Coherence-driven resolution of referential ambiguity: A computational model

STEFAN L. FRANK
Tilburg University, Tilburg, The Netherlands
and Radboud University Nijmegen, Nijmegen, The Netherlands

MATHIEU KOPPEN
Radboud University Nijmegen, Nijmegen, The Netherlands

LEO G. M. NOORDMAN
Tilburg University, Tilburg, The Netherlands
and Radboud University Nijmegen, Nijmegen, The Netherlands

AND

WIETSKE VONK
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
and Radboud University Nijmegen, Nijmegen, The Netherlands

We present a computational model that provides a unified account of inference, coherence, and disambiguation. It simulates how the build-up of coherence in text leads to the knowledge-based resolution of referential ambiguity. Possible interpretations of an ambiguity are represented by *centers of gravity* in a high-dimensional space. The unresolved ambiguity forms a vector in the same space. This vector is attracted by the centers of gravity, while also being affected by context information and world knowledge. When the vector reaches one of the centers of gravity, the ambiguity is resolved to the corresponding interpretation. The model accounts for reading time and error rate data from experiments on ambiguous pronoun resolution and explains the effects of context informativeness, anaphor type, and processing depth. It shows how implicit causality can have an early effect during reading. A novel prediction is that ambiguities can remain unresolved if there is insufficient disambiguating information.

Since natural language contains many ambiguities, a theory of language comprehension cannot be complete without giving an account of ambiguity resolution. Here, we present a computational model that explains how one type of ambiguity, *referential ambiguity*, is resolved.

A statement is referentially ambiguous if it contains an anaphor (often a pronoun) that has more than one possible referent. Selecting a referent results in one of the possible interpretations, or *readings*, of the statement. For instance, in the short text *Bob saw that Joe was running. He was in a hurry*, the pronoun *he* is ambiguous: It can refer to either Bob or Joe. This means that the second sentence has two readings: Either Bob was in a hurry, or Joe was in a hurry.

Several sources of information can be used to disambiguate a referentially ambiguous sentence or clause. One such source of information is formed by the event(s) described in the previous clause(s)—that is, the context the ambiguous clause appears in. Combined with the reader's general knowledge, context information can cause one of the readings to be considered more sensible than any other. In the example above, the context statement *Bob saw that Joe was running* results in a preference for interpreting the second sentence as "*Joe* was in a hurry," because being in a hurry is known to be a reason to be seen running, rather than a reason to see that someone else is running.

Other sources of information that can be used in disambiguation lead to an a priori (i.e., knowledge-independent) preference for one of the readings of the ambiguous statement. We will refer to this a priori preference as *bias*. In the example above, the fact that Bob is mentioned first will make him more accessible than Joe in the reader's mental representation of the discourse. This so-called first-mention effect (Gernsbacher & Hargreaves, 1988) results in an initial preference for choosing *Bob* as the referent of *he*. In other words, there is a bias for the context-inconsistent reading "*Bob* was in a hurry." For the comprehension process to result in the alternative, context-consistent reading, this bias must be overcome by the application of world knowledge.

S. L. Frank, s.frank@nici.ru.nl

An influential theory of discourse comprehension that also deals with referential expressions is Kamp and Reyle's (1993) discourse representation theory. This formal theory was extended by Gordon and Hendrick (1998) to incorporate the control of referent accessibility (i.e., bias) by structural factors. In their conception, referent accessibility drives the initial interpretation of the ambiguity before the plausibility of the ambiguity's readings is taken into account to select a final interpretation. However, evaluating the plausibility of a reading requires knowledge of the world, which neither Gordon and Hendrick nor Kamp and Reyle incorporate in their theories.

One of the first computational theories of referential disambiguation that takes the application of world knowledge into account is the one by Hobbs (1979), which was developed further by Kehler (2002). According to these theories, disambiguation is a side effect of a general process of coherence establishment. In the example above, increasing the coherence between the two statements comes down to finding the causal relation between running and being in a hurry. When world knowledge about this relation is applied, the ambiguity is resolved. In line with Hobbs's and Kehler's theories, the computational model we present in this article assumes that knowledge-based disambiguation is the result of a more general process that increases coherence. It simulates the cognitive processes involved in applying world knowledge to a referentially ambiguous clause and its context, thereby selecting the most likely reading of the ambiguity. In these simulations, the model allows for an account of how bias and context information interact.

In spite of the conceptual similarity between our model and those by Hobbs (1979) and Kehler (2002), there is an important difference: Hobbs's and Kehler's theories do not (nor do they claim to) include an inference mechanism that is analogous to a psychological process. Consequently, they do not lead to quantitative predictions of experimental data. The model we present here, on the other hand, does have an inference process that leads to predictions that are validated against empirical data, as well as novel predictions. The model predicts how context information and bias affect reading times and error rates and, thereby, gives an explanation of these data. Moreover, by having a parameter controlling depth of processing, the model accounts for the effect of processing depth.

Several models of knowledge-based inference have also been applied to referential disambiguation, but they often have given only limited accounts of empirical data. For instance, The Story Gestalt model (St. John, 1992) and the related, but much more comprehensive, story comprehension model by Miikkulainen (1993) lack a notion of processing time and, consequently, cannot predict reading times. In Kintsch's (1988) construction-integration model, successful simulation requires world knowledge to be selected in advance for each text that is processed. As we have argued elsewhere (Frank, Koppen, Noordman, & Vonk, in press), such an approach makes it difficult to validate a model against empirical data. Moreover, none of these models incorporate depth of processing.

A number of more recent models of ambiguity resolution make use of the metaphor of *attractors in state space*. In these models, the readings of an ambiguity are represented as attractors in a high-dimensional space, pulling toward them a vector representing the current interpretation of the ambiguity. The time needed for this dynamical process to stabilize is taken as a measure of reading time. For instance, in the word recognition model of Rodd, Gaskell, and Marslen-Wilson (2004), vectors in state space represent word meanings, making it a *semantic space*. Since each word sense corresponds to one semantic vector, an ambiguous word corresponds to multiple vectors in semantic space. When such a word is processed, its initial vector representation is a mix of its semantic vectors, all of which serve as attractors toward which the initial vector "falls" until it coincides with a vector representing one of the word's senses. The model by Kawamoto (1993) simulates lexical disambiguation as a similar dynamical process.

As for resolving syntactic ambiguities, McRae, Spivey-Knowlton, and Tanenhaus (1998) present a dynamical model but do not use the state space metaphor. In contrast, the visitation set gravitation (VSG) model (Tabor, Juliano, & Tanenhaus, 1997; Tabor & Tanenhaus, 1999) involves a state space in which (partial) sentence structures are represented. When the model processes a word that can be attached to the current structure in different ways (i.e., the sentence is structurally ambiguous at this point), the initial vector lies in between the attractors that correspond to the possible structures. It then falls toward one of these, resulting in a single interpretation.

Our model, too, views disambiguation as a selection process driven by attractors in some state space. In contrast to the spaces of the Rodd et al. (2004) and VSG models, the state space of our model does not hold any linguistic representations. Instead, vectors in this space represent situations in a simplified story world, making it a *situation space*. The situation vectors are taken from the distributed situation space (DSS) model of inferencing during story comprehension (Frank, Koppen, Noordman, & Vonk, 2003), summarized later in this article.

In the model, referential ambiguity is resolved by performing two subprocesses simultaneously. One is a knowledge- and context-independent attraction process, comparable to the disambiguation processes of the Rodd et al. (2004) and VSG models. The other is a knowledge-based and context-dependent, coherence-increasing, general inference process, implemented in the DSS model. Combining these two processes results in a unified model of inference and disambiguation, called DSS–AR (for *ambiguity resolution*).

Before turning to the model's description, the next section will present empirical data against which the model will be validated. These data come from experiments on the resolution of ambiguous pronouns. In these experiments, there are two possible referents for a pronoun, resulting in two readings of the referential ambiguity. The intended reading can be determined using context information and world knowledge, but bias may also affect which reading is selected. Many such experiments suitable for comparison with our model have been performed, making this a good testing ground for the model. Note,

however, that DSS–AR does not simulate the selection of a pronoun's referent but gives an account of the selection among the ambiguity's readings, because points in its state space represent story situations, rather than separate discourse entities.

Following the presentation of experimental data, two sections will describe the DSS model and its extension for disambiguation, respectively. Next, the setup of simulations will be described, after which their outcomes will be compared with the experimental data and the model's novel predictions will be presented. Finally, the Discussion section will give an interpretation of experimental findings in terms of the model's properties. Also, the model's (novel) predictions will be discussed in relation to several unresolved issues in reading time data—in particular, the timing of the effect of implicit causality and the possibility of fast processing of "unresolvable" ambiguities.

EMPIRICAL FINDINGS

The data discussed below fall into three categories. First, we will look into the time course of pronoun resolution. Second, results regarding sentence reading times will be discussed. Third, we will discuss some findings concerned with errors in pronoun resolution.

The Time Course of Pronoun Resolution

Several studies have indicated that an ambiguous pronoun can already be (partially) instantiated before disambiguating context information is available. Arnold, Eisenband, Brown-Schmidt, and Trueswell (2000) had subjects listen to short texts that brought one of two entities into focus. At the same time, these subjects looked at a picture corresponding to the situation described in the text. Eyetracking revealed that the subjects looked at the focused entity at the moment the text contained an ambiguous pronoun. It was not until somewhat later, when the pronoun could be disambiguated, that they looked at the pronoun's intended referent. It was concluded that focus can drive the initial instantiation of a pronoun (i.e., focus leads to bias for one of the readings) and that the intended reading is selected when context information becomes available. The same conclusion was drawn by Gordon and Scearce (1995) and Vonk (1985) on the basis of reading time data.

Reading Times

One of the phenomena most studied in relation to pronoun resolution is implicit causality, which is the ability of some verbs to bias the interpretation of a following pronoun. After reading *Bob lied to Joe because he . . .* , for instance, the pronoun is commonly taken to refer to Bob, whereas reading *Bob punished Joe because he . . .* leads to the assumption that Joe is the pronoun's referent (Caramazza, Grober, Garvey, & Yates, 1977; Garvey, Caramazza, & Yates, 1974). The verb *to lie to* is a so-called NP1-biasing verb (or NP1 verb for short), meaning that it biases toward the first noun phrase: The cause of lying is implicitly assigned to the liar and not to the person being lied to. The verb *to punish*, on the other hand, is an example

of an NP2 verb because it biases toward the second noun phrase: The cause of punishing is assigned to the punished and not to the punisher.

Implicit causality affects reading times. Congruent sentences are sentences in which the implicit cause agrees with context information, such as *Bob lied to Joe because he could not tell the truth*. These are generally read more quickly than incongruent sentences, such as *Bob lied to Joe because he could not handle the truth* (Garnham, Oakhill, & Cruttenden, 1992; Stewart, Pickering, & Sanford, 2000; Vonk, 1985).

At what point in the sentence does implicit causality affect the comprehension process? According to the so-called *focusing* account, implicit causality arises while the verb is read: An NP1 verb makes the subject more available to the reader, whereas an NP2 verb focuses attention on the object. Contrary to this, the *integration* account claims that implicit causality does not affect comprehension until later, after the explicit cause has been read, at the moment the two clauses of the sentence are integrated. There is disagreement among researchers concerning which account is correct: McDonald and MacWhinney (1995) found support for a focusing effect of implicit causality, whereas Garnham, Traxler, Oakhill, and Gernsbacher (1996) concluded that its effect takes place during integration. Long and De Ley (2000) claimed that it depends on reading skill: Skilled readers show an earlier effect of implicit causality than less skilled readers do.

The results of these three studies were based on probe recognition tasks, which Stewart et al. (2000) criticized for interfering with the normal comprehension process and for not being sensitive enough to the time course of processing. They argued for taking reading time measures instead and claimed that the two accounts make different predictions about the effect of congruency on reading times. First, the focusing account is claimed to predict *no* congruency effect in sentences in which the ambiguous pronoun is replaced by the name of the intended referent, whereas the integration account would predict a congruency effect regardless of whether the anaphor is an ambiguous pronoun or a proper name. Stewart et al. based this claim on the finding that pronouns are more sensitive to focus than are proper names (Garrod & Sanford, 1994; Gordon, Grosz, & Gilliom, 1993). Second, only the integration account would predict a smaller congruency effect when readers engage in shallower processing, because processing depth affects integrative processes but not focusing.

Stewart et al. (2000) conducted three self-paced reading experiments in which subjects read congruent and incongruent sentences containing an ambiguous pronoun and the same sentences with the pronoun replaced by the name of the intended antecedent, rendering the sentence unambiguous. It was found that there was a congruency effect in both conditions, without even an interaction between the congruency and the ambiguity variables. From this result, Stewart et al. concluded that the focusing account of implicit causality is incorrect.

They also investigated the effect of processing depth on the congruency effect by varying questions that the sub-

jects had to answer. In the deep condition, these questions could be answered only if the ambiguity was resolved correctly. In the shallow condition, the ambiguity did not need to be resolved to answer the questions, supposedly resulting in shallower processing. It was found that the congruency effect was smaller in the shallow-processing condition than in the deep-processing condition. Again, it was concluded that the focusing account is incorrect.

Error Rates

Bias does not always agree with context information on the preferred reading of the ambiguity. For instance, in *Bob lied to Joe because he could not handle the truth*, bias favors the reading in which Bob cannot handle the truth, whereas context information favors the other. Either of these sources of information may gain the upper hand during disambiguation. When the context-inconsistent reading is selected, this is considered an error in pronoun resolution. Alternatively, it is possible that the pronoun is not instantiated at all, or only partially (Greene, McKoon, & Ratcliff, 1992; Oakhill, Garnham, & Vonk, 1989).

Leonard, Waters, and Caplan (1997a) showed that the amount of context information affects the chance of making an error. They had subjects read sentences containing an ambiguous pronoun, as well as disambiguating context information. The amount of context was varied by sometimes adding another full sentence that strongly implied one of two readings of the ambiguity. The subjects had to decide as quickly as possible to whom the ambiguous pronoun referred. It was found that adding a context sentence resulted in fewer errors. Hirst and Brill (1980) found a similar effect of context informativeness on error rates.

Congruency between implicit causality and context information not only affects reading times, but also has an effect on error rates. Leonard et al. (1997a, 1997b) and Stewart et al. (2000) found that more errors in referential disambiguation are made when incongruent sentences are read than when congruent sentences are read. Leonard et al. (1997a) also found that this effect becomes weaker when extra context information is added.

THE DISTRIBUTED SITUATION SPACE MODEL

As was discussed in the introduction, Hobbs (1979) and Kehler (2002) have argued that the resolution of referential ambiguity is a side effect of a more general process of establishing coherence among discourse statements by applying world knowledge. The DSS model (Frank et al., 2003) simulates exactly such a process of coherence establishment[1] during story comprehension. It integrates a story's statements with each other, bringing them more in line with knowledge about the world in which the story takes place. This integration process leads to increased coherence and to the inference of information not given in the original story statements. Note that when signaled by textual cues in the form of concessive connectives such as *although* or *nevertheless*, comprehending a story requires finding a *contrast* between the given story and world knowledge. The model does not handle such stories.

The DSS model provides a formal framework for representing situations occurring in a story world and knowledge about that world. This knowledge concerns the likelihood of situations and the likelihood of one situation being followed by another. By being trained on a series of examples of situations occurring in the story world, the model obtains such knowledge.

Central to the model is the notion of a high-dimensional *situation space*. As formalized below, any possible story situation corresponds to a vector in this space. When a story is processed, these so-called *situation vectors* enter the model one by one. This means that DSS processing is incremental at the clause level, but not at a finer resolution (e.g., words or phrases). The integration of story statements comes down to finding a sequence of situation vectors that not only is consistent with the story-given sequence, but also is more likely according to knowledge about the story world. This process, which is modeled as the movement of vectors through situation space, leads to the encoding of additional information about the events taking place in the story—that is, to inferences.

At any time during the model's process, a measure for the amount of inference can be obtained by computing from the situation vectors the probabilities (so-called *belief values*) of particular events occurring in the represented situations. The time needed for the integration process to stabilize is used as the model's correlate of sentence reading times. The criterion for stabilization is manipulated to simulate the effect of processing depth. The model accounts for empirical findings regarding amounts of inference, reading times, and the effect of processing depth (Frank et al., 2003).

In the following subsections, we will discuss the story world to which the DSS model is applied, provide a detailed specification of the model itself, and explain how its outcome is interpreted. Its application to disambiguation will be discussed in the next section.

The Microworld

A major problem for models of discourse comprehension is the implementation of the large amount of world knowledge involved in the comprehension process. One solution is to preselect text-relevant knowledge on an ad hoc basis, meaning that different knowledge is implemented for each text being processed. This makes it difficult to run the model on several texts and statistically assess the relation between its results and experimental data. To make more thorough validation attainable, the DSS model makes use of a so-called *microworld*. This means that world knowledge is selected only once, prior to, and thus independently from, the texts that will be processed. Once all knowledge of the microworld is implemented, the DSS model can apply this general knowledge base on several texts, without requiring any text-specific knowledge selection and implementation.

The microworld used in this article is identical to the one used by Frank et al. (2003). It consists of two characters, named *Bob* and *Joe*,[2] who can engage in several activities, such as playing soccer or a computer game, and can be in different states, such as tired, inside, or outside.

In total, their activities and states can be described using the 14 basic events shown in Table 1.

A story situation is a (full or partial) description of the events occurring at one moment in the story. The number of possible story situations is, of course, much larger than the number of basic events, since these events can be negated and combined using conjunction and disjunction. There are probabilistic constraints on co-occurrences of events. For instance, Bob and Joe can play soccer only when they are outside, and they are more likely to be at the same place than at different places.

The situations of a story are assumed to follow one another in discrete *story time steps*, which are indexed by the variable t. Note that t is not a point in the *text* but a moment in time in the *story*. That is, the situation at time step $t - 1$ takes place before, and is a possible cause of, the situation at t. Likewise, the story situation at $t + 1$ follows the situation at t and is possibly its consequence. Examples of probabilistic constraints on the co-occurrence of situations at adjacent time steps are that someone who is tired at time step t is less likely to win at $t + 1$ and more likely to have been playing soccer than a computer game at $t - 1$.

Specification of the Model

Situation space and microworld knowledge. Microworld situations are represented as vectors in situation space, the number of dimensions of which is fairly arbitrarily set to 150. For any basic event p from Table 1, a vector denoted $\mu(p) = (\mu_1(p), \ldots, \mu_{150}(p))$ (with $0 \leq \mu_i(p) \leq 1$) results from training a self-organizing map (SOM; Kohonen, 1995) on a large number of example situations occurring in the microworld. These examples, constructed by hand, effectuated co-occurrence constraints among events. As a result of SOM training, knowledge about these constraints is encoded in the vector representations (see Frank et al., 2003, for details).

Each of the values $\mu_i(p)$ indicates the extent to which event p is represented in the ith dimension of situation space. In the field of fuzzy logic, such values are known as *membership values*. Using common equations from fuzzy logic, the vector representation of any situation can be constructed from the vectors for basic events, using the

Table 1
Fourteen Basic Microworld Events
and Their Intended Meanings

No.	Name	Meaning
1	Sun	The sun shines.
2	Rain	It rains.
3	B outside	Bob is outside.
4	J outside	Joe is outside.
5	Soccer	Bob and Joe play soccer.
6	Hide-and-seek	Bob and Joe play hide-and-seek.
7	B computer	Bob plays a computer game.
8	J computer	Joe plays a computer game.
9	B dog	Bob plays with the dog.
10	J dog	Joe plays with the dog.
11	B tired	Bob is tired.
12	J tired	Joe is tired.
13	B wins	Bob wins.
14	J wins	Joe wins.

Boolean operators of negation ($\neg p$), conjunction ($p \wedge q$), and disjunction ($p \vee q$):

$$\mu_i(\neg p) = 1 - \mu_i(p),$$

$$\mu_i(p \wedge q) = \mu_i(p)\mu_i(q),$$

and

$$\mu_i(p \vee q) = \mu_i(p) + \mu_i(q) - \mu_i(p)\mu_i(q), \qquad (1)$$

for $i = 1, \ldots, 150$. In this way, not only the basic events, but any microworld situation corresponds to a vector in situation space. For instance, the vector representing the situation described by *Bob and Joe play hide-and-seek outside* can be constructed by applying the conjunction rule of Equation 1 to the vectors $\mu(\text{B outside})$, $\mu(\text{J outside})$, and $\mu(\text{hide-and-seek})$.

From here on, we will use the symbol X to denote any situation space vector $(x_1, \ldots, x_{150})$, which does not need to correspond exactly to a basic or complex event. As an abuse of notation, the same symbol will be used to refer to the situation represented by the vector X. If it is relevant that the situation takes place at time step t, the situation (vector) is denoted X_t. To indicate that the vector represents an *original* story statement—that is, before any integration has taken place—a superscripted 0 is added. For instance, the miniature story over three time steps, *The sun shines. Bob and Joe play soccer. Joe wins.* is represented by the three vectors $X_1^0 = \mu(\text{sun})$, $X_2^0 = \mu(\text{soccer})$, $X_3^0 = \mu(\text{J wins})$.

Knowledge about temporal contingencies in the microworld is encoded in a 150×150-matrix W. The values in this matrix are computed by first constructing the sequence of situation vectors for the microworld example situations that were used as training input to the SOM. The value of element w_{ij} of W indicates the correlation between the values of vector dimensions i at time step t and j at $t + 1$, over all t (see Frank et al., 2003, for details).

Integration and inference. As input, the DSS model takes a sequence of original situation vectors $(X_1^0, X_2^0, \ldots)$, one vector at a time. After a vector has entered the model, all the vectors in the model so far are adjusted to bring them into closer correspondence to the temporal world knowledge encoded in W, simulating a reader's integration process. How exactly a sequence of vectors can be adjusted to increase its probability follows from the DSS model's mathematical basis, Markov random field (MRF) theory.

Given the situations at story time steps $t - 1$ and $t + 1$, it is possible to compute a vector $E_t = (E_{1,t}, \ldots, E_{150,t})$ that represents the expected situation at the intermediate time step t. From MRF theory, Frank et al. (2003) derived how E_t can be expressed in terms of $X_{t\pm1}$ and W.[3] It also follows from MRF theory that the most likely value of each variable $x_{i,t}$ is either its minimum or its maximum value. All minima are 0, and each $x_{i,t}$ has a maximum value $x_{i,t}^{\max} \leq 1$, which is needed to prevent the model from inferring facts that are inconsistent with text-given statements. In the DSS model, each $x_{i,t}^{\max}$ equals the initial value $x_{i,t}^0$.[4]

The integration process drives each variable $x_{i,t}$ toward one of its two extremes, depending on the variable's ex-

pected value, $E_{i,t}$. The process is defined by a differential equation that states how each value $x_{i,t}$ changes over processing time. This is the velocity of $x_{i,t}$, denoted $v_{i,t}$:

$$v_{i,t} = \begin{cases} \left(E_{i,t} - \dfrac{1}{2}\right)\left(x_{i,t}^{\max} - x_{i,t}\right) & \text{if } E_{i,t} > \dfrac{1}{2} \\[2mm] \left(E_{i,t} - \dfrac{1}{2}\right)x_{i,t} & \text{if } E_{i,t} \leq \dfrac{1}{2}. \end{cases} \tag{2}$$

So, if the expected value of $x_{i,t}$ is more than .5, its value changes by an amount $v_{i,t} > 0$, meaning that it moves toward its maximum. Likewise, if the expected value of $x_{i,t}$ is less than .5, its value is moved toward 0. Equation 2 makes sure the situation vectors adapt themselves to world knowledge while no value $x_{i,t}$ becomes negative or exceeds its maximum. Taken together, all values $v_{i,t}$ for a particular t form the velocity vector

$$V_{\text{wk},t} = (v_{1,t}, \ldots, v_{150,t}), \tag{3}$$

which gives the change in vector X_t resulting from the application of world knowledge (as indicated by "wk"). As a result of changing each X_t in the direction of $V_{\text{wk},t}$, the sequence of situations stays consistent with the original sequence and becomes more likely to occur in the microworld.

Stopping criterion and processing depth. Equation 2 is evaluated, using the original situation vectors X_t^0 as initial values, until the total change in vector values drops below a threshold value controlled by the depth-of-processing parameter θ:

$$\sum_{i,t} \left|v_{i,t}\right| < \frac{1}{\theta}, \tag{4}$$

where t ranges from 1 to the number of situations processed so far. Parameter θ indicates the extent to which the story's situations will be integrated. The larger the θ, the stricter the criterion for sufficient processing is, leading to stronger integration of story situations, which is assumed to correspond to deeper processing.

Interpreting the Model

Belief values. Although situation vectors are difficult to interpret by themselves, the meaning of any vector can be investigated by comparing it with specific vectors, such as those representing basic events. More precisely, the subjective probability that some p will occur in situation X can be computed from their vector representations. This probability is called the *belief value* of p in situation X because it indicates the extent to which a reader may believe that the event occurs in that situation. During the model's integration process, an increase in belief value of p reflects that it is inferred to have taken place. Also, it is possible to compute the subjective probability that p occurs at time step t given the previous and following situations, $X_{t\pm1}$. Frank et al. (2003) showed empirically that belief values, as defined below, indeed form good estimates of the actual probabilities in the microworld.

The a priori belief value of a (basic or complex) event p can be computed from its vector representation, $\mu(p)$. As a result of training the SOM, the a priori subjective

probability that p occurs in the microworld, denoted $\tau(p)$, equals the average value of its vector representation:

$$\tau(p) = \frac{1}{150}\sum_i \mu_i(p). \tag{5}$$

Likewise, the subjective probability that event p occurs in some situation X is the belief value $\tau(p|X)$. For example, given the vectors μ(B outside $\wedge$ J outside) and μ(sun), it is possible to compute τ(sun|B outside $\wedge$ J outside), indicating how likely a reader may find it that the sun shines at the moment in the story about which the story text states that Bob and Joe are outside. From the conjunction rule of Equation 1, the definition of $\tau(p)$ in Equation 5, and the definition of conditional probability, it follows that

$$\tau(p|X) = \frac{\tau(p \wedge X)}{\tau(X)} = \frac{\sum_i \mu_i(p)x_i}{\sum_i x_i}. \tag{6}$$

The subjective probability that event p occurs at time step t, given the previous and/or next situations, is denoted $\tau(p_t|X_{t\pm1})$. For instance τ(J wins$_t$|hide-and-seek$_{t-1}$) is the amount of belief that the current story event includes Joe's winning the game of hide-and-seek that was being played previously. Likewise, τ(B dog$_t$|B tired$_{t+1}$) is the subjective probability that Bob plays with the dog, given that he is tired a moment later. Since vector E_t represents the situation that can be expected from the situations at $t \pm 1$, the belief value of p_t given $X_{t\pm1}$ is computed by substituting E_t for the actual situation X in Equation 6:

$$\tau\left(p_t|X_{t\pm1}\right) = \tau\left(p_t|E_t\right) = \frac{\sum_i \mu_i(p)E_{i,t}}{\sum_i E_{i,t}}. \tag{7}$$

Coherence. Frank et al. (2003) defined a formal measure of story coherence by comparing a priori belief values (Equation 5) with a posteriori belief values (Equation 7). If the story situations at time steps t and $t \pm 1$ are in line with the temporal world knowledge in W (i.e., they follow expectations), the belief value of situation X_t in context $X_{t\pm1}$ will be larger than it is without context—that is, $\tau(X_t|X_{t\pm1}) > \tau(X_t)$—leading to a positive contribution to the coherence of the story containing the situations. Likewise, if $\tau(X_t|X_{t\pm1}) < \tau(X_t)$, the situations contribute negatively to the story's coherence because they violate expectations. As a result of the model's integration process, the coherence value of the story's interpretation generally increases during processing (Frank et al., 2003).

ADDING DISAMBIGUATION: THE DSS–AR MODEL

The DSS model shows how world knowledge is used to increase coherence between story statements, leading to the inference of additional information about the story's events. Since it simulates the cognitive process of coherence establishment, the model lends itself to being extended in order to form an implementation of Hobbs's (1979) and Kehler's (2002) proposal that establishing co-

herence gives rise to resolving referential ambiguity. As it is, however, the model cannot simulate disambiguation because it lacks a process that (attempts to) select among the ambiguity's readings. For instance, the statement *he wins* in a text about Bob and Joe is, for DSS, equivalent to the disjunction *Bob wins or Joe wins*, without a drive to infer one of the options. Ambiguity resolution, however, is the selection of one reading over the other(s), which can be thought of as a form of inference aimed at choosing from a small set of possibilities. In addition, bias often plays a role in ambiguity resolution. Therefore, DSS is turned into DSS–AR by adding formalizations of bias and of a disambiguation process.

Any biasing effect (e.g., the first-mention effect) is assumed to have taken place before the model starts resolving the ambiguity: Bias affects only the ambiguity's initial interpretation. When the vector representation of *he wins* in a text about Bob and Joe is constructed, bias for *Bob wins* is taken into account by letting that reading have a greater impact on the representation than does *Joe wins*.

As was mentioned above, the psychological process of disambiguation is modeled as a process of attraction in situation space. Each of the ambiguity's readings introduces into situation space a *center of gravity*, giving rise to an *attractor region* that drives the process of attraction. For example, the ambiguous statement *he wins* in a text about Bob and Joe results in two attractor regions, one being the "Bob wins" region, the other the "Joe wins" region. The "Bob wins" region consists of all the points in situation space that represent situations in which Bob wins, whereas the "Joe wins" region contains all situations in which Joe wins.

The process of attraction by the readings is now quite simple. It begins by taking the vector representing the ambiguous statement and modified by bias and lets each attractor region "pull" it into its direction, as will be explained in detail below. The influence an attractor region has on the vector increases as the vector gets closer to the region, making the vector "fall" toward one of the regions with increasing velocity and resulting in an increased belief value for the corresponding situation. As soon as the vector arrives in one of the attractor regions, a reading is chosen, and the ambiguity is resolved. From then on, the attractor regions do not affect the vector any longer.

To make disambiguation sensitive to context and world knowledge, no additional process is needed. Instead, the DSS model's integration process runs simultaneously with the disambiguation process. As a result, the ambiguity's vector moves toward the most likely reading and will often end up there; that is, the correct reading is usually selected as a side effect of integration.

As an example, take the sentence *Bob is tired and Joe is not, so he wins*, which describes two situations. Assuming that first mention makes Bob more accessible, resulting in bias for Bob to win, the second situation, in which someone wins, is initially represented by a vector closer to "Bob wins" than to "Joe wins" in situation space. As a consequence, it is forced more strongly toward the "Bob wins" region than toward the "Joe wins" region. However,

the context statement *Bob is tired and Joe is not*, in combination with knowledge about the relation between being tired and winning, makes the reading in which Joe wins more coherent. Therefore, the DSS model's integration process moves the vector toward the "Joe wins" region. Which of the two readings is eventually selected depends on the context: If it lends enough support to the inference that Joe is the winner, bias may be overruled so that "Joe wins" is chosen as the reading of the second clause. It is also possible that the ambiguity is resolved only partially, meaning that the vector does not end up inside any attractor region. This is especially likely when processing is shallow and the effects of bias and context are opposite.

Specification of the Model

Bias. Let $p, q, r, \ldots$ denote the situations to which an ambiguous statement could refer (i.e., these are its readings). Most often, there will be only two: p and q. In the DSS model, the ambiguous statement would simply be represented by the vector $\mu(p \vee q \vee r \vee \ldots)$—that is, as the disjunction of the readings without bias for any reading. In the DSS–AR model, on the other hand, bias affects the initial representation of the ambiguous statement.

To each reading $p, q, \ldots$, a *bias parameter* $\phi_p, \phi_q, \ldots$ is associated. They are all nonnegative, and their sum cannot exceed 1. An ambiguous situation is represented by a vector that is like $\mu(p \vee q \vee \ldots)$ but moved slightly into the direction of each reading for which there is a bias—that is, whose bias parameter is larger than 0. Formally, the initial vector representation of an ambiguous statement with two readings p and q equals

$$X^0 = \mu(p \vee q) + \phi_p \left[\mu(p) - \mu(p \vee q) \right]$$
$$+ \phi_q \left[\mu(q) - \mu(p \vee q) \right]. \quad (8)$$

As is shown in Figure 1, $\mu(p) - \mu(p \vee q)$ is a vector indicating a shift from $\mu(p \vee q)$ to $\mu(p)$. The size of this shift is controlled by bias parameter ϕ_p. Similarly, the term $\mu(q) - \mu(p \vee q)$ indicates a shift from $\mu(p \vee q)$ to $\mu(q)$, the size of which is controlled by ϕ_q. The initial vector X^0 is constructed by adding the shifts toward $\mu(p)$ and $\mu(q)$ to the vector representing $p \vee q$. Equation 8 can easily be generalized to cope with cases in which there are more than two readings.

Note that we do not attempt to explain how, when, or why bias arises. The extent to which certain interpretations are a priori preferred is simply taken as given and determines part of the model's parameter setting. Another simplifying assumption is that different sources of bias can be treated equally by simply summing their individual bias parameters. If, for instance, there are two sources of bias for p, which individually would lead to bias parameters $\phi_p = .3$ and $\phi_p = .2$, the resulting bias parameter is $\phi_p = .3 + .2 = .5$.

The model will also be used to compare the processing of ambiguous statements with that of unambiguous statements. No adjustment to the model is needed to deal with unambiguous cases. If it is simply stated that p is the case (i.e., the statement is unambiguous), the initial situation vector X^0 is, of course, not based on the disjunction $p \vee q$,
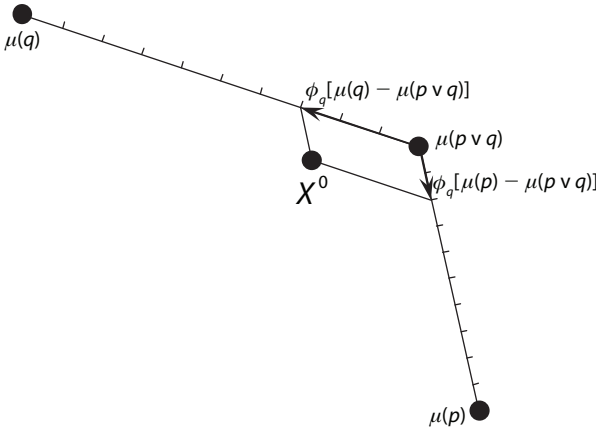
Figure 1. Construction of the initial vector representation X^0 of an ambiguous statement with readings p and q. Bias parameters are set to $\phi_p = .2$ and $\phi_q = .3$. The three points $\mu(p)$, $\mu(q)$, and $\mu(p \vee q)$ define a plane in situation space. Generally, the origin of situation space does not lie in this plane, so neither the origin nor the axes are shown here.

but on p. This means that in Equation 8, $p \vee q$ is replaced by p, so Equation 8 reduces to

$$X^0 = \mu(p) + \phi_q[\mu(q) - \mu(p)].$$

This shows that, if p is stated, bias for q even has an effect without any ambiguity. If there is no bias for q, $\phi_q = 0$, so $X^0 = \mu(p)$, as in the DSS model.

The attraction process. It follows from the conjunction rule of Equation 1 that any situation in which event p occurs must correspond to a situation space point $X = (x_1, \ldots, x_{150})$, with $x_i \leq \mu_i(p)$ for all i. The set of all such points forms the region of situation space representing all possible situations in which p is the case. Therefore, this is how the attractor region for p is defined.

The attraction process is implemented as follows. Let, for each attractor region p, $D_p = (d_{p,1}, \ldots, d_{p,150})$ be the vector pointing from a given point X to the nearest point in the p region. This means that $d_{p,i} = \min\{0, \mu_i(p) - x_i\}$. The distance from X to the p region equals the Euclidean length $\|D_p\|$ of this vector. The point X lies within the p region [i.e., $x_i \leq \mu_i(p)$ for all i] if and only if this distance equals 0. If this is the case for any attractor region, the ambiguity is resolved, and no region affects X. Otherwise, the region's gravitational force gives X a velocity, denoted V_p, in the direction of the p region. As the distance to the p region decreases, this velocity increases until it reaches an arbitrary maximum of 1 when X is infinitesimally close to the p region. The simplest expression for V_p that has such properties is

$$V_p = \frac{D_p}{\|D_p\|\big(1 + \|D_p\|\big)}.$$

That is, vector D_p is divided by its own length to result in the unit vector pointing from X toward the p region. Next, it is divided by $1 + \|D_p\|$ to make velocity depend on distance, with a maximum of 1. The velocity of X caused

by all attractor regions together is simply the sum of the velocities caused by the individual regions; that is,

$$V_{ap} = V_p + V_q + V_r + \cdots \qquad (9)$$

is the total velocity resulting from the attraction process (as indicated by "ap").

The DSS–AR model. The attraction process described above does not involve context information and world knowledge. It is affected only by the attractor regions corresponding to the readings and by the initial representation of the ambiguity, X^0, which depends only on initial preferences for each of the ambiguity's readings (i.e., on biases). To generally select the correct reading, however, context information and world knowledge need to be used.

The DSS model applies world knowledge to a series of story statements, resulting in a story interpretation that is more coherent than the original story. By adding the DSS model to the gravitational attraction process described above, sensible disambiguation becomes a side effect of the build-up of coherence. This addition is possible because both the application of world knowledge and the attraction process are expressed as the movement of vectors through situation space. By adding the velocities from the two sources, disambiguation comes to depend on context and world knowledge and often results in the most coherent reading. The change in X_t caused by the combination of the DSS model and the attraction process is defined by the differential equation

$$V_t = V_{wk,t} + \beta V_{ap,t}; \qquad (10)$$

that is, the change in vector X_t over processing time equals its velocity according to DSS ($V_{wk,t}$ from Equations 3 and 2) plus β times its velocity according to DSS–AR's attraction process (V_{ap} from Equation 9). Parameter β controls the strength of the influence of the ambiguity. The process halts when, according to Equation 4, the criterion for sufficient processing is reached.

Note that as soon as one of the attractor regions is reached, $V_{ap,t} = 0$ but X_t will nevertheless continue to be updated as long as $V_{wk,t}$ is large enough. Also, when there is no ambiguity, there is no need to choose a reading, so there are no attractor regions. This means that $V_{ap,t} = 0$ and $V_t = V_{wk,t}$, making the DSS–AR model's dynamics the same as those of DSS. Also, be reminded that in absence of ambiguity and bias, X^0 is the same as in the DSS model. That is, if there is neither ambiguity nor bias, the DSS–AR model reduces to DSS, showing that the DSS–AR model is a true extension of the DSS model and that the results of DSS remain valid for DSS–AR.

Interpreting the Model

Relation to experimental data. To determine the extent to which one of the readings p, q, ... is chosen over the others, belief values $\tau(p|X_t)$, $\tau(q|X_t)$, ..., as defined in Equation 6, can be computed at any moment during processing. The model's interpretation of the ambiguity is the reading with the largest belief value at the moment processing halts. If a context-inconsistent reading ends up with the largest belief value, this is regarded as an error in

disambiguation. Error rates according to the model will be compared with those from experiments. A more important measure for model validation is the time that was needed for the process to stabilize. As in the DSS model, processing times serve as the model's correlate of experimentally obtained reading times.

Degree of resolution. To investigate under which conditions ambiguities remain (largely) unresolved, we define a measure indicating the extent to which the ambiguity was resolved. This *degree of resolution* is the absolute difference between the final belief values of the readings:[5]

$$\text{degree of resolution} = |\tau(p|X_t) - \tau(q|X_t)|. \quad (11)$$

Context strength. For interpreting the model's outcome, it will be useful to have a measure of the helpfulness of context information for resolving the ambiguity. The influence that context is expected to have on disambiguation can be expressed in terms of belief values. Without any context, the belief value of p is the a priori value $\tau(p)$. Assuming that t is the story time step of the ambiguous statement, the influence of the previous or following context situation, $X_{t\pm1}$, changes the belief value by an amount $\tau(p_t|X_{t\pm1}) - \tau(p)$. Of course, the same is true for the belief value of q, another reading of the ambiguity. The context's preference for p over q equals the difference between the two context influences: $[\tau(p_t|X_{t\pm1}) - \tau(p)] - [\tau(q_t|X_{t\pm1}) - \tau(q)]$.

If this value is positive, context prefers reading p. If it is negative, the context points toward q. The context strength is simply the absolute value of the context's preference. A context strength of 0 means the context does not help at all to decide on the intended reading of the ambiguity. If, on the other hand, context strength is large, information from the context makes one of the readings much more likely than the other(s).[6]

SETUP OF SIMULATIONS

Model Input

The DSS–AR model, being applied to the microworld described earlier in this article, processed items corresponding to the ambiguous statement *he wins* at time step t in varying contexts. One example of a context is the story situation "B tired $\wedge\neg$(J tired)" at $t - 1$, which results in an item corresponding to the sentence *Bob is tired and Joe is not, so he wins*. Only one situation was used as context, so all items consisted of two situations, one being the ambiguity and the other its context. Since at least two situations are required for the integration process to start, the simulations presented here did not involve any incremental processing even though the DSS model can process incrementally at the level of clauses.

As contexts, all situations were used that satisfied the following three constraints: (1) The situation takes place at time step $t - 1$ or $t + 1$; that is, it occurs directly before or directly after the "winning" event; (2) the resulting context strength is at least .02; (3) the situation is either a (negation of a) basic event or a conjunction of (negation of) two basic events.[7]

This made a total of 368 different contexts. The average context strength was .041, with a maximum of .088. Each processed item consisted of the context situation and the situation corresponding to the ambiguous statement *he wins*. Since all items were assumed to be embedded in a text about Bob and Joe, the readings of the ambiguity are *Bob wins* and *Joe wins*. Unambiguous items were constructed from the ambiguous ones by replacing the ambiguity by the context-consistent reading.

Parameter Setting

There are four adjustable parameters in the model: depth-of-processing θ (of Equation 4), the strength of the ambiguity's influence β (Equation 10), and the biases ϕ_p and ϕ_q (Equation 8) for the two readings p and q of the ambiguity. No sophisticated search through parameter space was applied to find a suitable parameter setting.

The default value of θ was set to .3, which is also its default in the DSS model. Preliminary simulations revealed that varying the level of β between .1 and .9 had no important effect on model outcomes, so it was fixed at $\beta = .5$. No attempt was made to improve the fit between model outcomes and empirical data by adjusting β.

To find a proper value for the bias parameters, different levels for the total amount of bias $\phi = \phi_p + \phi_q$ were applied first. These levels of $\phi = .2, .3, .4, .5, .6,$ and .8 did not result in qualitative differences in model processing times. After visually comparing the resulting model processing times to reaction time data from Stewart et al. (2000, Experiment 1, Table 1, NP1 verb type condition), a value of $\phi = .5$ was selected. Next, total bias was divided into the two biases ϕ_p and ϕ_q, as explained below.

Simulated Conditions

As was mentioned in the description of the DSS–AR model, the bias parameters can be composed of contributions from several sources of bias. Here, two such sources are considered: the first-mention effect and implicit causality. These two sources can bias either toward the same reading or toward different readings. In the first case, implicit causality biases toward the reading that has the first-mentioned entity as the anaphor's referent, so the processed item can be viewed as *containing an NP1 verb*. If, on the other hand, implicit causality biases toward the reading that has the second-mentioned entity as the referent, the item can be interpreted as *containing an NP2 verb*.

Since there exist verbs that result in an NP2 bias, the effect of first mention must be somewhat weaker than implicit causality. In the model's simulations, the first-mention effect contributes a lesser amount to the bias parameter than does implicit causality: Without any fitting to experimental data, these amounts were chosen to be .2 and .3, respectively. If choosing the first-mentioned entity as referent would result in reading p, this means that $\phi_p = .2 + .3 = .5$ and $\phi_q = 0$ in the case of an NP1 verb, whereas an NP2 verb results in $\phi_p = .2$ and $\phi_q = .3$.

For obtaining model predictions of reading times and error rates, each of the 368 items was processed four times: There was either an NP1 verb or an NP2 verb, and context information was either congruent or incongruent with implicit causality. These different conditions are simulated by varying bias. Let p denote the correct (i.e.,

context-consistent) reading in the current item and q be the other, incorrect reading. As was mentioned above, an NP1 verb leads to bias for one reading only. In case of a congruency, therefore, only the correct reading p is biased, so $\phi_p = .5$ and $\phi_q = 0$. In case of incongruency, $\phi_p = 0$ and $\phi_q = .5$. To simulate the occurrence of an NP2 verb, bias is split between the readings; that is, $\phi_p = .3$ and $\phi_q = .2$ in congruent cases, whereas $\phi_p = .2$ and $\phi_q = .3$ in incongruent cases.

RESULTS

Time Course of Ambiguity Resolution

The left panel of Figure 2 shows how the belief values of "B wins" and of "J wins" develop over processing time, expressed in arbitrary units, as the item corresponding to the sentence *Bob is tired and Joe is not, so he wins* is disambiguated, with only "B wins" receiving bias, reflecting that Bob is mentioned first. Note that time index 0 corresponds to the moment that integration of the two clauses begins, so model processing time is not the full time over which the sentence is read. Clearly, the biased reading is preferred initially, but the application of context information eventually results in the most coherent reading. As the belief value of "Joe wins" increases, the belief value of "Bob wins" decreases with the same amount, because the two events exclude each other: Any evidence for Joe's winning must also be evidence against Bob's winning.

The right panel of Figure 2 shows the development of belief values when the context is "J tired $\wedge \neg$(B tired)" at $t - 1$ (*Joe is tired and Bob is not*) with only "J wins" receiving bias.[8] Again, the biased reading is preferred at first, but this is overruled by the context's preferring the other reading.

These results are in line with the experimental findings of Arnold et al. (2000), Gordon and Scearce (1995), and Vonk (1985) discussed in the section on empirical findings. They are not surprising, however, since bias affects the initial vector representation of the statement and context begins to have an effect when this vector is adjusted

by information present in the other statement. It is, however, important to show that the parameter setting allows for a noticeable effect of bias, and that the effect of context is large enough to overcome this bias. Note that this same parameter setting was used in all other simulations as well, unless the effect of varying parameter values was the issue under investigation.

Reading Times

We will first show how congruency and ambiguity affect model processing times. Next, it will be shown how the size of the congruency effect depends on processing depth and verb type. Finally, we will present the model's predictions regarding the effect of context strength on processing time for ambiguous items, with and without bias.

The effects of congruency and ambiguity. As was discussed in the section on empirical findings, congruent sentences are read more quickly than incongruent ones. Stewart et al. (2000, Experiments 1, 2, and 3) found that this congruency effect on reading times occurred not only in referentially ambiguous, pronoun-containing sentences, but also in sentences that were unambiguous because the pronoun was replaced by the name of the intended referent. There was no significant interaction between the congruency and ambiguity variables (i.e., whether or not the sentence contained a pronoun).

In two of these experiments, sentence stimuli were presented in two fragments, the first containing the words up to and including the anaphor, and the second containing the rest of the sentence. Integration of the two clauses could not begin until at least the verb of the second clause was read. Therefore, the time needed for integration (including disambiguation) can show up in the second-fragment reading times only. Since the model simulates the disambiguation process and not the word-by-word reading process, model processing times should be compared with these second-fragment reading times.[9]

As Figure 3 shows, the model predicts processing times structurally similar to the reading times found by Stewart et al. (2000). Incongruent items were processed
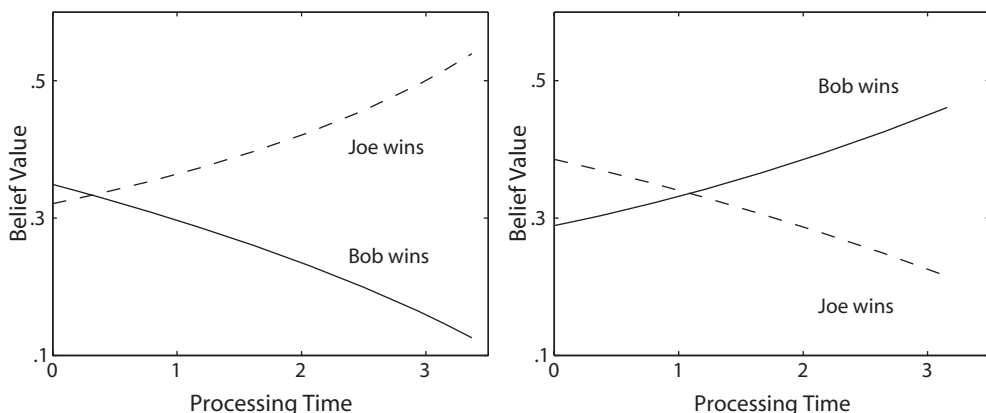

Figure 2. Belief values of *Bob wins* and of *Joe wins* during processing of items corresponding to *Bob is tired and Joe is not, so he wins* (left) and *Joe is tired and Bob is not, so he wins* (right), with bias pointing toward the reading in which the entity stated first in the sentence is the one that wins.
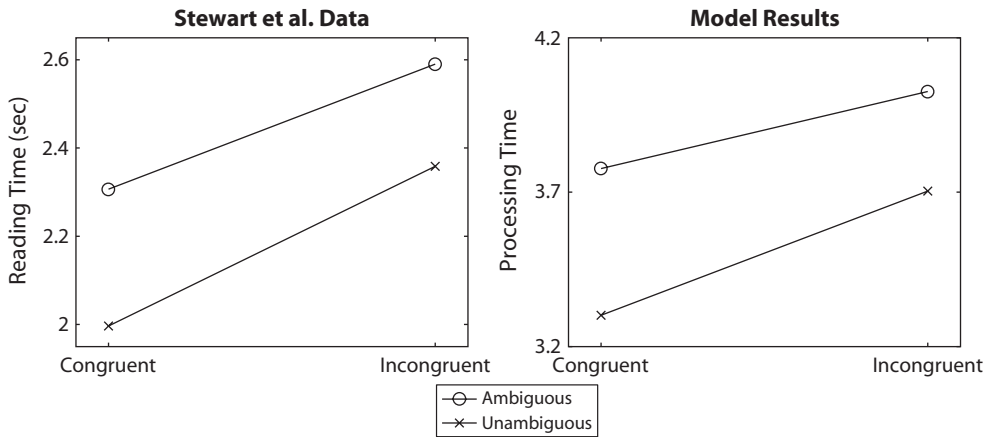
Figure 3. Left: Effect of congruency and ambiguity on reading times. Data are from Stewart, Pickering, and Sanford (2000, Experiment 3, Table 4, "deep question type" condition), Sentence Fragment 2, averaged over implicit cause conditions NP1 and NP2. Right: Effect of congruency and ambiguity on model processing times, averaged over the two verb type conditions.

more slowly than congruent ones $[F(1,2940) = 20.1, p < 10^{-5}]$, and ambiguous items were processed more slowly than corresponding unambiguous items $[F(1,2940) = 31.1, p < 10^{-7}]$. Also, there was no significant interaction between the congruency and ambiguity variables $[F(1,2940) = 1.1, p > .29]$.

The effect of processing depth. Another finding by Stewart et al. (2000) was that changing the subjects' reading task to one that did not require deep processing resulted in a smaller congruency effect. The model shows the same result (Figure 4). The congruency effect (defined as processing times on incongruent items minus processing times on congruent items) is smaller for values of the depth-of-processing parameter lower than $\theta = .3$, which was used in the previous simulations. This is true for both the ambiguous condition $[t(2942) = 9.27, p < 10^{-18}]$ and the unambiguous condition $[t(2942) = 14.9, p < 10^{-47}]$.

The model predicts an interaction between processing depth and anaphor type: The effect of θ is larger in the unambiguous condition than in the ambiguous condition $[t(5884) = 5.93, p < 10^{-8}]$. Stewart et al. did not report whether the same interaction was significant in their data. In the Discussion section, we will present an explanation for this interaction, accounting for both the experimental and the simulation data.

The effect of verb type. All the results for the figures above were averaged over the two implicit causality conditions (NP1 verbs and NP2 verbs). However, in all four of their experiments, Stewart et al. (2000) found that the congruency effect in the ambiguous condition depended strongly on verb type: The effect was much smaller (sometimes even absent) for NP2 verbs, although not always significantly so. For instance, in their third experiment, the congruency effect was 556 msec in the NP1
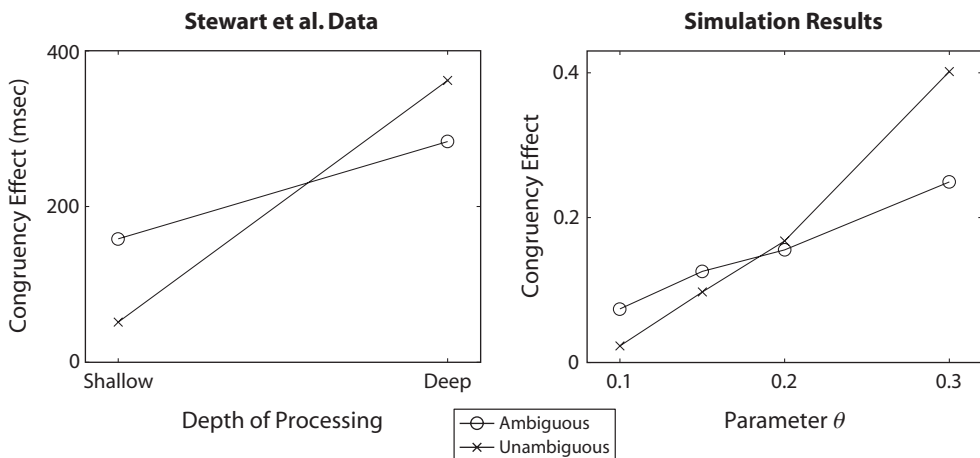

Figure 4. Left: Effect of processing depth on congruency effect. Data are from Stewart, Pickering, and Sanford (2000, Experiment 3, Table 4), Sentence Fragment 2, averaged over implicit cause conditions NP1 and NP2. Right: Effect of depth-of-processing parameter θ on congruency effect in model simulations, averaged over the two verb type conditions.

verb condition but only 11 msec in the NP2 verb condition (Table 4, Fragment 2 reading times in the "deep question type" condition). The same effect was found by Garnham et al. (1992, Experiments 4 and 5).[10] The model predicts such a differential effect: The predicted congruency effect on processing times was .398 for NP1 verbs (i.e., the two sources of bias agree) and only .098 for NP2 verbs (i.e., the two sources of bias disagree). The difference is statistically significant [$t(734) = 9.22, p < 10^{-18}$].

The effect of context strength. As for the effect of context strength on processing times, the model makes an interesting prediction. As might be expected, processing is faster when the context is strongly informative than when it is moderately informative (Figure 5, left panel). However, the absence of context information results in very fast processing.[11] Interestingly, Figure 5 also shows that removing the bias preference by setting both bias parameters to .25 does not affect processing times. This suggests that fast processing of items with very low context strengths can also result from not choosing a reading at all.

The right panel of Figure 5 shows the degree of ambiguity resolution (Equation 11) as a function of context strength, both with and without bias. In general, stronger contexts lead to higher degrees of resolution. The presence of bias does not affect the degree of resolution with high or intermediate levels of context strength. When context strength is very low, however, disambiguation takes place only if one of the entities receives more bias than the other. Otherwise, no reading is selected. However, as can be seen from the leftmost points of the left-panel graph in Figure 5, the model takes as little time deciding that trying to resolve the ambiguity is futile because there is no bias as it takes selecting a reading on the basis of bias only. An explanation for this will be presented in the Discussion section.

Error Rates

Error rate data were obtained from the same simulations as those that produced the processing time data presented above. Below, we will show how predicted error percentages depend on congruency and context strength and how verb type affects both error rates and the size of the congruency effect on error rates.

The effects of congruency and context strength. As was discussed in the section on empirical findings, Leonard et al. (1997a) found that the percentage of disambiguation errors depends on congruency and context informativeness and that those two factors interact. Figure 6 shows that although the model's error rate is too large when context is weak, it does predict these three effects. First, incongruent items result in more errors than do congruent ones [$\chi^2(1) = 100.5, p < 10^{-16}; n = 1,472$]. Second, after dividing the contexts into a *weak* and a *strong* class along the median of context strength (.0378), weak contexts result in more errors than do strong contexts. This is the case for both congruent [$\chi^2(1) = 5.05, p < .025; n = 736$] and incongruent [$\chi^2(1) = 30.8, p < 10^{-7}; n = 736$] items. Third, the effect of congruency is larger in weak contexts than in strong contexts [$\chi^2(1) = 19.5, p < .0001; n = 736$].

The effect of verb type. Unlike Leonard et al. (1997a), Stewart et al. (2000) presented data for NP1 verbs and NP2 verbs separately. In their ambiguous condition, they found a main effect of verb type on error rates. In general, more errors were made in the NP1-verb condition than in the NP2-verb condition. Averaged over the results for Experiments 1, 2, and 4,[12] NP1 verbs resulted in an error rate of 23.2%, whereas NP2 verbs resulted in 11.4% errors. The model makes fewer errors but predicts the same effect of verb bias: 15.3% errors in the NP1 verb condition and 3.1% errors in the NP2 verb condition [$\chi^2(1) = 64.9, p < 10^{-15}; n = 1,472$]. In the model, this effect is not present with congruent items [$\chi^2(1) = .33, p > .5; n = 736$]. Likewise, the Stewart et al. data show a smaller effect of verb type in the congruent condition than in the incongruent condition.

Stewart et al. (2000) also found that the congruency effect on error rates was larger for NP1 verbs than for NP2 verbs, as it was for reading times. Averaged over the
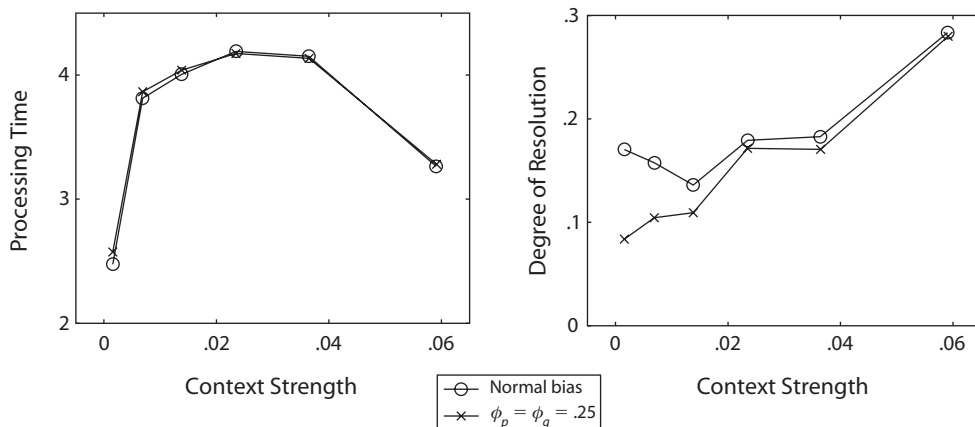

Figure 5. Effect of context strength on model processing time (left) and degree of ambiguity resolution (right), with normal bias and with both bias parameters fixed at .25. Every point in the graph is the average of 521 processed items when bias is normal and of 130 or 131 items when both bias parameters are set to .25.
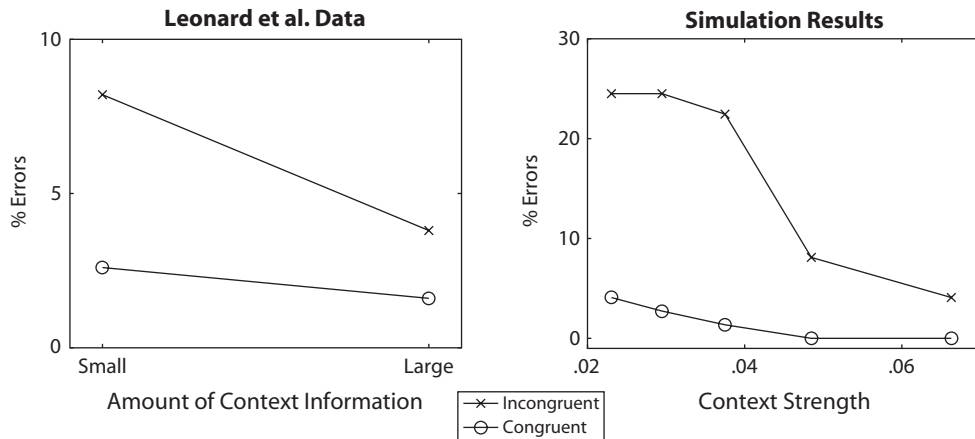
Figure 6. Left: Effect of congruency and context informativeness on error percentages. Data are from Leonard, Waters, and Caplan (1997a, Figure 6, "young adults" group). Right: Effect of congruency and context strengths on model error percentages, averaged over the two verb type conditions. Every point in the graph is the percentage of errors over 147 or 148 processed items.

results for Experiments 1, 2, and 4, the congruency effect was 13.1% in the NP1 verb condition but only 3.9% in the NP2 verb condition. The model predicts the same effect: The congruency effect on error rates is 27.9% in the NP1 verb condition and 2.5% in the NP2 verb condition [$\chi^2(1) = 92.1$, $p < 10^{-16}$; $n = 736$].

DISCUSSION

The DSS–AR model successfully simulates the influence of context information and bias on the resolution of referential ambiguity. In accordance with the assumption that the resolution of referential ambiguities results from a general, coherence-increasing, knowledge-based inference process, the model provides a unified account of inference and disambiguation by making two simple additions to the DSS model of knowledge-based inference. First, bias was assumed to have an effect only on the initial representation of a referentially ambiguous statement. Second, the selection of a reading of the ambiguity was driven by attractor regions resulting from introducing centers of gravity into situation space.

Explanations of Empirical Data

Reading times. The model accounts for a large amount of reading time data. In line with Stewart et al.'s (2000) experimental findings, referentially ambiguous sentences are processed more slowly than the corresponding unambiguous sentences, and both these processing times are affected by congruency between bias and context. It is clear why there should be a congruency effect in both the unambiguous and the ambiguous conditions. Whether an unambiguous name or an ambiguous pronoun is used, bias leads to an expectation that takes some time to neutralize when it turns out to be inconsistent with new information. In Stewart et al.'s experiments, items containing an ambiguous pronoun were turned into unambiguous items by replacing the pronoun with the name that resulted in the most coherent sentence. This means that any bias for the

other entity (i.e., any incongruency) decreased coherence. Similarly, in the model's input, ambiguous items are turned into unambiguous items by stating that either "Bob wins" or "Joe wins," whatever is consistent with the context. Any bias for the alternative reading decreases coherence. Since the DSS model is slower to process less coherent items (Frank et al., 2003), DSS–AR shows a congruency effect on unambiguous items.

It is less obvious why shallower processing decreases the congruency effect in the model. Considering that the depth-of-processing parameter θ affects only the point at which the integration process halts and that this process is not informed about congruency, θ might not be expected to affect the size of the congruency effect. Part of the effect of θ is caused by a general decrease in processing time for shallower processing, but especially in the unambiguous condition, this is not enough to explain all of the effect.

When processing is shallow, an item is easily considered sufficiently processed. In other words, there is not much of a need to strongly integrate the two clauses of the sentence, making context information less relevant. Since congruency concerns the relation between bias and context information, this means that congruency itself is less relevant when processing is shallow than when it is deep. In terms of the model's dynamics, incongruency usually leads to slow processing because of the initial bias that needs to be overcome. If θ is small enough, however, the process can halt before it has stabilized, leaving bias untouched and doing away with much of the effect of congruency. A similar explanation may account for the experimental findings.

Stewart et al.'s (2000) data also seemed to show a three-way interaction among the ambiguity, congruency, and processing depth factors (no test of this effect was reported). As can be seen from Figure 4, the congruency effect is more sensitive to processing depth in the unambiguous condition than in the ambiguous condition. This interaction was also visible in the processing times predicted by the model.

How to account for this effect? Presumably, reading an ambiguous pronoun triggers an attempt to apply context information to resolve the pronoun, regardless of processing depth. This means that the extent to which an ambiguous clause and its context become integrated will not depend strongly on processing depth. Integration of two unambiguous clauses, on the other hand, may occur only when processing is sufficiently deep, because there is nothing to resolve. When processing is shallow, unambiguous statements are simply taken for granted. Only when processing is deep does the reader integrate the clauses to check whether the statement is consistent with context. Since integration of the clauses is required for a congruency effect to occur, the congruency effect is more sensitive to processing depth in the unambiguous condition than in the ambiguous condition. The same holds in the model: Given an ambiguity, the attractor regions will immediately affect the vector representing the ambiguity, keeping its velocity above the threshold $1/\theta$ even if this threshold is high (i.e., processing is shallow). Without an ambiguity, there are no attractor regions, and only the DSS model's integration process takes place. As $1/\theta$ is increased, there will be more contexts in which this process lacks the impact to keep the vector's velocity above the threshold.

The DSS–AR model also explains the interaction between congruency and verb type found by Stewart et al. (2000) and Garnham et al. (1992). Stewart and Gosselin (2000) suggested that this interaction may be caused by the first-mention effect. An NP1 verb biases toward the subject, which is already biased because it is mentioned first. This makes the preference for the subject even stronger and results in a larger effect of congruency. An NP2 verb, on the other hand, biases toward the object, counteracting the first-mention effect. As a result, it does not matter much whether or not context information and implicit cause are congruent. Indeed, the model does assume that both first-mention and implicit causality affect bias and, thereby, explains why the congruency effect is larger for NP1 verbs than for NP2 verbs.

Error rates. The model predicted an effect of congruency on the percentage of errors in disambiguation, as well as a decrease in error rates as context strength increases (see Figure 6). Both these effects are in agreement with the data in Leonard et al. (1997a). When bias points toward the reading that is incorrect according to context information, there will be more cases in which this incorrect reading is chosen than when bias and context are congruent. However, strong contexts can overrule an incongruent bias more often than weaker contexts can, resulting in fewer errors.

Also, implementing both first-mention and implicit causality as a form of bias accounts for the main effect of verb type on error rates. Stewart et al. (2000) found that more errors are made when sentences with an NP1 verb are read than when sentences with an NP2 verb are read. In the model, the difference between the readings' bias parameters is much larger in the NP1 verb condition ($|\phi_p - \phi_q| = .5$) than in the NP2 verb condition ($|\phi_p - \phi_q| = .1$). This means that there will be more items in which context is too weak to overcome an incongruent,

bias-based interpretation in the NP1 verb condition than in the NP2 verb condition, resulting in the difference in error rates. This explanation is supported by the finding that the model's verb type effect is restricted to the incongruent items.

The DSS–AR model correctly predicts the interaction between congruency and verb type. Since the first-mention bias is combined with an NP1 verb bias but canceled out by an NP2 verb bias, the congruency effect on error rates is larger for NP1 verbs than for NP2 verbs, as was the case for processing times.

Timing of the Implicit Causality Effect

The model's reading time predictions are of special importance because of the ongoing debate about the timing of implicit causality. Be reminded from the section on empirical findings that Stewart et al. (2000) claimed that the integration account of implicit causality, unlike the focusing account, predicts that the congruency effect will occur with unambiguous name anaphors and will decrease with shallower processing. Indeed, they did find exactly these two effects, so they concluded that the focusing account is incorrect.

In the model, implicit causality is incorporated by making one of the readings of the ambiguity more preferred *before* any integration of clauses takes place. *During* the model's process, however, bias does not play a role any longer: The model is sensitive to biases only through their effect on the initial interpretation of the ambiguity. That is, the model implements the focusing account of implicit causality. Nevertheless, it predicts a congruency effect in the unambiguous condition and a smaller effect as depth-of-processing parameter θ is decreased. This clearly shows that Stewart et al.'s (2000) conclusion was not warranted. Of course, the model does not claim that implicit causality must have an immediate effect. It shows only that such an account is consistent with empirical findings (see also Koornneef & Van Berkum, 2006).

Novel Predictions: Leaving Ambiguity Unresolved

The DSS–AR model makes an interesting and experimentally testable prediction regarding the effect of context strength on processing times: Reading times will be longer for intermediate levels of context informativeness and shorter for both strongly informative contexts and uninformative contexts, whether biasing information is available or not. If there is bias, intermediate context strengths cause interference between bias and context information for incongruent items, resulting in slow processing. When context is either very strong or nearly absent, on the other hand, the ambiguity can be resolved quickly because either context or bias dominates the process. If there is neither bias nor context information, the situation vector representing the ambiguous statement does not move toward either of the attractor regions. The vector's rate of change will therefore fall below the threshold of $1/\theta$ quickly, resulting in short processing times but unresolved ambiguities.

We are not aware of any experimental findings regarding the effect of context informativeness on reading times

for referentially ambiguous sentences, but Kahn and Till (1991, Table 3, "young adult" condition) do report some data that seem to confirm the model's predictions. They had subjects read short texts in which an ambiguous pronoun appeared. Context informativeness was not manipulated directly, but their three *context expectedness* conditions can be interpreted as varying in informativeness as well. By varying the sentences preceding the ambiguous sentence, the context either strongly implied (*expected context*), weakly implied (*neutral context*), or did not imply (*unexpected context*) one of the readings of the ambiguity. It was found that reading times for the ambiguous sentence were longer in the neutral context condition (4.81 sec) than in the expected context and unexpected context conditions (4.03 and 4.48 sec, respectively). However, it was not stated whether the difference between the neutral and the unexpected conditions was significant.

In a study on syntactic ambiguity, van Gompel, Pickering, and Traxler (2001) found short reading times for sentences that could not be disambiguated on the basis of bias or world knowledge. They claimed that, in these cases, a reading is quickly selected at random and explicitly rejected the hypothesis that such ambiguities remain unresolved, because such a model "would also need a mechanism that determines when a sentence is sufficiently ambiguous that processing be suspended" (p. 251). The DSS–AR model shows that, at least for referential ambiguity, such an additional mechanism is not required at all. Instead, the ability to leave "unresolvable" ambiguities unresolved is an emergent property of the disambiguation process.

Limitations and Challenges

In spite of these successes, the model also suffers from some limitations that should not be overlooked. Its focus on world knowledge and on nonlinguistic representations entails two main drawbacks. First, a model dealing with world knowledge is necessarily constricted to a microworld if it is to avoid including knowledge subjectively and only on an ad hoc basis. The problem with any microworld is, of course, its small size. It is still an open question to what extent the model can be scaled up to handle more realistic amounts of knowledge. Since world knowledge is immediately available to the model in the form of the situations' vector representations, it does not need to search through a knowledge base. Therefore, there seems to be no a priori reason to expect the model to break down when much more knowledge is included. However, implementing this knowledge in the vector representations may raise problems because of technical limitations. If training very large SOMs turns out to be infeasible, some other implementation scheme may be used. For instance, a so-called competitive network is easier to train than a SOM and yields vectors whose relevant properties are similar (Frank, 2006).

The model's second limitation is its lack of linguistic input. The "sentences" it processes are represented as events in the microworld, while the text from which an event originates is ignored. As a result, it was necessary to implement all sources of bias in a common way—namely, as a bias parameter ϕ. In spite of this simplification, how-

ever, a large amount of data was accounted for. This generic scheme makes it easy to include sources of bias in addition to the first-mention and implicit causality effects, possibly resulting in further predictions of experimental data. For instance, Järvikivi, van Gompel, Hyönä, and Bertram (2005) found that for the resolution of ambiguous pronouns in Finnish, there exists a biasing effect of grammatical role, in addition to (and not interacting with) the first-mention effect.[13] Grammatical role could directly be incorporated as another source of bias in the DSS–AR model.

Although the model's nonlinguistic input suffices for the phenomena it attempts to explain, the ability to process actual texts would greatly improve its explanatory value. For instance, such a model might be able to simulate how textual factors lead to bias. Recent results by Frank and Haselager (2006) show that recurrent neural networks can transform sentences describing microworld situations into the vector representations of these situations. Although these networks do not deal with additional information in the form of linguistic cues, they form a first step toward real text input for the DSS and DSS–AR models.

REFERENCES

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, **76**, B13-B26.

Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal Learning & Verbal Behavior*, **16**, 601-609.

Frank, S. L. (2006). *Methods for implementing world knowledge in the DSS model* (Tech. Rep.). Nijmegen: NICI.

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: Erlbaum.

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. (in press). World knowledge in computational models of discourse comprehension. *Discourse Processes*.

Garnham, A., Oakhill, J., & Cruttenden, H. (1992). The role of implicit causality and gender cue in the interpretation of pronouns. *Language & Cognitive Processes*, **7**, 231-255.

Garnham, A., Traxler, M., Oakhill, J., & Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory & Language*, **35**, 517-543.

Garrod, S. C., & Sanford, A. J. (1994). Resolving sentences in a discourse context: How discourse representation affects language understanding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 675-698). San Diego: Academic Press.

Garvey, C., Caramazza, A., & Yates, J. (1974). Factors influencing assignment of pronoun antecedents. *Cognition*, **3**, 227-243.

GERNSBACHER, M. A., & HARGREAVES, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory & Language*, **27**, 699-717.

GORDON, P. C., GROSZ, B. J., & GILLIOM, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, **17**, 311-347.

GORDON, P. C., & HENDRICK, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science*, **22**, 389-424.

GORDON, P. C., & SCEARCE, K. A. (1995). Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory & Cognition*, **23**, 313-323.

GREENE, S. B., MCKOON, G., & RATCLIFF, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 266-283.

HIRST, W., & BRILL, G. A. (1980). Contextual aspects of pronoun assignment. *Journal of Verbal Learning & Verbal Behavior*, **19**, 168-175.

HOBBS, J. R. (1979). Coherence and coreference. *Cognitive Science*, **3**, 67-90.

JÄRVIKIVI, J., VAN GOMPEL, R. P. G., HYÖNÄ, J., & BERTRAM, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science*, **16**, 260-264.

KAHN, H. J., & TILL, R. E. (1991). Pronoun reference and aging. *Developmental Neuropsychology*, **7**, 459-475.

KAMP, H., & REYLE, U. (1993). *From discourse to logic: Introduction to model-theoretic semantics of natural language, formal logic, and discourse representation theory*. Dordrecht: Kluwer.

KAWAMOTO, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory & Language*, **32**, 474-516.

KEHLER, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

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

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

KOORNNEEF, A. W., & VAN BERKUM, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory & Language*, **54**, 445-465.

LEONARD, C. L., WATERS, G. S., & CAPLAN, D. (1997a). The influence of contextual information on the resolution of ambiguous pronouns by younger and older adults. *Applied Psycholinguistics*, **18**, 293-317.

LEONARD, C. L., WATERS, G. S., & CAPLAN, D. (1997b). The use of contextual information by right brain-damaged individuals in the resolution of ambiguous pronouns. *Brain & Language*, **57**, 309-342.

LONG, D. L., & DE LEY, L. (2000). Implicit causality and discourse focus: The interaction of text and reader characteristics in pronoun resolution. *Journal of Memory & Language*, **42**, 545-570.

MCDONALD, J. L., & MACWHINNEY, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory & Language*, **34**, 543-566.

MCRAE, K., SPIVEY-KNOWLTON, M. J., & TANENHAUS, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory & Language*, **38**, 283-312.

MIIKKULAINEN, R. (1993). *Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory*. Cambridge, MA: MIT Press.

OAKHILL, J., GARNHAM, A., & VONK, W. (1989). The on-line construction of discourse models. *Language & Cognitive Processes*, **4**, SI263-SI286.

RODD, J. M., GASKELL, M. G., & MARSLEN-WILSON, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, **28**, 89-104.

STEWART, A. J., & GOSSELIN, F. (2000). A simple categorisation model of anaphor resolution. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp. 930-935). Hillsdale, NJ: Erlbaum.

STEWART, A. J., PICKERING, M. J., & SANFORD, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory & Language*, **42**, 423-443.

ST. JOHN, M. F. (1992). The story Gestalt: A model of knowledge-intensive processes in text comprehension. *Cognitive Science*, **16**, 271-306.

TABOR, W., JULIANO, C., & TANENHAUS, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language & Cognitive Processes*, **12**, 211-271.

TABOR, W., & TANENHAUS, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science*, **23**, 491-515.

VAN GOMPEL, R. P. G., PICKERING, M. J., & TRAXLER, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory & Language*, **45**, 225-258.

VONK, W. (1985). The immediacy of inferences in the understanding of pronouns. In G. Rickheit & H. Strohner (Eds.), *Inferences in text processing* (pp. 205-218). Amsterdam: Elsevier.

NOTES

1. According to Kehler (2002), there are three types of coherence relations: causal, contiguity, and resemblance relations. Only the first two involve world knowledge, whereas resemblance requires linguistic knowledge. Therefore, only causal and contiguity relations are handled by the DSS model.

2. The only difference from the original microworld is that the second character used to be called Jilly.

3. To be exact, $E_{i,t} = \sigma(X_{t-1}W_i + W_i X'_{t+1})$, where W_i is the ith column of W, $W_{i\cdot}$ is its ith row, the column vector X'_{t+1} is the transpose of the row vector X_{t+1}, and $\sigma(x) = (1 - e^{-x})^{-1} - 1/x$ is a sigmoidal function [with $\sigma(0) = \frac{1}{2}$].

4. In DSS–AR, $x^0_{i,t}$ can be different from $x^{\max}_{i,t}$.

5. A similar measure can be defined for ambiguities with more than two readings—for instance, by considering only the two readings with the largest belief values.

6. If there are more than two readings, only the difference between the two readings for which context influence is largest could be considered.

7. Conjunctions of a basic event and its negation were excluded. Conjunctions that violate microworld constraints [e.g., soccer $\land \neg$(B outside)] were not excluded in advance, but almost all of them resulted in context strengths below .02, excluding them in practice.

8. As compared with the left panel, the right panel shows a larger effect of bias and a smaller effect of context, resulting in partial ambiguity resolution. This is partly caused by an asymmetry in microworld knowledge: A priori, Joe is considered slightly more likely to win than Bob.

9. For the data shown here, total sentence reading times were not qualitatively different from second-fragment reading times.

10. In their first experiment, Garnham et al. (1992) found an unexplained larger congruency effect for NP2 verbs than for NP1 verbs.

11. Note that these simulations, unlike the others, include contexts resulting in context strengths lower than .02.

12. Stewart et al. (2000) did not present error rates for Experiment 3. Also, tests of the statistical significance of verb type effects were not reported.

13. Confounding between subject and first-mentioned entity makes it difficult to perform such an experiment in English.