World Knowledge in Computational Models of Discourse Comprehension

Stefan L. Frank [a]; Mathieu Koppen [a]; Leo G. M. Noordman [a]; Wietske Vonk [bc]
[a] Nijmegen Institute for Cognition and Information, Radboud University Nijmegen, The Netherlands [b] Max
Planck Institute for Psycholinguistics, Nijmegen, The Netherlands [c] Center for Language Studies, Radboud
University, Nijmegen, The Netherlands

PLEASE SCROLL DOWN FOR ARTICLE

World Knowledge in Computational Models of Discourse Comprehension

Stefan L. Frank, Mathieu Koppen, and Leo G. M. Noordman
Nijmegen Institute for Cognition and Information
Radboud University Nijmegen, The Netherlands

Wietske Vonk
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
Center for Language Studies, Radboud University Nijmegen,
The Netherlands

Because higher level cognitive processes generally involve the use of world knowledge, computational models of these processes require the implementation of a knowledge base. This article identifies and discusses 4 strategies for dealing with world knowledge in computational models: disregarding world knowledge, *ad hoc* selection, extraction from text corpora, and implementation of all knowledge about a simplified *microworld*. Each of these strategies is illustrated by a detailed discussion of a model of discourse comprehension. It is argued that seemingly successful modeling results are uninformative if knowledge is implemented ad hoc or not at all, that knowledge extracted from large text corpora is not appropriate for discourse comprehension, and that a suitable implementation can be obtained by applying the *microworld strategy*.

To explain experimental findings in cognitive psychology, models of the underlying process are constructed. Often, such models are expressed only verbally—that is, without equations or completely specified algorithms. As long as such a verbal model is not too vague or too complex, it is possible to ascertain its

Correspondence concerning this article should be addressed to Stefan L. Frank, Institute for Logic, Language, and Computation, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands. E-mail: S.L.Frank@uva.nl

internal consistency and to use the model to make qualitative predictions. As a model's complexity increases, however, it becomes more difficult to test without turning it into a *computational* model—that is, specifying it in enough detail and with sufficient precision to implement and run the model as a computer program—the output of which can be compared to experimental results. For a computational model to be regarded as a simulation of some cognitive process, there should at least be a qualitative correspondence (i.e., a correspondence on an ordinal scale) between the model's outcome and the experimental data.

Simply demonstrating such a correspondence in one particular simulation is not enough to conclude that the model simulates the cognitive process. Running a simulation often involves choices and assumptions in addition to those for the construction of the model itself, and it may be those additional choices, instead of the model, which are (partly) responsible for seemingly successful modeling results. This issue is of particular importance when dealing with models of higher level cognitive processes. In general, the higher the level of the simulated process, the larger the gap between experiment and simulation that needs to be bridged. For instance, if the model does not include lower level, perceptual processes, its input must be a preprocessed version of experimental stimuli, if there is any relation between the two at all. The particular way the input is constructed will, of course, affect the model's behavior. Also, processes such as comprehension and common-sense reasoning often require the application of background knowledge, but no realistic amount of such knowledge can be made available to a model. Therefore, choices need to be made regarding the world knowledge that *is* included, and these can have a profound effect on the model's outcome.

We discuss different strategies for dealing with world knowledge in computational models. Potential pitfalls and opportunities arising from these strategies are illustrated by in-depth discussions of several models of one particular high-level cognitive process: the comprehension of discourse. As will become clear, the strong influence of world knowledge on the comprehension process has as a consequence that much of the alleged success in computational modeling hinges on details of knowledge implementation and cannot be attributed to the models themselves but only to choices made beyond those for model construction.

To clarify how the chosen strategy for world-knowledge implementation affects a models' outcome, it is necessary to go quite deeply into some of the details of the models. Nevertheless, note that our concern is the comparison of the different knowledge-implementation strategies—that is, our claims apply to these strategies and not to the particular models. Indeed, as we argue in the Discussion, some models can use strategies other than those with which they were originally designed. It is not our goal to identify the models' strengths and weaknesses or to make a direct comparison among the models. The latter would be close to impossible considering the large variance among the models' scopes and aims. The models are only investigated to illustrate the pros and cons of the

different knowledge-implementation strategies, and to show that problems that are likely to arise from some of these strategies do in fact occur in practice.

Although we only discuss models of discourse comprehension, our claims generally apply to computational models of cognitive processes that rely heavily on large amounts of world knowledge. In fact, the scope of the models we discuss is not necessarily restricted to discourse comprehension: Some of these models have been applied (or have at least inspired research) in fields as diverse as action planning (Doane & Sohn, 2000; Mannes & Kintsch, 1991), human–computer interaction (Kitajima & Polson, 1995), skill learning (Doane, Sohn, McNamara, & Adams, 2000), advertising (Luna, 2005; Yang, Roskos-Ewoldsen, & Roskos-Ewoldsen, 2004), and philosophy of mind (Frank & Haselager, 2006).

OVERVIEW

In the following sections, we argue:

1. Any model of discourse comprehension should, at least in principle, be able to handle all world knowledge that is potentially relevant to the simulated process. In practice, however, models that are not designed to deal with knowledge tend to break down when potentially relevant knowledge is added.
2. Limiting the implemented world knowledge to what is needed for the particular text under consideration makes it impossible to test a model's general properties or to validate it against experimental data.
3. Although generally applicable knowledge can be implemented by automatic extraction from large text corpora, this strategy results in knowledge of word meaning rather than knowledge of causal relations in the world, making it unsuitable for most aspects of discourse comprehension.
4. Because the implementation of realistic amounts of suitable world knowledge is infeasible, the applicability of models needs to be restricted to a simplified world. Apparent drawbacks of this *microworld strategy* apply to other strategies to the same extent at least.
5. Applying the microworld strategy to models that have not used it so far may facilitate model testing and improvement.

DISREGARDING WORLD KNOWLEDGE

Because world knowledge permeates almost every aspect of text comprehension (see Cook & Guéraud, 2005, for an overview), it clearly cannot generally be disregarded in models of this process. However, for particular low-level

subprocesses, the influence of world knowledge may seem small enough to be ignored. A model simulating only such a subprocess would be able to do without any implementation of world knowledge. If, however, world knowledge is potentially relevant to the simulated process, then the inclusion of such knowledge in the model should at least be feasible in principle, even when the current model does not deal with it. We illustrate this claim by discussing the Resonance model (Myers & O'Brien, 1998), which is intended to simulate only a low-level subprocess of discourse comprehension: the fluctuating activation of text items in the reader's mental representation of the text. As Myers & O'Brien acknowledge, the reader's knowledge can play a role in this process. Nevertheless, only information from the text is implemented in the Resonance model to simplify the analysis. In the following, we explain the model; illustrate that it may require world knowledge; and, crucially, show that it is likely to break down when extended with this knowledge.

The Resonance Model

Simulated process.   As reading proceeds, different parts of the representation of the preceding text differ in their accessibility to the reader. For instance, concepts and propositions that are central to the text can remain in working memory while less important elements are backgrounded. However, previously backgrounded text items can become reactivated if this is required or instigated by the sentence currently being read. The phenomenon is know as *reinstatement*. According to Albrecht and Myers (1995), reinstatement results from a bottom-up process by which previous text items "resonate" to the items in the current sentence. The Resonance model attempts to simulate this process of reinstatement-by-resonance.

The resonance process.   The Resonance model processes the sentences of a text one at a time. Each sentence is parsed into three types of items: concepts, propositions, and a sentence marker. Every time a sentence enters the model, these items are added as nodes to a network representing the text. These nodes can be connected to each other and to the nodes corresponding to previous text items. Such links depend only on the items' co-occurrence in a sentence and on propositional forms,[1] and not on the meaning of items or their relation to the reader's knowledge. Because all network nodes follow directly from the text and the links between them are based on formal, not semantic, considerations, world knowledge is not present in the Resonance model. This

---

[1]The sentence marker is connected to the propositions in the sentence, concepts are connected to the propositions that have them as an argument, and propositions are mutually connected if they have identical arguments (Myers & O'Brien, 1998, pp. 143–144).

absence of world knowledge does not follow from any theoretical claim, but is merely a simplification (see Footnote 2 in Myers & O'Brien, 1998).

After adding the items from the current sentence to the text network, *resonance values* and *signal strengths* of all network items are computed over a number of processing cycles. An item's resonance value can be interpreted as its availability in the reader's working memory, and its signal strength indicates the extent to which it can affect the resonance values of other items. Initially, all resonance values are set to 0, and items from the current sentence receive an initial signal strength. In each processing cycle, items that have a signal strength send a signal to all connected items. As a result, the resonance of a receiving item increases by the total amount of signal received. After computing the new resonance values, signal strengths are updated. An item's signal strength increases as its resonance increases, but decays over processing cycles. The rate of this decay is controlled by a parameter β that can take any positive value.

Resonances and signal strengths are repeatedly updated until all signal strengths fall below a small threshold value. At this point, the sentence has been processed. The four items that end up with the largest resonance values are said to remain in working memory. When processing the next sentence, not only the sentence's items but also the four items in working memory receive an initial signal strength.

Reinstatement.   The Resonance model has, for instance, been applied to the "captain text" (see Table 1 in Myers & O'Brien, 1998), which introduces a captain sitting at his desk to finish his ship's inventory. However, before he starts the inventory, he is called away, leaving the inventory forms on the desk. After this, the text has no more reference to the unfinished inventory or to the desk, until, at the end of the story, the captain is said to return to his office and sit down at his desk. In this sentence, the concept DESK is assumed to activate the previous occurrence of DESK in the text, reinstating the unfinished inventory. Indeed, Albrecht and Myers (1995) found experimental support for such reinstatement.

In the model, reinstatement consists in bringing certain *critical items* into working memory. Therefore, the amount of reinstatement can be defined as the increase in the number of critical items in working memory that is brought about by processing the reinstating sentence. Because working memory consists of four items, this is the maximum amount of reinstatement. When running the Resonance model on the captain text[2] with levels of decay rate β ranging from

---

[2]The text was parsed into 103 items, based on those provided by J. L. Myers (personal communication, September 20, 1995) and Weeber (1996). The critical items are the concept INVENTORY, the propositions of which it is an argument, and the markers of the sentences that contain them (Myers & O'Brien, 1998, p. 147).

0.005 to 0.7, we found reinstatement for $0.019 \leq \beta \leq 0.048$ only, with the amount of reinstatement being, at most, two items.

For a successful simulation of the experiment by Albrecht and Myers (1995), there should be less reinstatement in an alternative version of the story in which the desk is not mentioned in the reinstating sentence. Indeed, Myers & O'Brien (1998, p. 148) reported no reinstatement of critical elements at all when processing this alternative story. When we removed the proposition mentioning the desk from the reinstating sentence, we found a decrease in amount of reinstatement from 2 to 1, for values of β ranging from 0.027 to 0.032.

Recency and connectivity.   The accessibility of text items depends on more than just reinstatement. Two main experimental findings are that more recently read items are, in general, more available than older items, and that items which are central to the text are more accessible than less important items (Albrecht & Myers, 1991; O'Brien, 1987; O'Brien, Albrecht, Hakala, & Rizzella, 1995). The Resonance model simulates both these effects. First, items from the current sentence receive an initial signal and can, therefore, be expected to end up with larger resonance values than other items, resulting in a *recency effect*. Second, items that are more central to the text have many connections to other items and are, therefore, more likely to receive large resonance values, resulting in a *connectivity effect*.

The decay rate β can be expected to strongly affect the magnitudes of the recency and connectivity effects. If the decay rate is large, signals decay quickly and resonance will not spread far through the text network, resulting in a strong recency effect. For small values of β, the opposite happens: Signals keep spreading throughout the network, and most resonance will eventually settle on the items that have the largest number of connections.

The magnitude of the connectivity effect can be measured by the proportion of variance in resonance values explained by the items' numbers of connections. Likewise, the size of the recency effect is measured by the proportion of this variance explained by whether the items occurred in the current sentence or not. Figure 1 shows that, in our simulations, a clear trade-off between the recency effect and the connectivity effect, controlled by β, indeed occurs when processing the captain text.

The narrow range of decay rates (between 0.019 and 0.048) resulting in reinstatement corresponds to a very weak recency effect. This is not surprising because a weaker recency effect means that non-recent items have higher resonance values and, by definition of reinstatement, critical items are not recent when the reinstating sentence is being processed. However, a weak recency effect necessarily comes with a strong connectivity effect, as Figure 1 shows. If the value of β is too low, the effect of connectivity can become so strong that
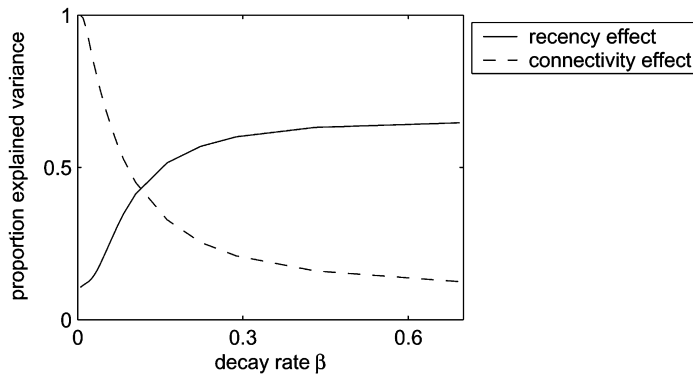
FIGURE 1   Proportion of variance in resonance values explained by recency or connectivity, as a function of decay parameter β, after processing the reinstating sentence of the captain text.

reinstatement is no longer possible because working memory is always occupied by the four most strongly connected items. For the lowest decay rate at which reinstatement occurs ($\beta = 0.019$), connectivity already explains almost 93% of variance in final resonance values.

World Knowledge in the Resonance Model

The captain text showed that bottom-up reinstatement can be simulated by a resonance process without including world knowledge. However, in another text to which Myers and O'Brien (1998) applied the Resonance model, one of the concepts CHURCH or BARN is to be reinstated by the mention of BUILDING. This reinstatement requires knowledge about *semantic* relations between these concepts. Contrary to the normal rules for connecting text items, connections between BUILDING and CHURCH and between BUILDING and BARN were added to the network (see Table 2 footnote in Myers & O'Brien, 1998), implementing precisely the piece of information needed for reinstatement. This example shows that world knowledge cannot completely be ignored for the simulated process. More important, if *some* piece of knowledge about the semantic relations between text concepts can be required for a particular case of reinstatement, then *any* semantic relation in the text could become relevant. Consequently, such semantic relations should be generally available to the model.

As previously shown, the model's ability to simulate reinstatement depends on a delicate balance between recency and connectivity effects. Crucially, this

balance is disturbed when semantic knowledge is included in the model. Connecting all pairs of semantically related concepts leads to a strong increase in the number of connections of some concepts in the network. Because of the strong connectivity effect that comes with the small effect of recency that is required, these concepts will often occupy working memory, preventing reinstatement. Indeed, after connecting all of the captain text's semantically related concepts to each other,[3] we no longer found any reinstatement, regardless of the level of β—that is, the Resonance model no longer functions when potentially relevant knowledge is included.

Conclusion

In particular instances, such as the captain text, the activation of text items by a bottom-up resonance process can be simulated without resorting to world knowledge. However, the fact that this is not so in general (as we have seen for the CHURCH–BUILDING–BARN example) entails that models of this process should at least be compatible with the presence of knowledge of word semantics. The Resonance model was never intended to function with a large amount of added knowledge. Consequently, it cannot deal with the highly connected network that results from including semantic relations, as was shown by our discussion of the recency–connectivity balance, coupled with the model's failure on the captain text when semantic knowledge was added.

More generally, any cognitive model should be compatible with the presence of all knowledge that is potentially relevant to the simulated process. However, it is to be expected that, by its sheer abundance, the inclusion of this knowledge will come to dominate the processing of any model that has not been designed to take knowledge into account. An obvious way out is to include only a small amount of knowledge such that the model's process will not become disrupted. For instance, the Resonance model achieved reinstatement of CHURCH or BARN by the *ad hoc* inclusion of nothing more than two required connections; namely, from BUILDING to CHURCH and BARN. As discussed in the next section, this "solution" of including only the specific pieces of knowledge needed for the case at hand is applied more widely with other models, and has the unwanted consequence of making proper investigation or validation of these models impossible.

---

[3]Semantic relations between concepts were found by using latent semantic analysis (LSA). Two concepts were connected if their semantic relatedness according to LSA's cosine score was at least as large as the cosine between BUILDING and CHURCH, which is .10. This resulted in 53.3% of concept pairs becoming connected. In total, the number of connections in the network increased by 60%.

AD HOC SELECTION OF WORLD KNOWLEDGE

All models of discourse comprehension require information in addition to the input text. Such additional information is encoded in the setting of parameters. Note that we use the term *parameter* in a fairly broad sense, to refer not only to quantitative variables (i.e., those having numerical values) but also to qualitative variables, such as whether a particular fact from world knowledge is made available to the model.

Parameters can also be classified by the role they play in the simulation. For our objective, an important distinction is between *free* and *fixed* parameters (each of which can be either quantitative or qualitative). Free parameters may vary over simulation runs, and can therefore be considered part of the input to the model rather than part of the model itself. Their values are set by the modeler on an ad hoc basis. A quantitative parameter like "activation decay rate," for instance, is a free parameter if its value is chosen dependent on the particular text to be processed. Ideally, a model should not include any free parameters because seemingly successful simulation results that depend crucially on the particular choice of free parameter values may incorrectly be attributed to the properties of the model itself.

Fixed parameters, in contrast, have values that remain the same whatever the model's input, so can be viewed as part of the model itself. Static aspects of the experimental context should correspond to fixed parameters in the simulation of the experiment. Qualitative fixed parameters are of great importance to models of discourse comprehension. Understanding a text always takes place within the context of the reader's background knowledge, so it is necessary for the model to have access to a knowledge base. To the extent that the reader's world knowledge is assumed constant among or within readers, its implementation in the model should be fixed—that is, take the form of fixed parameters.

When the choice between providing or not providing the model with some specific piece of world knowledge nevertheless depends on the text to be processed, this comes down to the setting of a qualitative free parameter. Several discourse-comprehension models use such a strategy: Pieces of world knowledge that seem relevant to the particular text under consideration are hand-picked, and only those are provided to the model. This prevents technical problems with running a model in the context of large amounts of knowledge, but introduces free parameters.

Although the outcomes of models that use free parameters for world-knowledge implementation have been claimed to support the models' validity, most of these results are in fact direct reflections of the ad hoc and subjective selection and implementation of world knowledge. The Construction–Integration (CI) model (Kintsch, 1988; Kintsch, 1998), which has been widely applied and has stimulated much experimental research, is the most influential discourse-

comprehension model using this strategy for world knowledge implementation. By a detailed discussion of the CI model, we intend to show that this strategy makes it impossible to determine whether the model's results reflect properties of the model or only of the parameter setting.

The CI Model

Simulated process.   Most texts do not explicitly mention everything that is needed for comprehension. To make sense of a text, inferences need to be made by applying world knowledge to the text-given information. The CI model simulates this inference process. It assumes that text comprehension takes place in two phases. In the first phase, called *construction*, world knowledge items that are potentially relevant to the text are selected. Next, from the resulting collection of text and knowledge items, less relevant or inappropriate items are discarded in the *integration* phase.

Construction.   The model takes as input a collection of concepts and propositions that correspond to the text being processed. For example, Kintsch (1988) parsed the short text *The lawyer discussed the case with the judge. He said "I shall send the defendant to prison."* into the five text propositions labeled "T" in Table 1. Two of these correspond to the text's incorrect interpretation in which the ambiguous pronoun *he* refers to the lawyer, and two others correspond to the correct reading in which *he* is the judge.

During the construction phase, each concept and proposition from the text retrieves a small number of items from the reader's world knowledge in a context-free manner. Knowledge items have a better chance at being retrieved

TABLE 1
Five Propositions, Labeled "T", Taken From the Text
*The Lawyer Discussed the Case With the Judge. He Said
"I Shall Send the Defendant to Prison."*, and the
Two Propositions, Labeled "A", That Become Associated
During the Construction Phase (Kintsch, 1988, p. 169)

Label	Proposition
T1	DISCUSS(LAWYER, JUDGE, CASE)
T2	SAY(LAWYER, T3)
T3	SEND(LAWYER, DEFENDANT, PRISON)
T4	SAY(JUDGE, T5)
T5	SEND(JUDGE, DEFENDANT, PRISON)
A1	SENTENCE(JUDGE, DEFENDANT)
A2	IMPLY(A1, T5)

if, according to the reader's knowledge, their association to the text item is stronger (Kintsch, 1988, p. 166). It is important to note that, in theory, only this association strength has an effect on the retrieval process. It is not possible for context information to influence which knowledge items are chosen. As Kintsch (1988) put it, "The construction process lacks guidance and intelligence; it simply produces potential inferences, in the hope that some of them might turn out to be useful" (p. 167). In practice, however, the construction process is strongly guided by the modeler's intelligence. Because the "real" association strengths between the items from the text and those from reader's knowledge are unknown, it is up to the modeler to decide which knowledge items are important to the text at hand. For instance, the five text propositions in Table 1 are assumed to retrieve two world-knowledge propositions in total: Text proposition (T3) SEND(LAWYER, DEFENDANT, PRISON) does not retrieve any items because the reader does not know anything about lawyers sending defendants to prison. On the other hand, the proposition (T5) SEND(JUDGE, DEFENDANT, PRISON) does retrieve two items from world knowledge. These propositions, labeled "A" in Table 1, are (A1) SENTENCE(JUDGE, DEFENDANT) and a proposition (A2) IMPLY(A1, T5), stating that sentencing the defendant implies sending him or her to prison. None of the other text propositions are associated to any world-knowledge item. Clearly, selecting (A1) and (A2), and just these two, comes down to a setting of qualitative free parameters.

So far, the construction process resulted in a collection of unconnected concepts and propositions, originating both from the text and from world knowledge. Next, these items are connected to each other, forming a network called the *enriched textbase* (Kintsch, 1988, p. 166) in which each connection is assigned a weight. The enriched-textbase network resulting from the "lawyer and judge" example, including connection weights, is shown in Figure 2.

There is no agreement about objective rules for choosing the weight values. Even the set of weights that is used varies strongly among different applications of the model. Kintsch (1988) used real values between -1 and $+1$, but in most other studies weights are taken from a small set of integer values. For instance, Kintsch and Welsch (1991) used only weights of 0, 1, and 2, whereas Tapiero and Denhière (1995) also included weights of -3. The integer values 0 to 5 are used in three different studies (Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Radvansky, Zwaan, Curiel, & Copeland, 2001; Schmalhofer, McDaniel, & Keefe, 2002), but Schmalhofer et al. used a different assignment of these weights to the connections. In Otero and Kintsch (1992), connections had a weight of -1, 0, 1, 2, or 10, but Singer (1996) used -1, 0, and 1, whereas Singer and Halldorson (1996) also included $\frac{1}{2}$. Also, it is not clear whether items should be connected to themselves. In Kintsch (1988) they are not, but according to Kintsch (1998) they are. Kintsch (1992) even used the weights of these self-connections to indicate which propositions are emphasized by the text, whereas
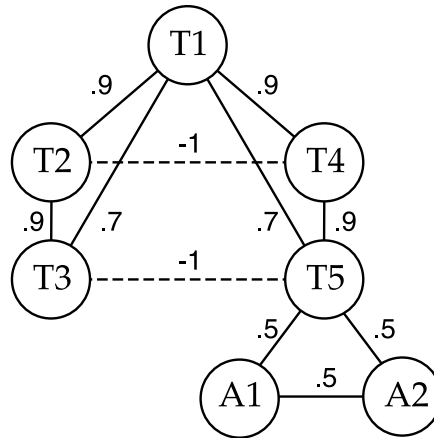
FIGURE 2   Enriched textbase network corresponding to the propositions in Table 1, including the connection weights (Kintsch, 1988, pp. 169–170). *Note*. Weights between text propositions depend on their distance in the text. The connections to, and between, associated propositions are assumed to have a weight of .5. Propositions corresponding to different interpretations of the pronoun are maximally negatively connected (indicated by dashed lines) because they exclude each other.

others varied them to simulate differences in the reader's goal (Schmalhofer et al., 2002), strategy (Otero & Kintsch, 1992), or age (Radvansky et al., 2001). This lack of consistency shows that the connection weights are quantitative free parameters.

Integration.   The integration phase of the CI model takes as input the enriched textbase that resulted from the construction phase and selects which nodes of this network can be discarded because they are less relevant to the text than others, or because they are inconsistent with the rest of the enriched textbase. This is accomplished by assigning an initial activation value to items originating from the text. The integration process spreads the activation among the network over a number of processing cycles until the average difference between the old and new activation values falls below some small fixed value. The items that now have the highest activation values are assumed to be the most relevant. Highly activated items originating from world knowledge are the inferences that were drawn from the text.

In the enriched textbase of Figure 2, proposition T5 has more positive connections than its competitor T3 and will, therefore, receive a higher activation value during integration. As a result, proposition T4 is connected to a more active node than its competitor T2, so it too will receive more activation. Eventually, the

activations of T2 and T3 (the LAWYER nodes) become 0, whereas the activations of T4 and T5 (the JUDGE nodes) remain high. The incorrect interpretation of the pronoun is discarded, and the correct one is kept: It is the judge, and not the lawyer, who says that he shall send the defendant to prison.

World Knowledge in the CI Model

Selection of associated knowledge items.   In theory, the particular knowledge items that become connected to the text during the construction phase are selected by a "dumb" process of association. In practice, exactly the items necessary for comprehension of the text, together with just a few distracting items, are carefully chosen by the modeler. Because this constitutes a setting of many free parameters, there exists a serious risk that a successful outcome of the integration process reflects the aptness of this choice rather than the aptness of the model.

To give an illustrative example of this, Kintsch, Patel, & Ericsson (1999; see also Kintsch, 2000a) apply CI to the ambiguous newspaper headline *Iraqi head seeks arms*. They report that, for the integration process to deactivate the incorrect reading of the sentence, several world-knowledge propositions were needed: "Iraq is a country," "Countries have governments," "Governments have heads," and "Arms are weapons." After applying the integration process to the enriched textbase network including these propositions, the items corresponding to the correct reading of the sentence ("The government of Iraq wants weapons") had higher activation values than the items corresponding to the incorrect reading. However, this result depends critically on the specific knowledge-based propositions that were chosen to be included because all of these were only connected to the text items corresponding to the correct reading. As a result, these "correct items" have more positive connections than the "incorrect items," so they receive more activation. If propositions related to the other reading had been selected (e.g., "An Iraqi is a person," "A person has a body," "Heads and arms are parts of the body"), the incorrect reading of the sentence would have resulted from the integration process.[4]

Connection weight setting.   As noted earlier, the choice of connection weights, like that of associated world knowledge items, is an ad hoc and sub-

---

[4]Kintsch, Patel, and Ericsson (1999) acknowledged that, in this example, the world-knowledge items in the enriched textbase do not result from a context-free construction process, but instead from a "goal directed inference process" (p. 197). This inference process must have disambiguated the sentence to come up with these particular world-knowledge items, raising the question of what use the subsequent integration process is.

jective setting of free parameters. Such freedom can lead to modeling results that are almost completely determined by the particular choices made. Consider, for example, the following sentence used in an experiment by Till, Mross, and Kintsch (1988): *The townspeople were amazed to find that all the buildings had collapsed except the mint.* Obviously, *mint* refers to the building and not the candy. Kintsch (1988) used the CI model to solve this lexical ambiguity. A small part of the enriched textbase is shown in Figure 3.

The concept MINT is assumed to retrieve from world knowledge the concepts CANDY and MONEY, which are connected by a weight of −1 because they exclude each other (only one of them can end up in the interpretation of the sentence). Also, a connection of −.5 is added between CANDY and BUILDING, "because the homophone *mint* contributed associations to the subset that refers to both of its senses" (Kintsch, 1988, p. 173). During the integration process, the activation of CANDY becomes 0, whereas MONEY ends up with an activation of .074, indicating that the model found the intended meaning of *mint* (or, rather, discarded the unintended meaning).

Had the negative connection between CANDY and BUILDING not been added, the network would be completely symmetric with regard to CANDY and MONEY, so these two concepts would have received the same final activation and the ambiguity would remain unresolved. Because there is no reason to assume there is a negative association between the concepts CANDY and BUILDING in world knowledge (see also under "LSA for CI" in a following section), the negative connection between them seems to have been added to make sure the integration process would produce the desired result.
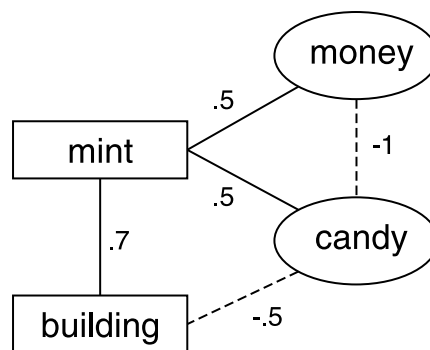

FIGURE 3   Part of the enriched textbase network of *The townspeople were amazed to find that all the buildings had collapsed except the mint*, including connection weights. *Note*. Rectangles are items originating from the text, ovals are the associated world-knowledge items. Dashed lines indicate negative connections (Kintsch, 1988, pp. 172–173).

Conclusion

Because different texts usually require different knowledge bases, implementing only text-relevant knowledge seems to make sense. However, this comes down to the introduction of many free parameters. With enough parameters to set, models can produce almost any desired output. As we have shown, in several applications of the CI model, presented results directly resulted from convenient choices regarding world knowledge. By implementing world knowledge subjectively and on an ad hoc basis, no conclusion about a model's properties or qualities can be drawn from its output because many other outcomes would also have been possible. This is not to say that such models must be flawed, but only that they cannot be properly validated until all parameters are fixed.

The "Connectionist Extension to CI" model (Sanjosé, Vidal-Abarca, & Padilla, 2006) deals with some of these issues by computing connection weights from activation values rather than choosing them subjectively. However, this does not solve the problems that associated world knowledge needs to be selected subjectively and that some initial connection weights are set by hand to encode the reader's prior knowledge (Sanjosé et al., 2006, p. 12).

So far, we have presented CI as a model of knowledge-based inference in which the activation of an item is interpreted as its relevance to the text. The model has also been widely applied to predict experimental data by comparing activation values to experimental measures. According to Schmalhofer, McDaniel, and Keefe (2002), a proposition's activation predicts the naming time of a word that is part of the proposition. They used CI to account for text priming effects on these naming times. For simulating experiments in which participants had to indicate whether a target sentence occurred in a previously read text, activation values are taken as a prediction of response times (Caillies, Denhière, & Kintsch, 2002) or sentence recognition probabilities (Caillies & Denhière, 2001; Kintsch et al., 1990; Radvansky et al., 2001). Tapiero and Denhière (1995) interpreted activation of a text proposition as its probability of free recall. For propositions denoting a fact from world knowledge, Singer (1996) and Singer and Halldorson (1996) took the activation to predict the time needed to verify whether the fact is true.

In these studies, the correspondence between model results and experimental data has been established extensively. Because qualitatively different inferences can easily be obtained by setting free parameters accordingly (as our discussion of several CI simulations has made clear), it stands to reason that the parameter setting also strongly affects the much subtler *quantitative* correspondences between model outcomes and experimental data. The fact that each text comes with its own parameters setting raises the question how persuasive these correspondences are.

This is also apparent in two other models that implement world knowledge as free parameters: the Landscape model (Van den Broek, Young, Tzeng, & Linderholm, 1999) and the model by Langston and Trabasso (1999). The latter model takes as input a network of nodes that represent text clauses. Text-specific knowledge about causality, as judged by the modeler, is implemented by weighted links that represent causal relations between the events referred to by the clauses. As we discussed extensively in a previous article (Frank, Koppen, Noordman, & Vonk, 2005), the model's output is virtually identical to this input. Consequently, the output also encodes the text's causal structure. Because causal structure is a major factor affecting discourse comprehension, the output of the Langston and Trabasso model corresponds to a lot of experimental data. However, this correspondence is no evidence in support of the model whatsoever because the same data is accounted for by the model's input. This makes the Langston & Trabasso model a very clear example of how the selection of world knowledge can have a larger impact on model results than does the actual model.

In the Landscape model, a similar problem arises when input values are mistakenly interpreted as model outputs. The model's input takes the form of activation values of items from the text and, if necessary, items that are assumed to be inferred from the reader's world knowledge. An item can receive activation because it is mentioned in the text or because it is somehow related to an item that is activated. Which items receive activation, and how much, can be chosen by the modeler to implement a particular theory of text processing or differences in reading goal.

In Linderholm, Virtue, Tzeng, & Van den Broek's (2004) simulation of the difference between reading for study and reading for entertainment, the difference between the two reading purposes is assumed to lie in the standard of coherence readers attempt to obtain. When reading for entertainment, only explicit and referential relations are identified in the text, while a study purpose also requires finding causal, logical, and contrastive relations (Linderholm et al., 2004, p. 173). Note that the differences in these standards of coherence entail a difference in required world knowledge.

To implement this, the explicit, referential, causal, logical, and contrastive relations in a text were analyzed. The choice of input activation was based on the relations that were found: Simulating the processing of a sentence from the text involved activating the propositions from earlier in the text that were explicitly or referentially related to the sentence being processed. In the study-purpose simulation, but not in the entertainment-purpose simulation, input activation was also provided to earlier propositions that had a causal, logical, or contrastive relation to the current sentence. As a result, more input activation is present in the study-purpose simulation, which, of course, means that there is more overall activation. According to Linderholm et al. (2004), "This finding suggests that,

in the computational simulation ... reading for study results in the activation of more elements throughout reading" (p. 173). However, this so-called finding is actually the manipulation of parameters (i.e., which propositions receive input activation depending on the task being simulated). The same is true for the second finding from this simulation, which is that propositions causally related to the current sentence receive activation in the study-purpose simulation but not in the entertainment-purpose simulation (see Figure 1 in Linderholm et al., 2004). Clearly, this difference in activation values between the two simulated tasks is precisely the applied manipulation of input activations rather than a prediction of the model.

EXTRACTING KNOWLEDGE FROM TEXT CORPORA

As argued in the previous section, implementing only text-specific world knowledge comes down to an unacceptable setting of free parameters. It would be better to have one fixed knowledge base that can be applied to many different texts—that is, to implement world knowledge as a setting of fixed, rather than free, parameters.

One method for constructing a generally applicable knowledge base is to compute, from large text corpora, word representations that encode semantic relations among the words. A well known technique that accomplishes this is latent semantic analysis (Landauer & Dumais, 1997) in which each word is represented as a vector in a high-dimensional semantic space. The similarity between two word vectors, measured by the cosine of the angle between them, is a value between -1 and $+1$ that indicates the semantic relatedness of the two words.

LSA has been used to reduce the number of subjective choices in CI-model simulations, and has served as the basis for an implementation of world knowledge to be used by computational accounts of discourse comprehension. However, as we argue, LSA cosines cannot directly replace the hand-picked connection weights in the CI model. More important, only knowledge of word meaning can be extracted from text corpora, whereas sentential meaning is more important to discourse comprehension, for one, because it is crucial for knowledge-based causal inferencing. Although it has been claimed that the information in LSA vectors can be used to drive causal inference (Kintsch, 2001), we show that such claims do not hold.

LSA for CI

Kintsch (1998, 2000a; Kintsch et al., 1999) discussed how hand coding in the CI model can be prevented by using LSA. Instead of subjectively deciding which

knowledge items become associated to the text items during the construction phase, the text items' nearest neighbors in LSA's semantic space can be selected objectively. The cosine between the vectors then becomes the weight of the connection between the corresponding network nodes. Because the LSA vectors do not depend on the particular text or task, this method does away with the ad hoc nature of world knowledge in the CI model.

However, this scheme will not generally result in a useful weight setting. For instance, in the MINT example discussed in our evaluation of the CI model, a negative weight between MINT's associates CANDY and BUILDING was needed to disambiguate the word (see Figure 3). However, the LSA Web Site (*lsa.colorado. edu*) reports a positive cosine of $+.14$ between them. The cosine between MONEY and BUILDING is also $+.14$, so letting LSA determine these connection weights would not have worked in this example. Moreover, LSA would not have come up with the associations MONEY and CANDY at all because they are, respectively, only the 25th and 79th nearest neighbor of MINT. The 10 nearest neighbors of MINT in semantic space are all related to the "money" sense of the word, and the 11th is related to the "herb" sense. The first word that could in any way be associated to the "candy" sense of MINT is CRISP, which is the 17th nearest neighbor. This shows that the number of associates a text item retrieves from world knowledge needs to be much larger than the CI model assumes.

Another problem with the LSA approach is that it fails when *propositions* need to be retrieved from world knowledge, as was the case in the "lawyer and judge" example of Table 1. Although, as discussed next, vector representations of propositions can be defined given the vectors of the proposition's predicate and arguments, there are no "nearest proposition neighbors" of a text item predefined in semantic space. As a result, choosing the associated world-knowledge propositions remains a setting of free parameters, again involving the risk that choices are made with the desired result in mind.

The standard approach for computing the LSA vector representation of a sentence or text is to simply take the (weighted) sum of the vectors of its individual words (e.g., see Foltz, Kintsch, & Landauer, 1998). However, because this method regards a sentence as a collection of individual words, it cannot account for sentential meaning in general. If the same approach would be applied to propositions, it would be impossible to take into account the different roles of predicates and arguments. The Predication model (Kintsch, 2001) offers a possible solution to this problem.

The Predication Model

Simulated process.   The Predication model computes a vector representation of a proposition given the LSA vectors of the proposition's predicate and arguments, but extends LSA's simple summing rule by not treating predicate

and arguments equally. More important, the resulting proposition vectors form an implementation of world knowledge that can be used for processing any text. The model itself does not simulate any cognitive process but does make claims about the mental representation of discourse statements and world knowledge. These representations have been applied to several aspects of cognition such as judgements of similarity, causal inferencing, and metaphor comprehension (Kintsch, 2000b; Kintsch, 2001; Kintsch & Bowles, 2002).

Constructing proposition vectors.   According to the Predication model, the vector representing a proposition of the form $P(A)$, with predicate P and a single argument A, is computed from the LSA vectors for P and A by first finding a small number of vectors closest to (i.e., having the largest cosine with) P. From those, a subset of vectors closest to A is selected. These vectors and the vectors for P and A themselves are summed to form the vector representing $P(A)$ (Kintsch, 2001, pp. 180–181). Vectors representing a proposition of the form $P(A_1, A_2)$, carrying two arguments, can be computed similarly (Kintsch, 2001, pp. 189–190).

One important thing to note is that the predicate P has a larger impact on the proposition's vector than does the argument A (or the two arguments A_1 and A_2). By first selecting the nearest neighbors of P and then selecting from these the vectors closest to A, predication biases the vector sum toward P. As a result, the vector for $P(A)$ usually is closer to P than to A.

Interpreting vectors.   To investigate the meaning of a proposition vector, it is compared to several so-called "landmark" vectors, which are the LSA vectors of words considered to be related to the proposition's meaning. For the propositions RAN(HORSE) and RAN(COLOR), Kintsch (2001, Table 2) used as landmarks the vectors for GALLOP and DISSOLVE. The first of these landmarks is closer to RAN(HORSE), whereas the second has a larger cosine with RAN(COLOR). It is concluded that the predication algorithm correctly resulted in a vector for RAN(HORSE) that "is like" GALLOP, and a vector for RAN(COLOR) that "is like" DISSOLVE. This result shows how the meaning of propositions is reflected in the distance between their vector representations, as was the case for LSA's word vectors.

World Knowledge in the Predication Model

The Predication model has access to the knowledge of word semantics that LSA extracted from text corpora. For discourse comprehension, however, *causal* knowledge is at least as important because causal inferences are often required to understand a text. Indeed, the Predication model has also been applied to causal inferencing, for example, to infer that the statement *The student washed*

the table implies that as a consequence *The table was clean*. If predication picks up this causal relation, the vector for WASHED(STUDENT, TABLE) should have a larger cosine with CLEAN(TABLE) than with, for instance, CLEAN(STUDENT). Kintsch (2001, Table 1) showed that this is indeed the case. More important, the difference between the two cosines is less when the sum of just the vectors STUDENT, WASHED, and TABLE is chosen to represent *The student washed the table*. This is taken as evidence that, for causal inferences, the Predication model performs better than LSA's method of simply summing the vectors of the words involved.

At first glance, predication seems to solve the world-knowledge implementation problem: Knowledge that is extracted from large text corpora can be applied generally, so computational models can use it as fixed parameters— for instance, for a task as crucial to text comprehension as causal inferencing. However, it turns out the alleged demonstrations of causal inference are not what they seem. Because interpreting the model's output requires comparing it to landmark vectors, the interpretation depends strongly on the landmarks that are chosen. By taking into account that proposition vectors are biased toward the predicate (in the previous example, the verb WASHED), it can be shown that in all nine cases of causal inference presented by Kintsch (2001), the particular choice of landmarks was responsible for incorrectly interpreting the model's outcome as reflecting causal relations.

The last column of Table 2 shows that the cosine between the words WASHED and TABLE ($cos = .25$) is much larger than between WASHED and STUDENT ($cos = .02$). Biasing the proposition vector toward the verb therefore results in a vector closer to TABLE than to STUDENT, and closer to CLEAN(TABLE) ($cos = .83$) than to CLEAN(STUDENT) ($cos = .62$). If the landmarks *The table was*

TABLE 2
Cosines Between Four Test Propositions and Landmark Propositions,
According to Kintsch (2001, Table 3); and Cosines Between the
Test Proposition's Predicate and the Landmarks' Arguments, According to the
Latent Semantic Analysis Web Site at *lsa.colorado.edu*

Proposition (pr)	Landmarks (lm)	cos(pr, lm)	$cos(P_{pr}, A_{lm})$
WASHED(STUDENT, TABLE)	CLEAN(STUDENT)	.62	.02
	CLEAN(TABLE)	.83	.25
DROPPED(STUDENT, GLASS)	BROKEN(STUDENT)	.87	.10
	BROKEN(GLASS)	.91	.23
DRANK(DOCTOR, WATER)	THIRSTY(DOCTOR)	.83	.05
	THIRSTY(WATER)	.78	.26
SHOT(HUNTER, ELK)	DEAD(HUNTER)	.73	.41
	DEAD(ELK)	.70	.24

TABLE 3
Cosines Between Five Landmark Propositions and Test Propositions, According to Kintsch (2001, p. 192); and Cosines Between the Predicates of the Test Propositions and the Landmark, According to the Latent Semantic Analysis Web Site at *lsa.colorado.edu*

Test Propositions (pr)	Landmark (lm)	cos(pr, lm)	$cos(P_{pr}, P_{lm})$
TOOK(SARAH, ASPIRIN)	WENT(PAIN, AWAY)	.89	.74
FOUND(SARAH, ASPIRIN)		.47	.39
EXPLODED(HARRY, PAPER BAG)	JUMPED(HE, in ALARM)	.32	.36
INFLATED(HARRY, PAPER BAG)		.28	.07
SHOT(HIKER, DEER)	DIED(DEER)	.74	.34
AIMED(HIKER, at DEER)		.56	.18
SCRUBBED(TED, POT)	SHONE(POT, BRIGHTLY)	.45	.25
FOUND(TED, POT)		.41	.19
LOST(CAMPER, KNIFE)	SAD(CAMPER)	.48	.37
DROPPED(CAMPER, KNIFE)		.37	.27

tired and *The student was tired* are chosen to test causal inferences, the model incorrectly concludes that washing the table is more tiring for the table than it is for the student: We found a cosine between WASHED(STUDENT, TABLE) and landmark TIRED(STUDENT) of .66, compared to a cosine with landmark TIRED(TABLE) of .81.

In three out of four cases in Table 2, the landmark closest to the test proposition was the one whose argument was closest to the proposition's predicate.[5] Only in the example DRANK(DOCTOR, WATER) did this not occur: Although DRANK is closer to WATER than to DOCTOR, the model seems to have concluded correctly that it was the doctor, and not the water, who must have been thirsty. However, changing the landmarks shows that, again, the interpretation of the model's output depends critically on the landmarks chosen. Drinking the water will refresh the doctor and not the water, but we found a much smaller cosine of DRANK(DOCTOR, WATER) with REFRESHED(DOCTOR) (cos = .38) than with REFRESHED(WATER) (cos = .69).[6]

Kintsch (2001, p. 192) gave five more examples of causal inference by the Predication model. In each of these examples, shown in Table 3, there was

[5]Note that the Predication model does not make the correct causal inference in the last example: The vector representing *The hunter shot the elk* is closer to *The hunter is dead* than to *The elk is dead*.

[6]In constructing the vector for REFRESHED(DOCTOR), latent semantic analysis came up with the following nearest neighbors: OCASSION [*sic*], HOGARTHIAN, GETHSEMANE, CHINOISERIE, and CARISSIMA. All of these occur only once in the corpus that was used to construct the semantic space, so their vectors have an extremely short length (0.03). Replacing them with the next five, more acceptable, words—SLEEPING, AWAKE, SOUNDLY, EVENING, and SLEEP—resulted in a vector for REFRESHED(DOCTOR) that had a cosine of .42 with DRANK(DOCTOR, WATER).

only one landmark (e.g., *The pain went away*), which was compared to two test propositions varying only in the predicate. One of these formed a direct cause for the landmark (e.g., *Sarah took the aspirin*), and the other did not (e.g., *Sarah found the aspirin*). In all five cases, Kintsch (2001) found that the causally related proposition was closer to the landmark than the non-related proposition. However, we found that all these cases could be explained by simply looking at the cosine between predicates of the test propositions and of the landmark (see Table 3). The landmark's predicate is always closer to the predicate of the causally related test proposition than to the other predicate. For instance, WENT is closer to TOOK (cos = .74) than to FOUND (cos = .39). Because a proposition's vector is usually closest to its predicate, this explains why these five landmark propositions are closer to the causally related propositions than to less related propositions.

Conclusion

Using LSA and the Predication model results in a knowledge implementation that is fixed and can be applied to many different texts. However, understanding text requires understanding its propositions, and there is an important difference between words and propositions that the Predication model ignores: Unlike words, propositions are statements to which a truth value can be assigned. Causal inferencing is based on such truth values and, therefore, requires propositions: If *it is true* that Sarah took the aspirin, then *it is likely* that the pain will go away. LSA constructs a semantic space in which the cosine between vectors is a measure for the semantic relatedness between the corresponding words. If the Predication model could indeed turn such semantic relations into causal relations, that would be a spectacular result. However, predication takes the LSA vectors and constructs from them other vectors *in the same space*. Relations that depend on truth values (such as causal relations) do not correspond to any measure in the semantic space. This is easily shown by comparing the vectors for CLEAN(TABLE) and its negation NOT(CLEAN(TABLE)). These vectors are very similar (their cosine is .91), showing that they do not represent the opposite meanings of the propositions. Instead, predication vectors correspond to words adjusted to a specific context. The vector that is claimed to represent the sentence *The horse runs* should actually be interpreted as meaning something like *horselike running* because it takes the verb *runs* and adjusts it to the context of *horse*.

This point is also made clear by results reported by Burgess, Livesay, and Lund (1998), who developed another method for automatically constructing vector representations of word meaning from text corpora: the Hyperspace Analog to Language (HAL) model. Burgess et al. (pp. 234–236) constructed vectors representing sentences by summing the sentence's word vectors (weighted by

inverse word frequency). These sentences described goals and corresponding means, such as *be less sarcastic* (goal) by *thinking twice before I say anything* (means). A comparison of the distances among vectors for goals and means revealed that knowledge about their relation was not captured in their vector representations. Because such knowledge is a kind of causal knowledge (means can cause a goal to be reached), this negative result is exactly what could be expected considering that HAL vectors encode information about words only. Knowledge about the causal relation between statements does not simply follow from the meaning of the individual words in these statements.

World knowledge about causality is of major importance to understanding text, so models of this process require an implementation of causal knowledge. Although LSA and HAL manage to encode a large amount of realistic knowledge, they do not represent causal knowledge, limiting their value to computational models of discourse comprehension. This is not just a technical problem with LSA and HAL, which could be overcome by using some other method. Such methods cannot extract causal knowledge from text corpora because *world* knowledge does not follow from *word* knowledge. Nevertheless, world knowledge can be obtained by a similar "extraction" strategy, as we discuss next.

THE MICROWORLD STRATEGY

Knowledge about the world cannot be automatically extracted from text corpora, so this question arises: What sort of corpus is needed for this? For such a scheme to work, the corpus should not consist of texts but of events or situations in the world. Unlike large and general corpora of naturally occurring texts, however, corpora of real-world events are not readily available. Therefore, event corpora need to be constructed from scratch. As a result, the knowledge that can be encoded is at best that of a tiny subset of the real world: a *microworld*.[7]

A model that uses this microworld strategy can process only texts dealing with the microworld. In contrast to models that use only text-relevant knowledge, however, microworld-based models can process several texts without requiring any parameter adjustment because *all* knowledge about the microworld is encoded in *fixed* parameters. Such models can be run on different texts while making sure that positive results are not simply caused by the modeler's interference.

---

[7]Note that, in principle, microworld knowledge could also be implemented directly—that is, without extracting it from an event corpus. However, we do not know of any discourse-comprehension model that obtains its world knowledge in such a way.

One model that makes use of microworld knowledge extracted from an event corpus is the Distributed Situation Space (DSS) model (Frank, Koppen, Noordman, & Vonk, 2003). As we hope to show from our following discussion of that model, using the microworld strategy leads to models whose properties can be thoroughly investigated.

The DSS Model

Simulated process.   The DSS model simulates the same process as the CI model: the knowledge-based drawing of inferences from a text. Nevertheless, the two models view the inference process very differently. Whereas CI simulates inference as activation and selection of world knowledge, DSS treats it as a process of pattern completion.

The microworld.   The microworld is very simple, having only two characters, called Bob and Jilly, who can engage in a small number of activities and be in a small number of states. For instance, it can be the case that "Bob is outside," "Bob and Jilly play soccer," or "Jilly wins." In total, 14 so-called "basic propositions" (listed in Table 4) are needed to describe any situation that can occur in the microworld. The number of possible microworld situations is much larger because the basic propositions can be combined using the boolean operators of negation, conjunction, and disjunction.

TABLE 4
Basic Propositions and
Their Intended Meanings

Label	Meaning
SUN	The sun shines.
RAIN	It rains.
B_OUT	Bob is outside.
J_OUT	Jilly is outside.
SOCCER	Bob and Jilly play soccer.
HIDE	Bob and Jilly play hide-and-seek.
B_COMP	Bob plays a computer game.
J_COMP	Jilly plays a computer game.
B_DOG	Bob plays with the dog.
J_DOG	Jilly plays with the dog.
B_TIRED	Bob is tired.
J_TIRED	Jilly is tired.
B_WINS	Bob wins.
J_WINS	Jilly wins.

Of course, some situations are more likely to occur than others. For instance, it is more likely that Bob and Jilly are both outside than that only one of them is. Some situations are even impossible. For instance, it never happens that Bob or Jilly wins while they are not playing a game. In addition to these differences in the likelihood of situations, there are also constraints regarding temporal sequences of situations. For instance, when someone is tired, the other is more likely to win the game they are playing. Also, when Bob and Jilly are tired, they are not very likely to start playing soccer.

World-knowledge representation.   Just like LSA vectors encode word knowledge based on the words' co-occurrences in texts, vectors in the DSS model encode world knowledge based on the events' co-occurrences in the microworld. The event corpus is a hand-crafted sequence of 250 microworld situations, which abides by the microworld's probabilistic constraints. Each of the situations is represented as a 14-element binary vector, with a 0 for each basic proposition that is not the case in the situation, and a 1 for each basic proposition that is. Next, a self-organizing map (SOM; Kohonen, 1995) is trained on these 250 vectors. As a result, each basic proposition comes to be represented by a pattern of activation over the 150 cells of the SOM, or equivalently, as a 150-element vector. Using well-known equations from fuzzy logic, these vectors can be combined into representations of any microworld situation. For this reason, the vectors are called "situation vectors."

Situations that are alike result in vectors that are alike. In practice, this means that from any situation vector X, the probability that any of the 14 basic propositions is the case can quite accurately be estimated by comparing X to the vector representation of the desired basic proposition. Such an estimated probability, or *belief value*, can also be computed for any combination of basic propositions—that is, for any microworld situation.

Knowledge about the temporal constraints between microworld situations is also extracted from the event corpus. This knowledge is encoded in a 150×150 matrix W.

The inference process.   A story is a description of a temporal sequence of situations occurring in the microworld. Usually, this description is incomplete: A story does not state *everything* that does and does not happen. The model's task is to take as input the vectors representing the incomplete story situations and infer what else is likely to be the case at each moment in the story.

Situations in the microworld are assumed to follow one another in discrete time steps, indexed by t. The vector representing the situation at some time step t is denoted X_t, so the input to the model is the sequence of vectors $\langle X_1, X_2, \ldots \rangle$. This is called the story's *trajectory* through situation space.

The situation vectors enter the model one at a time. After a new vector is added, the temporal world-knowledge matrix W is applied to the story trajectory so far. The DSS model adopts the mathematical basis from Golden and Rumelhart's (1993) story-comprehension model to compute the most likely trajectory given the current trajectory and matrix W. The trajectory is adapted toward this most likely trajectory, such that the vectors come to represent situations that are more fully specified. This simulation of the inference process adds information to the original, incomplete specifications of situations.

The process halts when the rate of change of the trajectory falls below a particular level, controlled by a "depth-of-processing" parameter. At that moment, the next story situation vector enters the model. The variable amount of processing time needed to update the trajectory serves as the model's analog of sentence reading time.

The model in practice.   Table 5 shows an example of a simple story. It consists of four consecutive situations. First, Bob and Jilly are stated as being outside. Next, it is claimed that they are playing together. In the microworld, this means that they either play soccer, hide-and-seek, both play a computer game, or both play with the dog. Because they were outside, it is unlikely that they play a computer game (it is not impossible, however, because they might have moved inside by $t = 2$). Integrating X_1 and X_2 should therefore result in an updated vector X_2 that represents a situation in which B_COMP $\wedge$ J_COMP is unlikely. Contrary to this, the information that Bob and Jilly are outside increases the likelihood of SOCCER, which is always played outside, but does not affect the probability of HIDE because hide-and-seek can be played inside or outside. Next, the story states that Bob or Jilly wins. This means that they could not have been playing with the dog at $t = 2$. Integrating X_2 and X_3 should therefore decrease the belief value of B_DOG $\wedge$ J_DOG in X_2. The story ends by stating that Bob and Jilly are not tired.

The result of processing the story of Table 5 is shown in Figure 4. When the first two story situations are integrated, the belief value of B_COMP $\wedge$

TABLE 5
A Sequence of Situations ("Story") Processed by the Distributed Situation
Space Model and a Corresponding Story Text

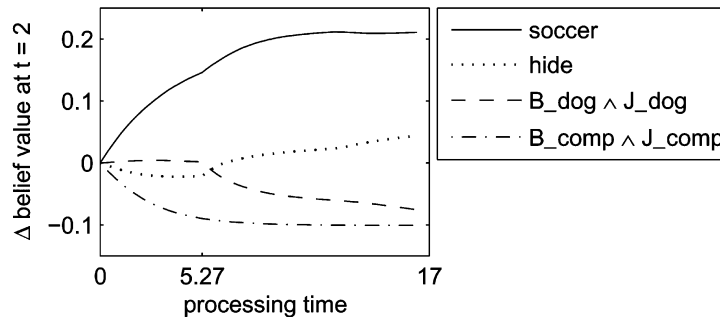| t	Microworld Situation	Possible Text
1	B_OUT $\wedge$ J_OUT	Bob and Jilly are outside.
2	SOCCER $\vee$ HIDE $\vee$ (B_COMP $\wedge$ J_COMP) $\vee$ (B_DOG $\vee$ J_DOG)	They are playing together.
3	B_WINS $\vee$ J_WINS	One of them wins.
4	$\neg$B_TIRED $\wedge$ $\neg$J_TIRED	They are not tired.

FIGURE 4   Result of processing the story of Table 5 by the Distributed Situation Space model. *Note*. Plotted is the change in belief values for four games at story time step $t = 2$, as a function of the model's processing time. Integration of the first two situations begins at processing time 0. At time 5.27, the third situation is added.

J_COMP at $t = 2$ decreases, and the belief value of SOCCER increases because of the information that they are outside. After 5.27 units of processing time, integration has stabilized enough to allow the third situation to enter the model. At this point, the belief value of B_DOG $\wedge$ J_DOG at $t = 2$ decreases because of the information that the game they played has a winner. In short, whenever a new story statement is included in the model, it makes the correct inferences by updating its representation of the story's situations in accordance with its temporal knowledge of the microworld.

World Knowledge in the DSS Model

The DSS model can process any story taking place in the microworld without the need for setting free parameters because the same world knowledge and other parameter values are used in many simulation runs. The model does have many knowledge-implementing parameters (i.e., the situation vectors and matrix W), but these are all fixed. This means that the pattern of results over many texts can be relied on to indicate properties of the model and not just of the parameter setting. For instance, Frank et al. (2003) showed that the model's processing time is lower for stories that are more coherent, which is in accordance with experimental data (Golding, Millis, Hauselt, & Sego, 1995; Myers, Shinjo, & Duffy, 1987). Because this result is based on processing many different stories and the parameter setting was the same for each of these, the predicted relation between coherence and processing time reveals an actual property of the model rather than the accidental outcome of some parameter setting.

As was the case for LSA's semantic vectors, the DSS model's situation vectors and temporal knowledge matrix result from an unsupervised training process—that is, the knowledge base is task-independent and can be probed in many different ways. Indeed, the DSS model's world-knowledge implementation has not only been applied to inferencing but also to modeling story retention (Frank et al., 2003) and ambiguity resolution (Frank, Koppen, Noordman, & Vonk, 2007). In these simulations, the model successfully predicted experimental data concerning free recall rates, reading times, and disambiguation errors.

Of course, the model also suffers from several limitations. For instance, it cannot infer at which story time steps situations take place so these must enter the model in chronological order. Also, the model does not simulate how text processing can lead to updated world knowledge. Moreover, it lacks an attention mechanism as well as an account of working-memory limitations. Note, however, that these shortcomings are specific to the model rather than resulting from the microworld approach to world-knowledge implementation.

Conclusion

Extracting world knowledge from a description of events in the world and implementing all of this knowledge for use by a computational model results in fixed rather than free parameters. Consequently, the model can be tested on several texts without changing the parameter setting, making it possible to investigate the model and its predictions more thoroughly. For this method to be feasible, the world to which the model is applied needs to be simplified considerably, making it a microworld. In contrast to knowledge extracted from text corpora, microworld knowledge deals with the situational rather than the word level, making it more useful for discourse comprehension.

Other microworld models.   When using the microworld strategy, care needs to be taken to make sure that the knowledge base is not too task specific. Two examples of microworld-based models that suffer from this problem are the Story Gestalt model (St. John, 1992) and the DISPAR (Distributed Paraphraser) model (Miikkulainen & Dyer, 1991). Both use recurrent neural networks that are trained on script-based stories, represented as a sequence of predicates and arguments. The models extract microworld (script) knowledge from these training stories.

The training task is different for the two models. Story Gestalt learns to answer specific types of questions about the story, whereas DISPAR is trained to paraphrase the stories it processes. Both these tasks require the drawing of inferences, using background knowledge, when gaps in the stories need to be filled in. The two models manage to learn the task, showing that they apply world knowledge when needed. However, it can be argued that the networks do not actually have any declarative knowledge about the microworld, but only

procedural knowledge that tells them how to answer questions or paraphrase stories. This is because, during the supervised training process, the neural networks adapt to the task they are to perform. The resulting connection weights (i.e., the implementation of world knowledge) do not encode knowledge in general because their only purpose is to be useful for either answering questions or paraphrasing stories. Consequently, the weights are not useful for any other task. For example, Story Gestalt is trained on questions that take the form of a predicate from the story. To answer the question, the network has to give the predicate's arguments. Because the network is never trained to answer a question like "what happened next?," it would need to be retrained completely if it is to answer such questions.

Possible drawbacks of the microworld strategy.   A question that arises when using the microworld strategy is whether the model can be scaled up to more realistic amounts of knowledge. To be regarded a viable cognitive model, it should be able to handle amounts of knowledge much larger than that of a small microworld. To what extent the DSS model can be scaled up is still an open question. This question is rarely posed when discussing models that only use text-specific knowledge, presumably because the presence of additional world knowledge is simply ignored. Nevertheless, the issue matters to such models as well because the amount of knowledge that is relevant to the text may be very large indeed. Using the microworld strategy forces the modeler to design the model to always take all existing knowledge into account. As a result, the scalability issue becomes apparent more easily when dealing with a microworld than with text-specific knowledge only. This may actually be an advantage of the microworld strategy, rather than a drawback.

Another issue for models that have only microworld knowledge is that they cannot be used to process texts dealing with events outside the microworld, such as most natural texts. This incapability of using natural text input seems like a major limitation because it makes detailed, quantitative comparisons between simulation results and experimental data impossible. However, qualitative comparisons can still indicate whether the model adequately simulates human discourse-comprehension processes. For example, without predicting reading times for realistic texts, the DSS model provides an account of the relation between text coherence and reading times. Moreover, it can be argued that models that only include text-specific knowledge (or none at all) do not process natural texts either because a text can only be considered natural when embedded in a large amount of world knowledge. For example, when some network node in a CI simulation is labeled SEND(JUDGE, DEFENDANT, PRISON), it is this label that activates our knowledge of the legal system, making it seem *to us* as if it represents a bit of natural text. To the model, however, there is no label. Within a simulation, a node can only acquire meaning to the extent that it is

connected to nodes representing world knowledge. If these are (largely) absent, the node remains meaningless and the text it represents is anything but natural.

To summarize, drawbacks of the microworld strategy apply as well to models that rely on free parameters. First, upscaling is an important open issue for *all* models of discourse comprehension. It is just harder to ignore when using a microworld. Second, although microworlds may provide only a small-scale embedding in world knowledge, an ad hoc selection of knowledge for each individual text amounts to an even smaller, disposable "nanoworld" instead of one reusable microworld.

DISCUSSION

Strategies for World-Knowledge Implementation

We identified four different strategies for dealing with world knowledge in computational models: disregarding world knowledge, treating it as a free parameter setting, treating it as a fixed parameter setting obtained from text corpora, and treating it as a fixed parameter setting obtained from event corpora.

Obviously, world knowledge is only left out of models to simplify them— that is, it is acknowledged that the simulated cognitive process is, in reality, embedded in world knowledge. Such a simplification comes with the risk that the model's algorithm only functions without world knowledge because the presence, and possible influence, of knowledge was not taken into account when the model was designed. In that case, it cannot be an adequate description of the cognitive process. We have shown this to be the case for the Resonance model, whose simulation of the "true" resonance process is ruined by the addition of knowledge about semantic relations, although such knowledge is used by the cognitive process.

Leaving knowledge out of the model makes it impossible to process texts for which some piece of knowledge is required. This can be solved by inserting bits of knowledge as needed. Whenever the knowledge that is implemented depends on the particular text being processed, world knowledge is treated as a setting of free parameters. The underlying idea is that any text activates or requires only a small amount of world knowledge, which is all that needs to be implemented. Of course, the problem is that nobody knows which particular bit of world knowledge becomes activated when processing a certain text. In practice, therefore, the modeler implements an ad hoc and subjective selection of world knowledge. Making this selection often comes down to performing the task that the model is supposed to simulate, resulting in correct model outcomes that nevertheless do not allow for any claims about the performance of the model itself. We have shown this to occur in simulations using the CI,

Landscape, and Langston and Trabasso (1999) models. This is not because of the models themselves but because of the presence of free parameters. Whenever a modeler has the freedom to steer each simulation individually by adjusting free parameters, it becomes unclear if the model's outcomes reflect its own internal processes or mainly those of the modeler.

If a discourse-comprehension model is to be validated against experimental data, or if its properties are to be rigorously investigated, it has to be applied to a large number of texts. Implementing a different subset of world knowledge for each of these comes down to setting many free parameters, making it impossible to draw conclusions about the model itself. We have seen an example of this in the Resonance model: Its failure on the captain text with semantic knowledge included indicates that its success on the CHURCH–BARN reinstatement depended critically on the inclusion of no more than two specific connections. In such a way, this "success" actually hides a shortcoming of the model. What is needed for ascertaining a model's general properties is a fixed parameter setting—that is, a single knowledge base to be applied to several texts.

By automatically extracting knowledge about word meaning from large text corpora, LSA and HAL construct meaningful and widely applicable vector representations of words. These vectors can serve as a fixed knowledge base, preventing the problems caused by the use of free parameters. However, when it comes to discourse comprehension, the meaning of propositions is more important than that of words and causal relations are at least as relevant as semantic relations. Such more sophisticated knowledge does not follow from mere word co-occurrences. Claims to the contrary turned out not to be warranted: The Predication model's alleged handling of causal inference was based on the misinterpretation of its output vectors, resulting from the particular choice of landmark vectors to which the model's output was compared.

Computational models of discourse comprehension are faced with the dilemma that no small subset of world knowledge can *a priori* be marked as "relevant to the text," in particular when dealing with narratives, whereas the implementation of (a substantial subset of) true human knowledge is far from feasible. By implementing knowledge about a small and clearly defined microworld, and restricting the input texts to those describing events taking place in this microworld, the introduction of free parameters is prevented. This is the strategy applied by the DSS model. As we have argued, its results do not crucially depend on a clever setting of free parameters—that is, on text-specific and subjective knowledge selection and implementation. Consequently, these results can be relied on to give information about the properties of the model.

Because the microworld strategy is not a model in itself, it is certainly not sufficient for successful modeling. Our claim is merely that it is a necessary approach to discourse-comprehension modeling as long as relevant amounts of human "macroworld" knowledge cannot be implemented.

Applying the Microworld Strategy to Other Models

The modeling problems we identified are first of all caused by the handling of world knowledge, and less so by the models themselves. In some cases, the microworld strategy may very well be applicable to models that have so far used some other method. If this is indeed possible, it is an appropriate step for testing and improving the models.

As we have previously argued, the Resonance model is not likely to be successful if it has to deal with larger networks; so simply plugging knowledge into Resonance will not work. However, the microworld strategy may help in developing an improved resonance algorithm that can deal with larger and more densely connected networks—that is, networks that include relevant amounts of semantic knowledge.

Other models may be better than the Resonance model at dealing with the relatively large amount of world knowledge that is implemented in the microworld strategy. The CI model, for instance, can quite easily be supplied with knowledge of a microworld. Of course, this is not implemented as a DSS-like situation space but should take the form of a large network of concepts and propositions. During the construction phase, many knowledge items could then be associated to the text items in a context-independent fashion, without any interference by the modeler. Also, connection weights in the microworld-knowledge network would be fixed, doing away with any free parameters. In fact, this is how the model is originally presented: Kintsch (1988, pp. 166–167) assumed that the reader's world knowledge is stored in a "knowledge net" consisting of concepts and propositions connected by weighted links. The probability that a particular knowledge item becomes associated to a text item is proportional to the connection weight between these two items in the knowledge net, and weights in the enriched text base are inherited from weight in the knowledge net. However, these ideas are not put into practice when the CI model is used. The reader's knowledge net is, of course, unknown and not implementable; so associated knowledge items and connection weights are chosen freely by the modeler. If, instead, a microworld knowledge net would have been implemented, it could have been used for processing several texts without changing the parameter setting. For any model of discourse comprehension, such a strategy is necessary to reveal the model's general properties.

ACKNOWLEDGMENT

REFERENCES

Albrecht, J. E., & Myers, J. L. (1991). Effects of centrality on retrieval of text-based concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 932–939.

Albrecht, J. E., & Myers, J. L. (1995). Role of context in accessing distant information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1459–1468.

Burgess, C., Livesay, K., & Lund, K. (1998). Explorations in context space: Words, sentences, discourse. *Discourse Processes*, *25*, 211–257.

Caillies, S., & Denhière, G. (2001). The interaction between textual structures and prior knowledge: Hypotheses, data and simulations. *European Journal of Psychology of Education*, *16*, 17–31.

Caillies, S., Denhière, G., & Kintsch, W. (2002). The effect of prior knowledge on understanding from text: Evidence from primed recognition. *European Journal of Cognitive Psychology*, *14*, 267–286.

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.

Doane, S. M., & Sohn, Y. W. (2000). ADAPT: A predictive cognitive model of user visual attention and action planning. *User Modeling and User-Adapted Interaction*, *10*, 1–45.

Doane, S. M., Sohn, Y. W., McNamara, D. S., & Adams, D. (2000). Comprehension-based skill acquisition. *Cognitive Science*, *24*, 1–52.

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, *25*, 285–307.

Frank, S. L., & Haselager, W. F. G. (2006). Robust semantic systematicity and distributed representations in a connectionist model of sentence comprehension. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the Cognitive Science Society* (pp. 226–231). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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. (2005). Predicting without modeling: A critique of Trabasso and Bartolone. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 374–377.

Frank, S. L., Koppen, M., Noordman, L. G. M., & Vonk, W. (2007). Coherence-driven resolution of referential ambiguity: A computational model. *Memory & Cognition*, *35*, 1307–1322.

Golden, R. M., & Rumelhart, D. E. (1993). A parallel distributed processing model of story comprehension and recall. *Discourse Processes*, *16*, 203–237.

Golding, J. M., Millis, K. M., Hauselt, J., & Sego, S. A. (1995). The effect of connectives and causal relatedness on text comprehension. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 126–143). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review*, *95*, 163–182.

Kintsch, W. (1992). How readers construct situation models for stories: The role of syntactic cues and causal inferences. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes, Volume 2* (pp. 261–278). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge, England: Cambridge University Press.

Kintsch, W. (2000a). The control of knowledge activation in discourse comprehension. In W. J. Perrig & A. Grob (Eds.), *Control of human behavior, mental processes, and consciousness: Essays in honor of the 60th birthday of August Flammer* (pp. 137–146). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kintsch, W. (2000b). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, *7*, 257–266.

Kintsch, W. (2001). Predication. *Cognitive Science*, *25*, 173–202.

Kintsch, W., & Bowles, A. R. (2002). Metaphor comprehension: What makes a metaphor difficult to understand? *Metaphor and Symbol*, *17*, 249–262.

Kintsch, W., Patel, V. L., & Ericsson, K. A. (1999). The role of long-term working memory in text comprehension. *Psychologia*, *42*, 186–198.

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 Bennet B. Murdock* (pp. 367–385). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kintsch, W., Welsch, D. M., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, *29*, 133–159.

Kitajima, M., & Polson, P. G. (1995). A comprehension-based model of correct performance and errors in skilled, display-based, human–computer interaction. *International Journal of Human–Computer Studies*, *43*, 65–99.

Kohonen, T. (1995). *Self-organizing maps.* Berlin: Springer.

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Langston, M. C., & 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: Lawrence Erlbaum Associates, Inc.

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.

Luna, D. (2005). Integrating ad information: A text-processing perspective. *Journal of Consumer Psychology*, *15*, 38–51.

Mannes, S. M., & Kintsch, W. (1991). Routine computing tasks: Planning as understanding. *Cognitive Science*, *15*, 305–342.

Miikkulainen, R., & Dyer, M. G. (1991). Natural language processing with modular pdp networks and distributed lexicon. *Cognitive Science*, *15*, 343–399.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes*, *26*, 131–157.

Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, *26*, 453–465.

O'Brien, E. J. (1987). Antecedent search processes and the structure of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 278–290.

O'Brien, E. J., Albrecht, J. E., Hakala, C. M., & Rizzella, M. L. (1995). Activation and suppression of antecedents during reinstatement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 626–634.

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in text: What readers believe versus what they read. *Psychological Science*, *3*, 229–235.

Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation models and aging. *Psychology and Aging*, *16*, 145–160.

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.

Singer, M. (1996). Comprehending consistent and inconsistent causal text sequences: A construction–integration analysis. *Discourse Processes*, *21*, 1–21.

Singer, M., & Halldorson, M. (1996). Constructing and validating motive bridging inferences. *Cognitive Psychology*, *30*, 1–38.

St. John, M. F. (1992). The story gestalt: A model of knowledge-intensive processes in text comprehension. *Cognitive Science*, *16*, 271–306.

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: Lawrence Erlbaum Associates, Inc.

Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition*, *16*, 283–298.

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 & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Weeber, M. (1996). *On modeling text processing*. Unpublished master's thesis, University of Nijmegen, Nijmegen, The Netherlands.

Yang, M., Roskos-Ewoldsen, B., & Roskos-Ewoldsen, D. R. (2004). Mental models for brand placement. In L. J. Shrum (Ed.), *The psychology of entertainment media: Blurring the lines between entertainment and persuasion* (pp. 79–98). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.