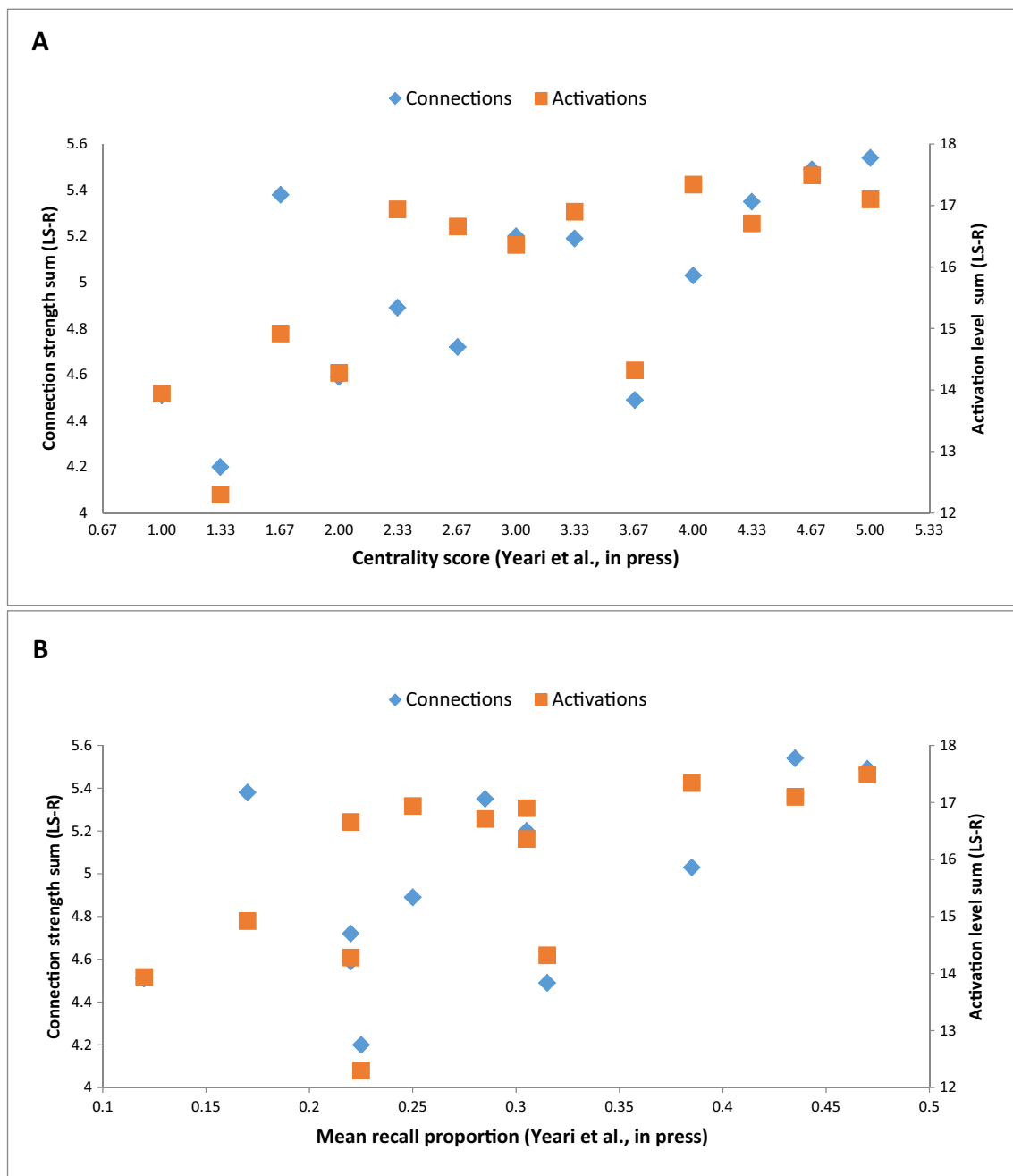


Fig. 1 Scatterplots showing the relation between the centrality scores (a) and recall proportions (b) obtained by humans in Yeari et al. (in press) and the connection strengths and activation levels computed in the present simulations

when the majority of the text was neutral with regard to the probe items (simulation of fourth experiment). Finer analyses of the computational results suggested that predictive inferences are more probable when a sufficient amount of activation is spread from conceptually related textual ideas (see the resonance theory: Myers & O'Brien, 1998). Specifically, in the simulation of the first experiment, more activation was spread from the inference text ideas than from the control text ideas to the probe items. In the simulation of the third experiment, the reactivation of the unrelated introductory segments

(and the possible competing predictions they activated, which were not simulated in the present simulations), prior to the probing of the predictive inferences, competed for their activation and attenuated it. In the simulation of the fourth experiment, the unrelated introductory segments were reactivated to a lesser extent prior to the probing point, and thus interfered less with the activation of the predictive inferences.

The second set of simulations partially captured the pattern of the findings of Singer and Halldorson (1996), who used a question-probing procedure to examine the activation of

bridging inferences. Similar to Singer and Halldorson, we found higher activations for probe items that conceptually linked (i.e., bridged) the preceding two-sentence inference passages, as compared to the control passages. Moreover, these activations were attenuated, although they remained active (in contrast to Singer and Halldorson's findings), when the passages included only one sentence instead of two, and the probe items represented predictive rather than bridging inferences. Analyses of the computational results revealed that the second sentences in the inference passages were more strongly connected to the first sentences (which differed in the inference and control passages), and were thus more active than the second sentences in the control passages. Apparently, the amount of activation spread from the two-sentence inference passages was sufficiently high to be detected by the answering-time procedure in the study by Singer and Halldorson.

The third set of simulations captured the pattern of the findings that Yeari et al. (*in press*) exhibited regarding human centrality estimations and free recall of textual information. Simulations yielded higher activations and stronger interconnections for the central information units, which were recalled better than the peripheral information units. Moreover, positive correlations were found between human centrality scores and recall proportions, on the one hand, and computational activation levels and connection strengths, on the other, with stronger correlations for the connection strengths than for the activation levels. The computational findings are consistent with well-known theoretical models in reading comprehension (e.g., Trabasso & Sperry, 1985; van den Broek, 1988), which hold that the centrality level and retrieval probability of textual information is a function of the number and strength of conceptual connections that an information unit establishes with other units in the text during reading. The more extensive and stronger the interconnections of a given textual unit, the higher the probability that that unit will be perceived as central information and retrieved after reading.

It is worth noting that the present successful simulations relied on significant differences between experimental conditions that were quite small in magnitude (e.g., differences of .002 significantly distinguished between the simulations of Exps. 3 and 4 of Klin, Guzman, et al., 1999, and the simulations of the one-sentence and two-sentence conditions of Singer & Halldorson, 1996). These small but significant differences indicate that the within-item variances were also small, presumably due to the following causes: (a) minor differences (one word, in some cases) between the inference and control versions of the simulated texts, and (b) successful simulations that captured those minor differences in the same direction they had been originally designed by the researchers to affect human readers. When estimating the magnitude of the differences observed between conditions in the proportions to the mean values (i.e., activation levels) observed, we found that, proportionally, the computational differences (around 3 % in the different

simulations) were not much smaller than the response time differences observed in the behavioral studies (around 3 % and 10 % in Klin, Murray, et al., 1999, and Singer & Halldorson, 1996, respectively). Thus, considering these proportions and the effect sizes obtained in the present analyses, we believe that the present findings are reliable and support the validation of the LS-R model.

To summarize, the present LS-R model successfully simulated previous behavioral findings in six out of seven different simulations. Both online reading processes, such as inference generation, and offline performances of reading comprehension, such as centrality estimation and free recall, were simulated. In contrast to previous simulations, the present ones were conducted over the entire set of original texts used in the behavioral studies. The analyses of these simulations contributed theoretical insights regarding the mechanisms underlying simulated comprehension phenomena. Particularly, they demonstrated the role played by other simulated mechanisms, such as within-text interconnections, spreading (re)activation, and activation competition, along with the central role played by (LSA) semantic connections (see Yeari & van den Broek, 2015), in simulating comprehension processes and products. More practically, the LS-R model might be applied as a text analyzer that assesses predictability, accessibility to inferential information, coherency, readability, and so on. By manipulating the model parameters (e.g., working memory capacity, learning rate), the LS-R model might also be applied to evaluate the appropriateness of a text to a reader.

Limitations and future directions

Beyond the implications of the present findings for the theoretical, computational, and practical aspects of reading comprehension, the computational implementation of LS-R raises several questions that need to be addressed in the future. First, the LS-R model has largely built on the LSA computations of semantic relations between individual text units, particularly in simulating inference generation from semantic knowledge. Although LSA has proved to be an adequate simulator of inference activation in numerous simulations (Yeari & van den Broek, 2015), it is limited in capturing semantic meanings that are derived from word order and grammatical (e.g., subject vs. object, passive vs. active) and logical (e.g., negation, comparison) variations (Jones & Mewhort, 2007; Kuperberg, Paczynski, & Ditman, 2011; Landauer et al., 1998; Perfetti, 1998; Wolfe, Magliano, & Larsen, 2005). The LS-R model has partially compensated for this weakness by forming episodic relations between text units in a sequential manner (i.e., the computation of activations and relations were made for each text unit in the order they were presented within cycles and between cycles). Yet the success of the present simulations can mostly be attributed to the relatively large samples of sentences and texts in the different simulations. These samples

allowed the model to overcome the grammatical and logical variations that add unique interpretations that could not be captured by the semantic content of the individual text units alone. A future development of the LS-R model might adopt a computational procedure that arranges the text units on the basis of their grammatical roles (in contrast to their original order in the sentences) and gives precedence to important roles such as predicates (see the predication model: Kintsch, 2000, 2001).

The sequential nature of the LS-R model enabled the simulation of word order, which LSA completely ignores. However, this sequential aspect also prevented the model from considering the act of rereading of prior text segments that naturally occurs during reading (Yeari, van den Broek, & Oudega, 2015). This limitation characterizes many of the computational models, because rereading takes place due to multiple reasons that are difficult to classify and predict. It also characterizes many of the behavioral studies that have displayed texts sequentially, one text unit (e.g., word, clause) at a time, to attain greater control of the experimental design, or to allow thinking aloud while reading (e.g., Linderholm & van den Broek, 2002). Although the present simulations have overcome the possible effect of rereading on comprehension processes and products, the simulation of rereading should be considered an important challenge for future developments of this and other computational models.

A further development that should be considered in future simulations is the modeling of individual differences (e.g., reading skill, knowledge, and motivation). The findings of the present simulations were based only on within- or between-item differences, comparing the simulations of similar texts in the different conditions and experiments. The simulations actually represented an average reader whose reading processes and products were influenced by the semantic content and structure of the texts. Simulating behavioral findings that have been found in a large sample of typically developed readers enabled the LS-R model to simulate successfully the influence of texts on the average reader. However, readers' characteristics should not be ignored in future simulations, given that comprehension processes and products are the result of the interaction between reader-based and text-based factors (e.g., Ozuru, Dempsey, & McNamara, 2009; van den Broek, Bohn-Gettler, Kendeou, Carlson, & White, 2011).

Finally, the present simulations were conducted on online behavioral data that were collected by a probing procedure. The probing procedure is an adequate methodology to examine the immediate spontaneous activation of textual and inferred information. However, probes provide minimal data and are possibly not representative of the diverse wealth of information that a reader activates spontaneously or strategically during reading (Magliano, Trabasso, & Graesser, 1999). Simulating data that were collected by a think-aloud procedure, for example, could greatly enrich the computational simulations of online processes and enhance their validity. This type of

simulation would also promote the integration of text-based with reader-based factors, because each simulation would be conducted on a specific text and a specific think-aloud protocol that was collected by a specific reader. In the present research, we introduced the LS-R model and demonstrated some of its basic capabilities for simulating human comprehension phenomena. Future studies could utilize these capabilities for the simulation of other comprehension phenomena that have been studied using diverse behavioral methodologies.

Author note This research was supported by Grant No. PIEF-GA-2009-254607-IEF from the European Union to M.Y. Facilities for conducting the research were provided by the Brain and Education Lab, the Leiden Institute for Brain and Cognition, and the Department of Education at Leiden University. We thank Eva Leusink, Sanne Rovers, and Adi Avramovich for help setting up and running the simulations.

Appendix

One text example from Yeari et al. (in press), divided into information units

Underlined are the central information units, and in italics are the peripheral information units that received the highest agreement, and thus were included in the analyses.

Mount Vesuvius

Mount Vesuvius is a volcano located between the ancient Italian cities of Pompeii and Herculaneum.

It has received much attention because of its frequent and destructive eruptions.

The most famous of these eruptions occurred in A.D. 79.

The volcano had been inactive for centuries.

There was little warning of the coming eruption, although one account unearthed by archaeologists says that a hard rain and a strong wind had disturbed the celestial calm during the preceding night.

Early the next morning, the volcano poured a huge river of molten rock down upon Herculaneum,

completely burying the city

and filling in the harbor with coagulated lava.

Meanwhile, on the other side of the mountain, cinders, stone and ash rained down on Pompeii.

Sparks from the burning ash ignited the combustible rooftops quickly.

Large portions of the city were destroyed in the conflagration.

Fire, however, was not the only cause of destruction.

Poisonous sulphuric gases saturated the air.

These heavy gases were not buoyant in the atmosphere and therefore sank toward the earth and suffocated people.

Over the years, excavations of Pompeii and Herculaneum have revealed a great deal about the behavior of the volcano.

By analyzing data, much as a zoologist dissects a specimen animal,

scientists have concluded that the eruption changed large portions of the area's geography.

For instance, it turned the Sarno River from its course and raised the level of the beach along the Bay of Naples.

Meteorologists studying these events have also concluded that Vesuvius caused a huge tidal wave that affected the world's climate.

In addition to making these investigations,

archaeologists have been able to study the skeletons of victims

by using distilled water to wash away the volcanic ash.

By strengthening the brittle bones with acrylic paint,

scientists have been able to examine the skeletons

and draw conclusions about the diet and habits of the residents.

Finally, the excavations at both Pompeii and Herculaneum have yielded many examples of classical art,

such as jewelry made of bronze,

which is an alloy of copper and tin.

The eruption of Mount Vesuvius and its tragic consequences have provided us with a wealth of data about the effects that volcanoes can have on the surrounding area.

Today volcanologists can locate and predict eruptions, saving lives and preventing the destruction of cities and cultures.

References

- Albrecht, J. E., & O'Brien, E. J. (1991). Effects of centrality on retrieval of text-based concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 932–939. doi:10.1037/0278-7393.17.5.932
- Beeman, M. J., Bowden, E. M., & Gernsbacher, M. A. (2000). Right and left hemisphere cooperation for drawing predictive and coherence inferences during normal story comprehension. *Brain and Language*, *71*, 310–336.
- Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, *11*, 177–220.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717–726.
- Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, *29*, 850–859.
- Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*, 257–273.
- Cook, A. E., & Guéraud, S. (2005). What have we been missing? The role of general world knowledge in discourse processing. *Discourse Processes*, *39*, 265–278.
- Cook, A. E., Limber, J. E., & O'Brien, E. J. (2001). Situation-based context and the availability of predictive inferences. *Journal of Memory and Language*, *44*, 220–234.
- Fincher-Kiefer, R. (1996). Encoding differences between bridging and predictive inferences. *Discourse Processes*, *22*, 225–246. doi:10.1080/01638539609544974
- Frank, S. L., Koppen, M., Noordman, L. G. M., & Vonk, W. (2003). Modeling knowledge-based inferences in story comprehension. *Cognitive Science*, *27*, 875–910.
- Frank, S. L., Koppen, M., Noordman, L. G. M., & Vonk, W. (2007). Modeling multiple levels of text representation. In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes* (pp. 133–157). Mahwah, NJ: Erlbaum.
- Frank, S. L., Koppen, M., Noordman, L. G. M., & Vonk, W. (2008). World knowledge in computational models of discourse comprehension. *Discourse Processes*, *45*, 429–463.
- Gaddy, M. L., Van den Broek, P., & Sung, Y. (2001). The influence of text cues on the allocation of attention during reading. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 89–110). Amsterdam, The Netherlands: Benjamins.
- Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, *117*, 227–247. doi:10.1037/0096-3445.117.3.227
- Golden, R. M., & Rumelhart, D. E. (1993). A parallel distributed processing model of story comprehension and recall. *Discourse Processes*, *16*, 203–237.
- Goldman, S. R., & Varma, S. (1995). CAPPING the construction-integration model of discourse comprehension. In C. A. Weaver III, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 337–358). Hillsdale, NJ: Erlbaum.
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*, 163–189. doi:10.1146/annurev.psych.48.1.163
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395. doi:10.1037/0033-295X.101.3.371
- Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*, 1–37. doi:10.1037/0033-295X.114.1.1
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149. doi:10.1037/0033-295X.99.1.122
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review*, *95*, 163–182. doi:10.1037/0033-295X.95.2.163
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, *7*, 257–266.
- Kintsch, W. (2001). Predication. *Cognitive Science*, *25*, 173–202.
- Kintsch, W., & Welsch, D. M. (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 of Bennett B. Murdock* (pp. 367–385). Hillsdale, NJ: Erlbaum.
- Kintsch, W., Welsch, D. M., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*, 133–159.
- Klin, C. M., Guzman, A. E., & Levine, W. H. (1999). Prevalence and persistence of predictive inferences. *Journal of Memory and Language*, *40*, 593–604.
- Klin, C. M., Murray, J. D., Levine, W. H., & Guzmán, A. E. (1999). Forward inferences: From activation to long-term memory. *Discourse Processes*, *27*, 241–260.

- Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, 23, 1230–1246.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240. doi:10.1037/0033-295X.104.2.211
- Landauer, T. K., Foltz, P., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25, 259–284. doi:10.1080/01638539809545028
- Langston, M., & Trabasso, T. (1999). Modeling causal integration and availability of information during comprehension of narrative texts. In H. van Oostendorp & S. Goldman (Eds.), *The construction of mental representations during reading* (pp. 29–69). Mahwah, NJ: Erlbaum.
- Langston, M. C., Trabasso, T., & Magliano, J. P. (1998). Modeling on-line comprehension. In A. Ram & K. Moorman (Eds.), *Computational models of reading and understanding* (pp. 181–225). Cambridge, MA: MIT Press.
- Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes*, 34, 259–280.
- Linderholm, T., & van Den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, 94, 778–784. doi:10.1037/0022-0663.94.4.778
- Linderholm, T., Virtue, S., Tzeng, Y., & Van den Broek, P. (2004). Fluctuations in the availability of information during reading: Capturing cognitive processes using the landscape model. *Discourse Processes*, 37, 165–186.
- Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, 91, 615–629. doi:10.1037/0022-0663.91.4.615
- McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114, 159–188.
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, 22, 247–288.
- McNamara, D. S., & O'Reilly, T. (2009). Theories of comprehension skill: Knowledge and strategies versus capacity and suppression. In A. M. Columbus (Ed.), *Advances in psychology research* (Vol. 62, pp. 113–136). Hauppauge, NY: Nova Science.
- Miller, A. C., & Keenan, J. M. (2009). How word reading skill impacts text memory: The centrality deficit and how domain knowledge can compensate. *Annals of Dyslexia*, 59, 99–113.
- Millis, K., & Larson, M. (2008). Applying the construction–integration framework to aesthetic responses to representational artworks. *Discourse Processes*, 45, 263–287.
- Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, 26, 131–157. doi:10.1080/01638539809545042
- Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in text: What readers believe versus what they read. *Psychological Science*, 3, 229–235. doi:10.1111/j.1467-9280.1992.tb00034.x
- Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, 19, 228–242.
- Perfetti, C. A. (1998). The limits of co-occurrence: Tools and theories in language research. *Discourse Processes*, 25, 363–377.
- Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation models and aging. *Psychology and Aging*, 16, 145–160. doi:10.1037/0882-7974.16.1.145
- Sanjosé, V., Vidal-Abarca, E., & Padilla, O. (2006). A connectionist extension to Kintsch's Construction–Integration model. *Discourse Processes*, 42, 1–35.
- Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes*, 33, 105–132.
- Shapiro, A. M. (2004). How including prior knowledge as a subject variable can change outcomes of learning research. *American Educational Research Journal*, 41, 159–189.
- Shears, C., & Chiarello, C. (2004). Knowledge-based inferences are not general. *Discourse Processes*, 38, 31–55.
- Shears, C., Miller, V., Ball, M., Hawkins, A., Griggs, J., & Varner, A. (2007). Cognitive demand differences in causal inferences: Characters' plans are more difficult to comprehend than physical causation. *Discourse Processes*, 43, 255–278. doi:10.1080/01638530701226238
- Singer, M. (1996). Comprehending consistent and inconsistent causal text sequences: A construction–integration analysis. *Discourse Processes*, 21, 1–21. doi:10.1080/01638539609544947
- Singer, M., & Halldorson, M. (1996). Constructing and validating motive bridging inferences. *Cognitive Psychology*, 30, 1–38.
- Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences. *Journal of Memory and Language*, 31, 507–524.
- Singer, M., & Kintsch, W. (2001). Text retrieval: A theoretical exploration. *Discourse Processes*, 31, 27–59. doi:10.1207/S15326950dp3101_2
- Tapiero, I., & Denhière, G. (1995). Simulating recall and recognition by using Kintsch's construction–integration model. In C. Weaver, S. Mannes, & C. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 211–232). Hillsdale, NJ: Erlbaum.
- Thibadeau, R., Just, M. A., & Carpenter, P. A. (1982). A model of the time course and content of reading. *Cognitive Science*, 6, 157–203.
- Trabasso, T., & Bartolone, J. (2003). Story understanding and counterfactual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 904–923. doi:10.1037/0278-7393.29.5.904
- Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language*, 24, 595–611.
- Trabasso, T., van den Broek, P., & Suh, S. Y. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes*, 12, 1–25. doi:10.1080/01638538909544717
- Trabasso, T., & Wiley, J. (2005). Goal plans of action and inferences during comprehension of narratives. *Discourse Processes*, 29, 129–164.
- Tzeng, Y., van den Broek, P., Kendeou, P., & Lee, C. (2005). The computational implementation of the landscape model: Modeling inferential processes and memory representations of text comprehension. *Behavior Research Methods*, 37, 277–286. doi:10.3758/BF03192695
- van den Broek, P. (1988). The effects of causal relations and hierarchical position on the importance of story statements. *Journal of Memory and Language*, 27, 1–22.
- van den Broek, P. (1994). Comprehension and memory of narrative texts: Inferences and coherence. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 539–588). San Diego, CA: Academic Press.
- van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science*, 328, 453–456.
- van den Broek, P., Bohn-Gettler, C., Kendeou, P., Carlson, S., & White, M. J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 123–140). Greenwich, CT: Information Age.
- van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: The role of co-activation in confronting misconceptions. *Applied Cognitive Psychology*, 22, 335–351. doi:10.1002/acp.1418
- van den Broek, P., Ridsen, K., Fletcher, C. R., & Thurlow, R. (1996). A “landscape” view of reading: Fluctuating patterns of activation and

- the construction of a stable memory representation. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 165–187). Mahwah, NJ: Erlbaum.
- van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Erlbaum.
- Wolfe, M. B. W., Magliano, J. P., & Larsen, B. (2005). Causal and semantic relatedness in discourse understanding and representation. *Discourse Processes*, *39*, 165–187.
- Yeari, M., Oudega, M., & van den Broek, P. (in press). The effect of highlighting on processing and memory of central and peripheral text information: Evidence from eye movements. *Journal of Research in Reading*. doi:10.1111/1467-9817.12072
- Yeari, M., & van den Broek, P. (2015). The role of textual semantic constraints in knowledge-based inference generation during reading comprehension: A computational approach. *Memory*, *23*, 1193–1214. doi:10.1080/09658211.2014.968169
- Yeari, M., van den Broek, P., & Oudega, M. (2015). Processing and memory of central versus peripheral information as a function of reading goals: Evidence from eye-movements. *Reading and Writing*, *28*, 1–27.