Discourse Processing

What makes us human is really our imaginations. I think we are probably not actually homo sapiens. I think we are Pan narens. We are a chimpanzee that’s good at telling stories.

TERRY PRATCHETT

To interpret clauses and sentences, comprehenders engage in complex processes that produce intermediate products before a final interpretation is reached. During interpretation, comprehenders do a substantial amount of mental work to structure the input. So comprehenders do not treat sentences as mere lists of words. Comprehenders also engage in complex cognitive processing involving intermediate stages as they interpret discourse—interrelated sets of sentences—including narratives (stories) and expository texts (texts that try to explain how things work, like this one). This process of understanding involves building mental representations that capture features of the text itself (such as the exact words that it contains, the order in which words appear, and the syntactic structures that the speaker or author used), but the comprehender must do much more than this to figure out what the text is about. If the comprehender does not do this additional work and go beyond representing the text itself, her representations of the text will resemble an incoherent list of clauses and sentences, and she will not appreciate the ideas that the text conveys. This chapter focuses on describing the mental representations and processes that comprehenders bring to bear.
to create meaning while listening to or reading discourse. The chief focus will be on narrative texts (stories), because this is the area that has received the most attention from language scientists. This focus is not accidental, however, as the comprehension of narrative text is most closely related to the kind of attributional (explanatory) processing that people do in their daily lives (Singer, Graesser, & Trabasso, 1994). As people go about their business, they try to understand why and how events happen in the world, why people behave the way they do, and what is likely to happen in the future. Similarly, when people read stories, they try to figure out why events happen, how the different events in the story fit together, why characters do what they do, and why and how characters react to the events in the story. To do so, comprehenders combine information that is directly and explicitly signaled by the text with information that they supply themselves, including pre-existing ideas and knowledge about physical and psychological causation. This combination of explicit and implicit information helps comprehenders make sense of the sequence of events in the story, characters’ actions, their emotional responses, and so forth. Rather than being a passive process, comprehension resembles other forms of active cognitive processing. “Seeing, hearing, and remembering are all acts of construction” (Neisser, 1967, p. 10).

Current approaches to discourse comprehension focus on four main aspects of processing. First, there are a whole set of processes that are responsible for identifying the exact content of the clauses and sentences that make up the text itself. Second, there are processes that connect the actual words in the text with the ideas, objects, or events that those words refer to (these are called referential processes). Third, there are processes that are responsible for connecting the different pieces of the text to one another (these are the processes that establish textual cohesion or coherence). Finally, there are processes that are responsible for building a representation of what the text is about (these are processes involved in building a discourse representation or mental model). A good way to describe these interrelated processes is to look at some current theories that seek to explain how extended texts are processed and interpreted. The next sections will review Walter Kintsch’s construction–integration theory and related approaches, Morton Ann Gernsacher’s structure building framework, and Rolf Zwaan’s event indexing model.

**Construction–Integration Theory**

Perhaps the best known and most widely studied theory of discourse comprehension is Walter Kintsch’s construction–integration theory (Kintsch, 1988, 1998; Kintsch & van Dijk, 1978). The construction integration processing model represents a kind of production system, not because it has anything to do with speaking (although Kintsch has proposed that the model can be adapted as an account of spoken-language planning), but rather because the system is built like a particular kind of computer program—a production system. In a production system, the contents of an active memory buffer (short-term or working memory) are scanned. A set of if–then rules (or productions) is applied based on the contents of the active memory buffer. For example, one production rule might be: “If the contents of the memory buffer are empty, then input another unit of text.” Another rule might be: “If two units of text in the working memory buffer have overlapping content, then connect those two units into one larger unit.” Thus, the discourse processing system consists of a set of productions that manipulates the contents of working memory and builds coherent, structured mental representations that can be stored in a stable form in the comprehender’s long-term memory.
The construction–integration production system interprets texts by building three distinct kinds of mental representations. The least abstract mental representation is the *surface model*, which is essentially a phrase structure tree that captures the exact words in the text, along with their syntactic relations. To build the second model, interface processes take the surface model as input, do some work on that input, and output a set of *propositions* that the surface model represents. The mental representation that describes the propositions represented by the text is called the *text-base*. The text-base is close to the verbatim form of the text itself, but it can include some information that was not explicitly mentioned in the text (some examples appear below), and it does not include information about the precise wording of the original text—so some surface information is lost as the construction–integration system builds the text-base representation. Finally, at the highest level of abstraction, the construction–integration theory proposes that comprehenders build a *situation model*. The situation model describes the ideas and/or events that the text is about. This is the ultimate goal of the construction–integration system, as it is with most comprehenders. People are usually not particularly interested in the exact wording of a text (unless the person in question is a proofreader, a poet, or a playwright). Normally, people read texts because they want to know what happened and why, and they are not particularly concerned about how, exactly, that knowledge is conveyed.

We have already spent a considerable amount of time discussing how words are identified and how sentences are parsed, so we don’t need to spend additional time here discussing how the surface model is built. Just take your favorite parsing model and plug it in to do that job. Instead, let’s take some time to think about propositions and the text-base level of representation.

The text-base representation consists of a set of connected *propositions*. Propositions are defined in two ways in construction–integration theory. The first definition of a proposition is, “a predicate and its arguments.” Basically, that boils down to a verb (the predicate) and the role-players that go along with the verb (the arguments). So, in sentence (1)

(1) The customer wrote the company a complaint.

*wrote* is the predicate, *customer* is the subject/agent argument, *the company* is the indirect object/recipient argument, and *a complaint* is the direct object/theme argument. If we wanted to express that proposition in a convenient, generic notation form, it would look like this (Kintsch, 1994, p. 71):

**Proposition 1**: predicate [agent, recipient, theme]

The specific proposition in the preceding sentence could be represented in the following form:

**Proposition 1**: write [customer, company, complaint]

Propositions capture the action, state, or change of state that is being expressed in an utterance, and the arguments of the predicate indicate which characters or objects are involved in the action, as well as other information that elaborates on how the action is taking place. (Note that the definition of *argument* in Kintsch’s construction–integration theory is different than the definition of *argument* that is more common in linguistic theories. According to Kintsch, all of the partners of the predicate count as arguments. According to linguistic theories, optional partners, like locations, the specific time that the action takes place, and so forth, would be called *adjuncts* rather than arguments.)

The other way that *proposition* is defined in construction–integration theory is, “The smallest unit of meaning that can be assigned a truth value.” Anything smaller than that is
a predicate or an argument. Anything bigger than that is a macroproposition. So, wrote is a predicate, and wrote the company is a predicate and one of its arguments. Neither is a proposition, because neither can be assigned a truth value. That is, it doesn’t make sense to ask, “True or false: wrote the company?” But it does make sense to ask, “True or false: The customer wrote the company?” To answer that question, you would consult some representation of the real or an imaginary world, and the statement would either accurately describe the state of affairs in that world (i.e., it would be true) or it would not (i.e., it would be false).

Although the precise mental mechanisms that are involved in converting the surface form to a set of propositions have not been worked out, and there is considerable debate about the specifics of propositional representation (see, e.g., Kintsch, 1998; Perfetti & Britt, 1995), a number of experimental studies have supported the idea that propositions are a real element of comprehenders’ mental representations of texts (van Dijk & Kintsch, 1983). In other words, propositions are psychologically real—there really are propositions in the head. For example, Ratcliff and McKoon (1978) used priming methods to find out how comprehenders’ memories for texts are organized. There are a number of possibilities. It could be that comprehenders’ memories are organized to capture pretty much the verbatim information that the text conveyed. In that case, we would expect that information that is nearby in the verbatim form of the text would be very tightly connected in the comprehender’s memory of that text. So, for example, if you had a sentence like (2) (from Ratcliff & McKoon, 1978)

\[(2) \text{ The geese crossed the horizon as the wind shuffled the clouds.} \]

the words horizon and wind are pretty close together, as they are separated by only two short function words. If the comprehender’s memory of the sentence is based on remembering it as it appeared on the page, then horizon should be a pretty good retrieval cue for wind (and vice versa).

If we analyze sentence (2) as a set of propositions, however, we would make a different prediction. Sentence (2) represents two connected propositions, because there are two predicates, crossed and shuffled. If we built a propositional representation of sentence (2), we would have a macroproposition (a proposition that is itself made up of other propositions), and two micropropositions (propositions that combine to make up macropropositions). The macroproposition is:

as (Proposition 1, Proposition 2)

The micropropositions are:

Proposition 1: crossed [geese, the horizon]
Proposition 2: shuffled [the wind, the clouds]

Notice that the propositional representation of sentence (2) has horizon in one proposition, and wind in another. According to construction–integration theory, all of the elements of that go together to make a proposition should be more tightly connected in memory to each other than to anything else in the sentence. As a result, two words from the same proposition should make better retrieval cues than two words from different propositions. Those predictions can be tested by asking subjects to read sentences like (2), do a distractor task for a while, and then write down what they can remember about the sentences later on. On each trial, one of the words from the sentence will be used as a retrieval cue or reminder. So, before we ask the subject to remember sentence (2), we will give her a hint. The hint (retrieval cue) might be a word from proposition 1 (like horizon) or a word from proposition
2 (like clouds), and the dependent measure would be the likelihood that the participant will remember a word from the second proposition (like wind). Roger Ratcliff and Gail McKoon found that words that came from the same proposition were much better retrieval cues (participants were more likely to remember the target word) than words from different propositions, even when distance in the verbatim form was controlled. In other words, it does not help that much to be close to the target word in the verbatim form of the sentence unless the reminder word is also from the same proposition as the target word (see also Wanner, 1975; Weisberg, 1969).

Other studies using reaction time methods also support the psychological reality of propositions. If memory for texts is organized around propositions, which would mean that people extract propositions as they comprehend stories, then people should be able to access information from one proposition faster than information from two separate propositions. To test this hypothesis, Ratcliff and McKoon (1978) had people read pairs of unrelated sentences like (3) and (4):

(3) Geese crossed the horizon as the wind shuffled the clouds.

(4) The chauffeur jammed the clutch when he parked the truck.

They then had their participants perform a probe recognition task. In a probe recognition task, participants are presented with a list of words. Their task is to say as fast as possible, yes or no, whether each word appeared in a text that they had read previously. Unbeknownst to the participants, the list of words was organized so that sometimes a pair of adjacent words in the list was from the same proposition (e.g., horizon and crossed), sometimes the pair of words was from a different proposition, but the same sentence (e.g., horizon and wind), and sometimes the pair of words was from a different sentence (e.g., horizon and clutch). The dependent measure in a probe recognition study is how long it takes the participant to answer the yes-or-no question. If the representation of the test word is very active, or very accessible, people should respond very quickly; otherwise, they should be slow. In Ratcliff and McKoon's experiment, the first word in the pair serves as a retrieval cue. If the text is organized into propositions, then presenting the first word should activate other information from the same proposition more than it activates other information from the same sentence; and information from the same sentence should be more activated than information from a different sentence. This prediction was confirmed. Reading and responding to cue words like horizon caused participants to respond much faster to target words from the same proposition, the cue words had some effect (but not as big) on target words from the same sentence, and they had no effect at all on target words from the other sentence in the pair.

Other evidence for the psychological reality of propositions includes the fact that the number of words recalled from a sentence depends on the number of propositions in the sentence, when length is held constant (Forster, 1970). Error rates on recall tasks also depend on the number of propositions in the sentence. Errors increase geometrically as the number of propositions to be remembered increases, irrespective of the length of the text (Barshi, 1997; in Kintsch, 1998). Propositions tend to be recalled in an all-or-nothing fashion. That is, if any part of the proposition is recalled it is very highly likely that the entire proposition will be recalled (Goetz, Anderson, & Schallert, 1981). Overall reading time for a text depends on the number of propositions in the text, again when the number of words in different texts containing different numbers of propositions is held constant (Kintsch & Keenan, 1973). Finally, when people read stories, are given an individual word from the story, and are asked to say the first word that comes to mind (a form of free association test), the most likely response will be a word from the same proposition (Weisberg, 1969).
The final type of mental representation that people build while reading texts is the situation model (sometimes called a mental model, Johnson-Laird, 1983). The situation model is a mental simulation of the events in a story, and it captures a number of different features of the real or imaginary world that the text is about, including space, time, causality, and characters’ emotional states. One way to appreciate the importance of the situation model is to see what happens to text processing when the situation model cannot be built. Read the following paragraph and see whether you can make sense of it (from Bransford & Johnson, 1972, p. 719; see also Johnson, Doll, Bransford, & Lapinski, 1974):

*If the balloons popped, the sound wouldn’t be able to carry since everything would be too far away from the correct floor. A closed window would also prevent the sound from carrying, since most buildings tend to be well insulated. Since the whole operation depends on a steady flow of electricity, a break in the middle of the wire would also cause problems. Of course, the fellow could shout, but the human voice is not loud enough to carry that far. An additional problem is that a string could break on the instrument. Then there could be no accompaniment to the message. It is clear that the best situation would involve less distance. Then there would be fewer potential problems. With face to face contact, the least number of things could go wrong.*

If you are like most people, you will find it very difficult to understand the preceding paragraph, even if you read it over several times. The main problem here is that it is nearly impossible to figure out what the paragraph is about. That is, it is impossible to build a model of the situation or context that the words in the paragraph refer to. What does the author mean by *problems*? What kind of problems? What kind of stringed instrument is the paragraph about? What do balloons have to do with it?

It turns out that when people listen to paragraphs like this, they view them as being incoherent, and their memory for the contents of the paragraph tends to be very poor (Bransford & Johnson, 1972). Now, read the paragraph again after looking at the picture in Figure 5.1. If you are like most people, you will find it much easier to make sense of the text after looking at the picture. Why the big difference? One answer is that, without the picture, it is impossible (or nearly so) to build a situation model that captures what the text is about, so your representation of the text lacks global coherence. Without an overarching situation model, it is difficult to figure out what the words in the paragraph refer to (so you have trouble establishing reference), and it is difficult to figure out how individual sentences in the paragraph relate to preceding and following sentences, so your representation of the text lacks local coherence. With the picture in mind, you can bring to bear all of your general world knowledge about instruments and the problems that are involved in trying to impress a dream woman. This allows you to establish reference (e.g., *instrument* refers to electric guitar), and you understand how and what kinds of *problems* might arise (if the balloons pop, the speaker will fall). So the situation model, which is where general world knowledge meets the specifics of the text itself, is a vital element of discourse comprehension.

**Construction and integration**

The ultimate goal of the construction–integration system is to build a situation model describing relevant aspects of what a text is about. The system builds a surface form representation, converts that to a text-base, and then builds a situation model that reflects the contents of the text-base combined with information from general world knowledge. How does this all take place? The construction–integration account proposes that discourse processing is divided into discrete cycles. This is because there are limits on the capacity of
active or working memory, so only a small part of the text can be worked on at any given time. During any given processing cycle, only a small portion of the text is being worked on. Each processing cycle itself consists of different subcycles. The first of these subcycles is the construction phase; and the second is the integration phase.

In the construction phase, new text is brought into the system for processing. A surface form representation is built, propositions are extracted, and knowledge is activated to the degree that it is associated with the words in the text and the activated propositions. This knowledge activation phase is conceptualized as being largely or entirely automatic. That is, so long as the comprehender is paying attention and is trying to comprehend, she has little or no control over what information becomes active and available to the comprehension system. We have already seen that the semantic associates of ambiguous words become active, regardless of their relationship to the context of the sentences that they appear with (see Chapter 3). Construction–integration theory proposes similarly that general knowledge that is associated with the current contents of the active memory buffer also becomes

Figure 5.1  Contextual prerequisites for understanding: Some investigations of comprehension and recall (from Bransford & Johnson, 1972)
activated automatically. So, for example, if people are reading a story about a musical concert and they see an unambiguous word like piano, all of the properties of piano become accessible (activated), whether those properties are relevant to the current context or not (but see Tabossi, 1988). So, immediately after people see the word piano they respond quickly to probe words like heavy (because pianos are heavy), even though that property is not particularly useful to understanding a story about a concert. It is only at later stages of processing that non-useful or irrelevant associated information becomes deactivated.

Construction–integration theory adopts a model of knowledge activation very close to the TRACE account of lexical processing. According to Kintsch (1998, p. 76), “Meaning has to be constructed by activating nodes in the neighborhood of a word. This activation process is probabilistic, with activation probabilities being proportional to the strengths of connections among the nodes, and it may continue for a variable amount of time, spreading outward into the knowledge net from the source node.”

Once knowledge has been activated promiscuously, constraint satisfaction processes reduce the pattern of activated nodes, so that the remaining activated nodes are most closely relevant to the overall context or theme of the text. These activated nodes are conceptualized as a set of activated propositions in the working memory buffer. During the integration phase of processing, the propositions that are active in working memory are connected to one another and to the contents of the preceding text.

Let’s look at a specific example to see how integration works in the construction–integration system. Assume that the active memory buffer contains two propositions extracted from this mini-text (from Kintsch, 1994):

John traveled by car from the bridge to the house on the hill. A train passed under the bridge.

If the first sentence were parsed into its component propositions, the propositional representation would look like this:

(P1) predicate: TRAVEL
    agent: John
    instrument: car
    source: bridge
    goal: house
    modifier: on hill

If the second sentence were parsed into its component propositions, the propositional representation would look like this:

(P2) predicate: PASS
    object: train
    location: under bridge

With these two propositions active in the working memory buffer, the construction integration system now tries to find a way to integrate them, to build a representation that establishes a relationship between the two propositions. According to Kintsch, the production system applies an argument overlap strategy to integrate the propositions. The argument overlap strategy says, “When you have two propositions active in working memory, look for arguments in each proposition that represent the same concept. When you find overlapping arguments, use those arguments to tie the two propositions together.”

In our example mini-text, the proposition extracted from the first sentence can be integrated with the proposition extracted from the second sentence, because they share the
argument *bridge*. So what makes the two sentences fit together is that readers assume that the *bridge* that John started at is the same *bridge* that the train passed under. If readers do not make this assumption, then they will not be able to figure out how the first sentence relates to the second, their representation of the mini-text will be incoherent, and, if they remember both propositions, they will most likely remember the two as reflecting entirely separate and independent events.

In order for two propositions to be related or *integrated*, both propositions must be in an active state in the working memory buffer at the same time. However, given that working memory capacity is limited to about seven independent chunks of information (Baddeley, 1972; Miller, 1956), and given that the processes used to manipulate the contents of working memory also use up some of the available resources, only a small number of propositions can be active in working memory simultaneously. Sometimes, the comprehender will have a proposition that is active in working memory that does not relate to, and can not be integrated with, other active propositions in working memory. When this happens, comprehenders can search their long-term memory to try to find a proposition from earlier in the discourse that does relate to the “orphaned” new proposition. These *reinstatement searches* are sometimes needed to maintain coherence, but they are costly in terms of processing resources, as indicated by increased reading times at points in texts where incoming propositions cannot be directly related to immediately preceding text (Fletcher, 1981, 1986; Fletcher & Bloom, 1988; Fletcher, Hummel, & Marsolek, 1990). Sometimes, the reinstatement search will fail to supply an old proposition that can connect to the new proposition. In that case, the orphaned proposition will be purged from working memory. There is some chance that the proposition will be stored as an independent unit in long-term memory, but it is more likely that the proposition will simply be lost or forgotten. Purging propositions from working memory frees up capacity that will be needed for new propositions in the next processing cycle.

Once a text-base representation has been built and propositions have been integrated, comprehenders can update their situation model—their representation of what the text is about. As comprehenders update their situation models, they include information that is directly stated in the text, but they also use their general world knowledge to add information to the situation model that is not directly stated in the text. This process of *inference* can take many forms, and there is a debate about when and how different kinds of inferences are made (see below), but there is no question that inferred information is an important aspect of comprehenders’ situation models. For example, texts do not always explicitly state how two different propositions are related, and comprehenders must supply the “missing” information themselves (or else their representation of the text will be incoherent). For example, consider this brief story from Haviland and Clark (1974):

Mary unpacked some picnic supplies. The beer was warm.

To integrate these two sentences, the reader has to determine how they fit together. Because there is not any explicit overlap between the two sentences at the level of arguments, the comprehender needs to do some extra work to *bridge* the two sentences (so this kind of inference is called a *bridging* inference). In this case, the comprehender infers that the *beer* and some *picnic supplies* go together (because general world knowledge tells them that people often take beer along when they go on a picnic), and so the two sentences can be integrated on that basis. This process of inferencing takes time and uses up some of the available processing resources, however, and so *The beer was warm* takes longer to read in the context of *picnic supplies* than it would if the preceding sentence explicitly mentioned *beer*. Generally speaking, researchers working on discourse processing agree that bridging inferences are made “on-line,” during the process of discourse interpretation, and do not
Discourse Processing depend on any kind of special strategy on the comprehender’s part (e.g., Graesser, Singer, & Trabasso, 1994; McKoon & Ratcliff, 1992).

Previously, we reviewed evidence that propositions are a psychologically real aspect of comprehenders’ mental representations of texts. But what about the other elements of the construction–integration representational scheme? Do we really need three levels of representation: the surface form, the textbase, and the situation model? Or can we just stop at propositions? There is, in fact, considerable evidence from studies of how people remember texts that supports the distinction between surface form, the textbase, and the situation model. These experiments also provide evidence that comprehenders construct all three kinds of mental representations when they process extended discourse.

The idea that we represent more than just the verbatim form of the text comes from classic memory studies from the early 1970s. These experiments provide evidence that, while the situation model is developed on the basis of the verbatim or surface form, there really are at least two different representations being built. In these studies, people read either sentence (5), (6), (7), or (8) (Bransford, Barclay, & Franks, 1972; see also Franks & Bransford, 1974; Johnson, Bransford, & Solomon, 1973):

(5) Three turtles rested on a log and a fish swam beneath them.
(6) Three turtles rested on a log and a fish swam beneath it.
(7) Three turtles rested beside a log and a fish swam beneath them.
(8) Three turtles rested beside a log and a fish swam beneath it.

They were told beforehand that their memories for the sentences would be tested later on. The participants’ task was to try to memorize the sentences exactly as they were written. Notice that the meaning of sentences (5) and (6) is essentially the same, even though there is a slight difference in the wording of the two sentences (them in (5) is replaced by it in (6)). Both meanings are conveyed by the leftmost configuration of objects in Figure 5.2. The surface forms of sentences (7) and (8) are distinguished by the same small change in wording (it replaces them), but this small change in wording also changes the meaning. Sentence (7) goes with the middle configuration of objects in Figure 5.2, while sentence (8) goes with the rightmost configuration. If people remember the surface form of the sentences they read, then they should be equally accurate when asked to remember the exact wording of a sentence that they have previously read. So, after a short retention interval (a few minutes doing a distractor task), people were presented with pairs of sentences, one of which they had read, and one of which they had not, and they were asked to pick out the exact sentence that they had read. People were much less accurate picking between sentences...
(5) and (6), which express the same meaning. They were much more accurate picking between sentences (7) and (8), which express different meanings. These findings indicate that the memory for surface form decays very rapidly (within a couple of minutes), but memory for meaning or gist is more durable. A person can identify the exact wording of a sentence she read previously if her situation model representation is consistent with only one of the available choices. But if two sentences convey the same meaning (i.e., they map onto the same situation model), a person has much more difficulty remembering which version she read, even when she tries to memorize the exact wording of the sentences.

More recent research shows that, in addition to representing spatial relations in their situation models, people also represent temporal relations, that is how events in a story are laid out in time (Rinck, Hähnel, & Becker, 2001). This experiment used the same study-test procedure as the classic Barclay experiments, but in addition to conveying spatial information, some of the test sentences conveyed temporal information. The test sentences (which were presented in German) described two events as occurring simultaneously, with a third event following, as in sentence (9):

(9) The piano was heard together with the harp and the soprano sang along with it.

In German, the feminine gender pronoun ihr (“it”) can only refer to the harp. In half of the test sentences, the feminine pronoun ihr was replaced by the masculine pronoun ihm, which can only refer to the piano. But whether the pronoun refers to the harp or the piano, both versions map onto the same sequence of events (or the same “temporal model”). As was the case when to-be-memorized sentences referred to the same spatial model, subjects were not able to recognize which version (the one with ihr or the one with ihm) that they had actually studied when they were tested following a brief distractor task.

Further evidence for the independent existence of surface, text-base, and situation models come from studies of more lengthy texts. For example, consider the following brief paragraph (from Fletcher & Chrysler, 1990; see also Kintsch, Welsch, Schmalhofer, & Zimny, 1990):

George likes to flaunt his wealth by purchasing rare art treasures. He has a Persian rug worth as much as my car and it’s the cheapest thing he owns. Last week he bought a French oil painting for $12,000 and an Indian necklace for $13,500. George says his wife was angry when she found out that the necklace cost more than the carpet. His most expensive “treasures” are a Ming vase and a Greek statue. The statue is the only thing he ever spent more than $50,000 for. It’s hard to believe that the statue cost George more than five times what he paid for the beautiful Persian carpet.

In Randy Fletcher and Sue Chrysler’s study, participants read the preceding story. Later, they were asked to say whether test sentences had appeared verbatim in the story or not. The test sentences were designed to probe the surface form, the text-base, or the situation model. To test the surface form, the test sentence either contained the exact wording that had appeared in the text, or else one of the words was replaced by a synonym (e.g., rug in He has a Persian rug worth as much as my car would be replaced with the word carpet). If participants correctly report that He has a Persian carpet is a new phrase that did not appear in the story, then they have an accurate surface form representation. To test the text-base, the set of propositions that the participants built from the text, the word necklace in his wife was angry when she found out that the necklace cost more than the carpet was replaced by the word painting. It is true (according to the story) that the painting cost more than the carpet, but that is not what the wife was angry about. So, if participants correctly reject the sentence with painting in place of necklace, they must have remembered the proposition the wife was upset that the necklace cost more than the carpet. Finally, to test the
situation model, the word carpet was replaced by the word vase to make the test sentence his wife was angry that the necklace cost more than the vase. If participants falsely recognize this last test sentence, this violates the correct situation model, because the vase is stated to be worth much more than the carpet. When participants’ recognition memory was tested, they almost never made mistakes about the situation model, they sometimes made mistakes about the specific propositions that were in the story, and they frequently made mistakes about the specific wording of the story. Differences between the error rates for different kinds of questions shows that the questions were tapping into different kinds of mental representations.

Kintsch and his colleagues (1990) also found evidence for separate memory strengths for different kinds of text representations. In this study, participants read short paragraphs and, after varying retention intervals, judged whether test sentences were exactly the same as sentences they had read in stories previously. When the test sentences violated the situation model, participants almost never said they appeared in the stories, even after a delay of four days between reading the story and being tested (see Figure 5.3, the bottom line labeled “New inap.” reflects how often participants false alarmed and said that a new sentence that violated the original situation model was really an old sentence from the original story). By contrast, participants were very likely to say “yes” to new sentences when the new sentences were paraphrases of sentences that had appeared in the stories, and this likelihood increased as the delay between initial reading and test increased (see Figure 5.3, the open circles labeled “Para.” for “paraphrase”). This shows that the surface form representation is fairly weak and decays very rapidly. The “Inf.” curve represents new test sentences that reflect information that was not directly stated in the original text, but could be inferred from it. (For example, if you read that John sent Susan a love letter, you might infer that John wanted...
to go on a date with Susan.) The graph shows that, immediately after reading the story, participants knew the difference between information that was explicitly stated in the story and inferences that they drew from the explicitly stated information. But as time elapsed, participants were more and more likely to “remember” that inferred information was directly stated in the original story. This kind of mistake is often called a “source memory” error by memory researchers, because although the information in the inference is accurate, people make mistakes about how they acquired the information (see, e.g., Jacoby, Woloshyn, & Kelley, 1989). Surface form, propositions, and the situation model are each represented separately and remembered for different lengths of time by people who read stories. The situation model is the strongest and longest lasting representation, the text-base is the next strongest, and the surface form is the weakest and shortest lived (but some surface information does survive, even over the long haul; Gernsbacher, 1985; Keenan, MacWhinney, & Mayhew, 1977).

So, to summarize, the construction–integration production system builds three separate mental representations. The surface form model represents the exact words in the text and their syntactic relations. The text-base represents a set of connected propositions extracted from the surface form. The situation model includes information directly and explicitly stated in the text plus information that comprehenders supply themselves in the form of inferences. Text is processed in cycles. In each cycle, comprehenders input a few propositions’ worth of text and knowledge associated with the inputted text becomes automatically activated. In the integration phase comprehenders connect new propositions to previously processed propositions, draw inferences, and update their situation models.

The Structure Building Framework

The structure building framework (Gernsbacher, 1990) is a theory of discourse processing that has influenced researchers for the past two decades. Like the construction–integration account, the structure building framework seeks to explain how comprehenders build mental representations of extended discourse. But while Kintsch’s theory is largely restricted to spoken and written text, Morton Ann Gernsbacher’s theory can be applied to the comprehension of both verbal and non-verbal materials (like picture stories). And while Kintsch’s account explains discourse comprehension as being the product of special purpose discourse comprehension mechanisms, the structure building framework appeals to general purpose cognitive mechanisms to explain how discourse is interpreted and remembered. According to Gernsbacher, the processes that are responsible for discourse comprehension are also responsible for other cognitive tasks that may not be directly related to language. Some specifics will help illustrate these points.

The structure building framework proposes that general purpose cognitive mechanisms are responsible for discourse comprehension. To understand a story, comprehenders begin with the process of laying a foundation. The foundation is based on the information that arrives first, just as laying a foundation is the first thing that happens when you build a house. Two additional general processes, mapping and shifting, are used to continue building the structure once the foundation has been laid. The mapping process connects incoming information to the foundation as long as the incoming information is related to, or coheres with, the preceding information. If the new information is not related to the preceding information, comprehenders undertake the process of shifting to build a new substructure. Thus, the comprehenders’ mental representation of a story, or any set of events with a coherent structure, consists of a foundation, plus an appropriate number of connected
substructures. The story representation as a whole then, consists of “several branching substructures (Gernsbacher, 1995, p. 49),” with the branches terminating at the foundation.

According to the structure building framework, the metaphorical building blocks that comprehenders use to create their representations for stories consist of activated memory nodes (similar to the nodes in Collins and Quillian’s semantic network theory; or propositions in Kintsch’s construction–integration framework). When memory nodes become activated by a text, they send out processing signals. The processing signals that the memory nodes send out lead to either enhancement or suppression of other memory nodes. The enhancement mechanism increases the activation of memory nodes that are related to the input and the currently activated set of memory nodes. The suppression mechanism decreases the activation of memory nodes.

This fairly simple account can be used to explain a wide variety of experimental results. For example, if the process of laying a foundation is psychologically real, then we should see evidence that first-mentioned parts of texts are dealt with differently than later-arriving parts. One piece of evidence is that comprehenders take longer to process a given word if it appears as the first word in a sentence than if it appears later (Aaronson & Scarborough, 1976). People also process the first sentence in a paragraph more slowly than other sentences (Cirilo & Foss, 1980; Haberlandt, 1980). Similar effects occur for non-verbal materials. For example, when viewing a picture story, people spend more time looking at the first picture in the story than the other pictures (Gernsbacher, 1996). This slow-down would be expected if comprehenders are dedicating special effort to encoding the first-mentioned parts of a story or text (i.e., laying a foundation).

Given this extra effort at encoding, we might expect that information arriving first in a text should enjoy some kind of special status in the mental representation of the text, and it turns out that it does. This advantage of first mention has been demonstrated in experiments where people read or listen to sentences like (10) (Gernsbacher & Hargreaves, 1988; Gernsbacher, Hargreaves, & Beeman, 1989).

(10) Tina beat Lisa in the state tennis match.

After reading this sentence, people verify that the first-mentioned person (Tina) appeared in the sentence much faster than they verify that the second-mentioned person (Lisa) did, suggesting that in the comprehender’s mental representation of sentence (10), the first-mentioned person enjoys a higher degree of activation. One might offer a counter-explanation to the advantage of first mention hypothesis by appealing to the fact that Tina is the grammatical subject of the sentence (a prominent and important syntactic position), but verification times for Tina are still faster than verification times for Lisa when both are part of a conjoined subject (as in Tina and Lisa beat Susan and Marsha in the state tennis match). Similarly, the semantic role that the individual characters play also does not override the advantage of first mention. In sentence (10), Tina is the subject of the sentence and she is also the thematic agent (do-er, or initiator) of the action. In (11), Tina is still the first-mentioned participant, but now she is the patient (do-ee, or recipient) of the action:

(11) Tina was beaten by Lisa in the state tennis match.

People still verify that Tina appeared in sentence (11) faster than they verify that Lisa was in sentence (11), suggesting again that people pay special attention to first-mentioned participants, whether the first-mentioned character is the initiator or recipient of the action described in the text.

Evidence for mapping and shifting processes can be found by looking at how textual cohesion and coherence affect processing and memory for texts. Mapping occurs when incoming information is highly related to the part of the text that is currently being
processed. Shifting occurs when incoming information does not closely relate to the material currently being processed. Shifting should produce two effects. First, because the process of shifting takes up processing resources, comprehenders should slow down at points in a text where coherence breaks occur, because it takes more time for them to shift and start building a new substructure than it does to map the same information onto an ongoing substructure. Second, because shifting results in the construction of a new substructure, information from previous portions of the text should become less available after shifting has occurred. Both of these predictions enjoy support from experiments on text processing. The first prediction is verified by experiments showing that parts of texts that follow topic shifts are processed more slowly than texts that maintain the current topic (Mandler & Goodman, 1982).

Other experiments show that information becomes less available after the comprehender has shifted and started a new substructure. For example, clause structure determines how accessible concepts are after a sentence has been processed (Caplan, 1972). Sentences (12) and (13) each contain two substructures, each of which consists of a clause.

(12) Now that artists are working fewer hours, oil prints are rare.
(13) Now that artists are working in oil, prints are rare.

In (12), the word *oil* appears in the second clause (the second substructure). In (13), *oil* appears in the first clause (the first substructure). So, according to the mapping and shifting hypotheses, *oil* should be more accessible at the end of sentence (12), because it is part of the most recently constructed substructure, than at the end of the sentence (13), because there it is part of the first substructure and comprehenders will have shifted and built a new substructure after they read the word *oil*. Participants more quickly verified that the word *oil* appeared in the test sentence at the end of sentence (12) than at the end of sentence (13), consistent with the mapping and shifting hypotheses. Note that the word *oil* is followed by the exact same three words in both (12) and (13), so the pure effect of recency cannot explain this outcome. Similar loss of information occurs at important boundaries in non-verbal picture stories (Gernsbacher, 1985). Participants are less able to remember the left-right orientation of pictures (a surface feature that does not affect meaning) from picture stories after they finish “reading” the story than while they are in the process of comprehending the story. Further, the process of shifting can help explain some of the differences between good and poor comprehenders. Comprehenders who score lower on tests of general verbal ability, such as the Verbal section of the SAT, are less able to access recently comprehended information, which might indicate that they shift and build new substructures more often than comprehenders who are better able to comprehend what they read (Gernsbacher, Varner, & Faust, 1990).

The hypothetical processes of enhancement and suppression have also been supported by research findings. *Enhancement* is viewed as an automatic process whereby knowledge that is related to the current text is activated rapidly, without conscious volition (as long as the comprehender is paying minimal attention), and proceeds in an uncontrolled fashion, at least initially. Findings from probe recognition experiments and reading time studies indicate that information is activated as people listen to or read stories, whether that information is relevant to the context or current topic of the discourse or not (e.g., Duffy, Henderson, & Morris, 1989; Onifer & Swinney, 1981; Rayner & Duffy, 1988; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Swinney, 1979). For example, if comprehenders read sentence (14),

(14) The teacher has a lot of patience.

and they are asked to judge whether a target word is semantically related to the sentence, they have a hard time rejecting the word *hospital*. That is because when they see the word...
patience, they activate the phonologically related word patients, which is semantically related to the word hospital (Gernsbacher & Faust, 1991). So, even though the word patients has a completely different meaning than the word patience, the similarity in the way they sound creates an associative relationship, and this associative relationship means that we have to activate both meanings, even though we only need one.

This automatic, widespread activation of associated knowledge occurs very rapidly, within a few hundred milliseconds after a word is encountered, but if the comprehender is to end up with a coherent representation of what the text is about, this automatically activated but irrelevant information must be removed from the comprehender's representation of the discourse. That is where the process of suppression comes in. Suppression operates on those activated memory nodes that are not closely related to the topic or theme of the ongoing discourse. When the unrelated information is removed, the resulting discourse representation is less cluttered with irrelevant details, and hence more coherent. The process of suppression is conceptualized as a less automatic process than enhancement. It takes longer to work than enhancement does (e.g., Neely, 1977; Wiley, Mason, & Myers, 2001) and it is more variable across individuals. In fact, differences in suppression ability may underlie differences in people's ability to understand texts.

Differences between individuals in how well they are able to suppress irrelevant information have been demonstrated in a series of studies involving lexically ambiguous words like spade. As we saw in Chapter 3, reading or listening to words like spade leads to the automatic activation of all of the meanings related to the word. So, when readers see the word spade, they activate both the playing card meaning and the shovel meaning. This automatic activation of multiple meanings occurs even when the word spade appears in a context that makes only one of the meanings appropriate. In Sentence (15), only the shovel meaning of spade fits with the rest of the sentence:

(15) The gardener dug with the spade.

However, if the word ace is presented, and people are asked to judge whether its meaning is related to the preceding sentence, they take a long time to say "no", because the playing card meaning is activated by the word spade. To get rid of this irrelevant meaning, the mechanism of suppression starts to operate on the activated playing card meaning, and after a short period of time, a few hundred milliseconds, the "playing card" meaning of spade is no longer activated—it has been suppressed.

How does suppression work? Is it as automatic as enhancement? There are a number of reasons to think that suppression is not just a mirror image of enhancement. First, suppression takes a lot longer to work than enhancement does. Second, while knowledge activation (enhancement) occurs about the same way for everyone, not everyone is equally good at suppressing irrelevant information, and this appears to be a major contributor to differences in comprehension ability between different people (Gernsbacher, 1993; Gernsbacher & Faust, 1991; Gernsbacher et al., 1990). For example, Gernsbacher and her colleagues acquired Verbal SAT scores for a large sample of students at the University of Oregon (similar experiments have been done on Air Force recruits in basic training, who are about the same age as the college students). Verbal SAT scores give a pretty good indication of how well people are able to understand texts that they read, and there are considerable differences between the highest and lowest scoring people in the sample. This group of students was then asked to judge whether target words like ace were semantically related to a preceding sentence like (15), above. Figure 5.4 presents representative data from one of these experiments. The left-hand bars show that the ace meaning was highly activated for both good comprehenders (the dark bars) and poorer comprehenders (the light bars) immediately after the sentence. After a delay of one second (a very long time in language
The good comprehenders had suppressed the contextually inappropriate “playing card” meaning of *spade*, but the poor comprehenders still had that meaning activated (shown in the right-hand bars of Figure 5.4).

Further evidence for the non-automatic nature of the suppression process comes from studies like the Gernsbacher experiments, but using slightly different experimental tasks. One potential problem with the semantic judgment task used by Gernsbacher is that it can encourage participants in the experiment to think about different meanings of the test words and it requires an explicit, conscious judgment of how the sentence and the target word go together. Thus, people’s responses might be influenced more by their decision-making processes than by processes involved in interpreting the text itself. To try to get around this problem, Debra Long and her colleagues manipulated the kind of task that participants engaged in, and they tried to find tasks that could be done on the basis of more automatic mental processes (Long, Seely, & Oppy, 1999). When Long and her colleagues used a naming task (which is thought to be relatively immune to strategic or controlled mental processes; McKoon, Ratcliff, & Ward, 1994; Potts, Keenan, & Golding, 1988), rather than a semantic judgment task, both better and poorer comprehenders showed the same pattern of meaning activations. Both groups showed a high degree of activation for contextually inappropriate meanings immediately after reading sentences like (15). Both groups appeared to suppress the contextually inappropriate meaning after a delay of 850 ms. The experimental task was changed to lexical decision in a second experiment and meaning judgment in a third. Performance on both lexical decision and meaning judgment is thought to reflect strategic mental processes. Under these conditions, less skilled comprehenders again appeared to keep the contextually inappropriate “playing card” meaning active even after significant delays, consistent with the original Gernsbacher results. Hence, it appears that less skilled comprehenders do have trouble suppressing contextually inappropriate meanings, and this deficit is especially apparent in tasks where performance requires comprehenders to manage conflict between different sources of information, such as an ambiguous word and the context it appears in, in order to produce a response.

**Figure 5.4** Estimated activation of inappropriate meanings based on a semantic judgment task (from Gernsbacher & Faust, 1991). RT = reaction time; hphone = homophone; nonhphone = nonhomophone.
To sum up, the structure building framework says that we use general purpose cognitive mechanisms to process and understand discourse. Comprehenders begin by laying a foundation, and then they either map incoming information onto the current structure, or they shift and build a new substructure. Processes of enhancement and suppression manipulate the activation levels of memory nodes. Shifting too often can lead to incoherent mental representations of texts. Failing to suppress activated but irrelevant information can also impair comprehension.

The Event Indexing Model

The Event Indexing Model (EIM) is first and foremost a theory about how people build situation models from narrative texts (Zwaan, Langston, & Graesser, 1995). According to the EIM, the purpose of the discourse comprehension system is to understand the “goals and actions of protagonists … and events that unfold in the real world or some fictional world” (Zwaan et al., 1995, p. 292; see also Zwaan & Rapp, 2006). To represent these story elements, five core aspects of stories are tracked, and each event in the story is indexed or tagged according to each of the five core features: The time frame over which the event occurs (time), the characters that are involved in the event (protagonists), the causal connection of the current event to preceding and following events (causation), the spatial location(s) where the events occur (space), and how the event relates to a protagonist’s goals (motivation). Similar to the structure building framework, the EIM conceptualizes events as activated memory nodes, and the representation of a story consists of a set of memory nodes and the connections between them. Each memory node is coded for the five previously mentioned features, and as each new piece of the text is processed, it is evaluated as to how it relates to previously activated memory nodes. So, each time a new piece of text is processed, the comprehender updates the situation model to reflect the information provided by the text. Different pieces of text can require updating of different features of the event index.

Sometimes, new information in a text elaborates on the elements of a previously activated event node. But sometimes, new information indicates a break between the previously activated event and the new information. According to the EIM, if there is a discontinuity on one or more of the five features, the current event node is deactivated and a new node is activated. This process is similar to the shifting process in the structure building framework, and discontinuities in stories should produce measurable processing costs (because shifting to work on a new event node is more complicated than continuing to map incoming information onto a previously activated event node). So, one way to test the EIM is to see how people respond when a new piece of text creates a discontinuity between the activated memory node and the new information provided by the text. In fact, people do process parts of texts that create discontinuities slower than parts of texts that can be mapped directly onto a previously activated memory node (Zwaan, Magliano, & Graesser, 1995). For example, when the text explicitly signals a temporal discontinuity (e.g., it says, *A day later* …), concepts mentioned just before the discontinuity are less accessible than in an equivalent story that does not signal a temporal discontinuity (e.g., it says, *A moment later* …; Zwaan, 1996). Texts can indicate that currently described events are part of flashbacks, and so occurred a long time ago in the virtual world described by the text. When the recency of the flashback episode is manipulated (i.e., the text either says the flashback episode happened recently or a long time ago), information from episodes that is described as taking place a long time ago is less accessible than information that is described as happening more recently (Claus & Kelter, 2006; see also Kelter, Kaup, & Claus, 2006).
The Event Indexing Model

Another way to assess the EIM is to see how comprehenders organize and remember the events in texts. According to the EIM, parts of texts that have overlapping values in the event index should be connected together in the comprehender’s mental representation of the story. For example, two sub-events that occur at the same time should be represented more closely than two sub-events that occur at different times. Take a minute to read the story “The Czar and His Daughters.” Notice that the action of _dragging_ takes place at the same time as _crying_. Normally, these two concepts would be unrelated—they are semantically very different, and people tend not to associate the two. But because the two events are connected by the temporal structure of the story, comprehenders should create a connection between _crying_ and _dragging_ when they read the story. By contrast, the actions of _walking_ and _crying_ are unrelated both in general terms (for the same reasons _crying_ and _dragging_ are unrelated), and the event structure of the story also does nothing to bring the two actions together. To test whether the story affects how people view the relationship between _crying_ and _dragging_, Rolf Zwaan and his colleagues (Zwaan et al., 1995) asked people to read stories like “The Czar and His Daughters” and then to perform a categorization task. To perform the categorization task (also known as a _clustering_ task), the subjects read a list of verbs and placed the verbs inside a set of boxes. The subjects were told to place two verbs in the same box if they thought the verbs “belonged together” (Zwaan et al., 1995, p. 294). Subjects who had read the story were far more likely to place _dragging_ and _crying_ in the same box than they were to place _walking_ and _crying_ together, but subjects who had not read the story were just as likely to place _walking_ and _crying_ together as they were to place _dragging_ and _crying_ together. Pairs of verbs that were related on the other event indexing dimensions (space, causation, entities, and goals) were also likely to be grouped together by subjects who had read the story and not by subjects who had not read the story. These results indicate that people use all five of the event indexing dimensions to organize their representations of stories.

Recent neuroimaging results also support separate indexing of different story characteristics (Ferstl, Rinck, & Von Cramon, 2005). Evelyn Ferstl and her colleagues manipulated whether words in a story conveyed temporal information (_Markus’ train arrived at the station 20 minutes before Claudia’s_) or emotional information (_Sarah couldn’t remember that she had ever been so sad_; Ferstl et al., 2005, p. 726). Emotion-conveying words led to increased brain activity in posterior ventromedial prefrontal cortex (see the green-marked regions in Plate 8). Words that conveyed temporal information produced increased brain activity in a different set of brain regions (marked in yellow and red in Plate 8), including parts of the frontal and parietal cortices on both sides of the brain. Additional fMRI and positron emission tomography (PET) studies also suggest that different kinds of indexing processes are supported to different degrees by different

---

**THE CZAR AND HIS DAUGHTERS (GRAESSER, 1981)**

Once there was a Czar who had three lovely daughters. One day the three daughters went walking in the woods. They were enjoying themselves so much that they forgot the time and stayed too long. A dragon kidnapped the three daughters. As they were being dragged off, they cried for help. Three heroes heard their cries and set off to rescue the maidens. The heroes fought the dragon. Then the heroes returned the daughters to their palace. When the Czar heard of the rescue, he rewarded the heroes.
networks of brain regions. Stories that call on people to infer characters’ mental states produce different patterns of brain activity when compared to stories that require inferences about physical causes (Ferstl, Neumann, Bogler, & Von Cramon, 2008; Fletcher et al., 1995; Mason, Williams, Kana, Minshew, & Just, 2008). Of course, when people understand a story, these different indexing dimensions are normally integrated into a coherent whole (Rapp & Taylor, 2004).

**Modeling space, time, protagonists, and motivation**

**SPACE**

The EIM proposes that we use our general perceptual apparatus to build situation models from texts. One of the main tasks that our perceptual apparatus does for us is modeling three-dimensional space, so that we can navigate through the world, pick out perceptual targets for detailed processing and evaluation, predict how objects will move, and so forth. It is not surprising, then, that spatial models are an important aspect of discourse understanding. To comprehend stories, we build an internal representation of the space that the events in stories take place in, and we track the movements of characters through this virtual space (e.g., Black, Turner, & Bower, 1979; Bower & Morrow, 1990; Bower & Rinck, 2001; Glenberg, Meyer, & Lindem, 1987; Morrow, Bower, & Greenspan, 1989; Morrow, Greenspan, & Bower, 1987; Rapp, Klug, & Taylor, 2006). Languages provide us with numerous ways to mark spatial relations (e.g., over there, in, behind, next to, to the left, and so forth). Although different languages have different ways of expressing spatial relationships (Choi, McDonough, Bowerman, & Mandler, 1999), spatial modeling is a general feature of discourse comprehension. Comprehenders assume that the information in stories will be consistent throughout with regards to the way space is described, and it is more difficult to process parts of texts that are inconsistent with previously inferred spatial relations than parts of texts that are consistent with previously inferred spatial relations (de Vega, 1995).

In some of the classic experiments on spatial modeling of texts from Gordon Bower’s lab, research subjects start by memorizing the layout of a fictitious space (Bower & Morrow, 1990; Bower & Rinck, 2001; Morrow et al., 1987, 1989). Subjects continue to look at and memorize the layout until their accuracy on a test reaches a very high level. After they successfully memorize the layout of the imaginary building, including the locations of particular objects within the building, participants then listen to a story about a character moving through the memorized space. During the story, the research subjects are interrupted periodically by a visual target word on a computer screen. Their task is to react as quickly as possible to the word (by saying whether the object is present in the imaginary building, or by indicating whether the target on the computer screen is a word or not, or simply by naming the object). People’s reaction time on the test depends on how far the main character is from the named object in the virtual, or imaginary space. If the main character in the story is close to the tested object, then people respond quickly. If the main character is distant from the tested object, then people respond more slowly. Findings like these indicate that people represent space in stories in an anological form, which means that their mental representation of space in stories preserves features of actual, three-dimensional, real-world space (as shown in the general mental representation literature in mental rotation and image scanning experiments; e.g., Kosslyn, 1973; Shepard & Metzler, 1971). So, people who listen to stories about a character moving through an imaginary space behave as though there is a “spotlight of attention” following the main character through the space. Objects that are in close proximity to the main character enjoy higher than normal activation because the comprehender’s attention is focused on the main character, but attention is not limited to the main character (hence the spotlight metaphor).
One of the benefits of having a spotlight (rather than a laser beam) of attention is that objects that are likely to be mentioned in the near future enjoy a higher than normal degree of activation, and so comprehenders can access and integrate those concepts more quickly once they are mentioned.

One of the potential concerns about Bower’s map task experiments is that subjects had to spend a considerable amount of time memorizing the layout of the spaces they would hear about later, and this is not a normal part of the way people experience language. People also sometimes object that the probe recognition task is unnatural. However, when stories refer to spaces that people are already familiar with (so they do not have to memorize anything special for the experiment), the same kinds of effects occur (Glenberg et al., 1987), whether activation of a concept is assessed by the probe recognition task or a naturalistic reading task. Objects that are in close virtual proximity to a currently focused character enjoy a higher degree of activation than objects that are further away from the currently focused character in the comprehenders’ mental model of the situation.

**TIME**

Comprehenders model temporal relations in narratives by taking advantage of real-world knowledge and explicit cues in the discourse (e.g., de Vega, Robertson, Glenberg, Kaschak, & Rinck, 2004). Speakers and listeners share schematic knowledge about the temporal organization of commonly occurring events (e.g., eating food comes after ordering food at a restaurant), and this schematic organization affects the way people tell stories (Barsalou & Sewell, 1985). Comprehenders model time in stories by applying a default strategy, modified by the presence of explicit linguistic cues. The default strategy is to assume that events in the world occurred in the order that they were described in the story, but that default assumption can be overridden by schematic knowledge or by explicit cues in the discourse. If the information conveyed explicitly by the text activates a schema, comprehenders will assume that the events conform to the order specified by the schema, unless the text directly indicates something else. When there is no pre-existing schema, or the schema leaves the order of the events unspecified, comprehenders apply the *temporal iconicity* heuristic, and assume that the events occurred in the order they were mentioned in the story.

Speakers can decide to explicitly signal that events are described in a different order than they occurred using connectives like *before* and *after*, or using explicit time tags (e.g., *at noon*, *at twilight*). So, a speaker could say *Before Megan ate lunch, Kristen walked to the park*. However, when a text describes events in a different order than they happened, people take a little extra time to process that text (Mandler, 1986). Violations of temporal iconicity also produce different patterns of neural activity, as indicated by ERP measures (Münte, Schultz, & Kutas, 1998; Figure 5.5). Münte’s study showed that sentences that violate temporal iconicity (the *before* sentences) produce a greater negative deflection in the ERP signal than sentences that describe events in the same order that they happened (the *after* sentences). The size of the difference correlated with subjects’ working memory capacity, so this supports the idea that comprehenders model the events in the order that they are described, but when temporal iconicity is violated, they have to mentally flip the order of the events. This flipping process uses up working memory resources.

**PROTAGONISTS**

Stories are centered around characters, so it is not surprising that comprehenders include a wide variety of features of protagonists in their mental models in stories. Characters’ mental states are modeled extensively as comprehenders process stories. So, for example,
in the Bower map-task experimental paradigm, objects that are in a room that a main character is thinking about enjoy a higher degree of activation and accessibility than other objects that may be associated with the virtual story space. This modeling of characters’ mental states extends to modeling of their perceptual experiences. In other words, comprehenders adopt the visual perspective of characters in stories and view the virtual story world through their eyes. As a result, objects in the story that are out of the focused character’s line of sight are less accessible to the story comprehender (are less activated in the story comprehender’s situation model) than objects that are “visible” to the focused character, and comprehenders’ reaction times to virtually occluded objects are slower than reaction times to objects that are visible from the focused character’s perspective (Horton & Rapp, 2003).

Comprehenders also model character’s emotional states, and use them to draw inferences about how characters will respond to events in stories (Gernsbacher, Goldsmith, & Robertson, 1992; Gernsbacher, Hallada, & Robertson, 1998; Gernsbacher & Robertson, 1992). So, if you are reading a story about someone winning the state tennis match, you would be surprised and respond more slowly if that character were described later as being sad or depressed. Comprehenders also model aspects of characters’ personalities, this information is accessible when the characters are mentioned, and may serve as the basis for inferencing. Have a look at the story about the character Carol (from Peracchi & O’Brien, 2004; p. 1046; see also Rapp, Gerrig, & Prentice, 2001):

**Figure 5.5** ERP data from Münte et al., 1998 (p. 71)
Test sentences: After/before the scientist submitted the paper, the journal changed its policy. The ERP signal in the circle shows increased negativity for the before sentences during processing of the second clause.
Context Version 1:
Carol was known for her short temper and her tendency to act without thinking. She never thought about the consequences of her actions, so she often suffered negative repercussions.

Context Version 2:
Carol was known for her ability to peacefully settle any confrontation. She would never even think to solve her problems with physical violence.

Continuation:
Carol was fed up with her job waiting on tables. Customers were rude, the chef was impossibly demanding, and the manager had made a pass at her just that day. The last straw came when a rude man at one of her tables complained that the spaghetti she had just served was cold. As he became louder and nastier, she felt herself losing control. Without thinking of the consequences, she picked up the plate of spaghetti, and raised it above the rude man’s head.

Test Word: dump

In Peracchi and O’Brien’s experiment, comprehenders either read version 1 of the context or version 2 of the context before reading the continuation. The continuation describes a sorely trying day at the restaurant, and the question is whether comprehenders represent information about the focused character’s personality, and use that information to predict how the story will turn out. When the target word dump was presented to readers after the sentence She picked up the plate of spaghetti, and raised it above the rude man’s head, subjects responded to it faster if Carol had previously been described as having a short temper, but no faster than a control condition if she had previously been described as being calm and peaceful. So it appears as though comprehenders do model characters’ personality traits, they activate those models when the character is mentioned in the text, and they use the modeled personality traits to predict how things will turn out in the story.

MOTIVATION

Another way comprehenders organize their representations of narrative texts is by keeping track of protagonists’ (main characters’) goals (Egidi & Gerrig, 2006; Singer & Halldorson, 1996; Singer, Halldorson, Lear, & Andrusiak, 1992; Singer & Richards, 2005; Suh & Trabasso, 1993; Trabasso & Suh, 1993; Trabasso, van den Broek, & Suh, 1989). Narratives frequently have a complex goal structure, and so information about characters’ goals increases and decreases in accessibility as comprehenders process the narrative. Sometimes, goals are nested within goals. In Soyoung Suh and Tom Trabasso’s seminal work on goal inferences, they presented people with stories that had complex goal structures. For example, a character called Jimmy wanted a new bike (first-mentioned and superordinate goal), so he tried to get a job (second-mentioned and subordinate goal). Half of the time (goal success condition), the superordinate goal was described as being satisfied (Jimmy’s mom agrees to buy him a bike) before the subordinate goal was mentioned. The other half of the time, the superordinate goal was thwarted (Jimmy’s mom does not agree to buy him a bike) before the subordinate goal was mentioned. When the superordinate goal is thwarted, presumably Jimmy still has that goal (at least, the text does not state that he has abandoned it). Under those conditions, the superordinate goal remains more active in the comprehender’s mental representation of the story. This is demonstrated by the fact people mention the superordinate goal more in the goal failure condition when they are asked to talk out loud while reading the story (see Figure 5.6, “goal failure” condition). When the goal succeeds (Timmy’s mom buys a bike), that superordinate goal becomes less active in comprehenders’
Discourse Processing

representations of the story (see Figure 5.6, “goal success” condition). These activation differences are also reflected in differences in reaction times to goal-related target words between the goal failure and goal success conditions. Subjects respond to target words related to the superordinate goal faster in the goal failure condition (where the main character is still pursuing the goal) than in the goal success condition (where the main character has moved on and is now motivated by some other goal).

Comprehenders’ ability to track goals in narratives is actually very sophisticated, as they appear to be able to track multiple goals of multiple characters, and do so to the extent that the characters are salient, prominent, or central to the narrative. Further, comprehenders are able to recognize when different characters’ goals are in conflict (Magliano, Taylor, & Kim, 2005; Richards & Singer, 2001).

Causation, Cohesion, and Coherence in Discourse Encoding and Memory

The modeling of causal relations is an important aspect of the EIM, but it is also a core issue in discourse processing, interpretation, and memory more generally. More specifically, causal connections between propositions seem to be the glue that holds different parts of narratives together. Recall that, in reviewing the construction–integration theory, the notion of argument overlap was introduced. Argument overlap occurs when two adjacent pieces of text contain information that refers to the same object, character, or concept. When this occurs, two propositions can be linked together on the basis of this overlapping information. This kind of textual connection is sometimes referred to as cohesion—different parts of a text go together by virtue of having common elements that refer to the same thing in the world. Cohesion is one factor that can contribute to textual coherence—the notion that different parts of a text fit together somehow. Cohesion is an important element of the mental representation of extended discourse, but some researchers think
that construction–integration’s focus on cohesion as the prime factor that ties propositions together is a mistake. They note that there are things besides argument overlap that can tie propositions together and that texts that are highly cohesive are not always highly coherent. Consider, for example, the following paragraph:

*It was a sunny day in the city on the hill. Hills are higher than valleys. The central valley contains a number of interesting museums. Museums often have antique weapons. Weapons of mass destruction are a threat to security. Linus’ security blanket needs washing.*

If we asked people to judge whether the preceding paragraph was coherent, most of them would say it’s pretty incoherent. This is so even though each proposition in the paragraph can be linked to the preceding proposition because of argument overlap. According to construction–integration theory, this paragraph should be rated as highly coherent, but it patently is not. What is missing from the preceding paragraph is any kind of causal structure. Although adjacent sentences refer to the same concepts, there is no other reason why one sentence follows the other. You could scramble the order of the sentences at random and the result would be nearly as coherent as the original.

Some researchers have therefore proposed that a critical element of discourse coherence and discourse understanding is the discovery of the causal structure of the story (Fletcher, 1986; Fletcher & Bloom, 1988; Fletcher et al., 1990; Trabasso & van den Broek, 1985; Trabasso et al., 1989; van den Broek & Trabasso, 1986). Rather than looking for overlapping arguments, readers analyze the statements in narrative according to two criteria:

1. Is this statement a cause of events that occur later in the story?
2. Is this statement a consequence of events that occurred earlier in the story?

To determine whether something is a cause, comprehenders apply the necessity in the circumstances heuristic (which is based on the causal analysis of the philosopher Hegel). The necessity in the circumstances heuristic says that “A causes B, if, in the circumstances of the story, B would not have occurred if A had not occurred, and if A is sufficient for B to occur.” Let’s look at an example (from Keenan, Baillet, & Brown, 1984):

*Timmy’s brother punched him again and again. The next day Timmy’s body was covered in bruises.*

Event A (Timmy’s brother punching him) is considered a cause for event B (Timmy having a lot of bruises the next day), because in the context of this mini-drama, being punched is enough all by itself for bruises to happen (it is sufficient), and absent some other named cause, event A is necessary for event B to occur (take away punching and the bruises also go away).

When people read or listen to stories, they are not so much concerned that adjacent units of texts (sentences, paragraphs, or episodes) have common arguments (although causally connected elements often will have common arguments, for example “Timmy” in the preceding mini story). What is important to comprehenders is to figure out why events follow one another in a given sequence, and discovering a causal relation answers the question of why event B followed event A in the story.

The importance of causal structure in the mental processing of texts can be demonstrated in a variety of ways. First, the propositional structure of texts can be described as a network of causal connections. Some of the propositions in a story will be on the central causal chain that runs from the first proposition in the story (*Once upon a time …*) to the last (*… and they lived happily ever after*). Other propositions will be on causal dead-ends or side-plots. In *Cinderella*, her wanting to go to the ball, the arrival of the fairy godmother, the loss of the
glass slipper, and the eventual marriage to the handsome prince, are all on the central causal chain. Many of the versions of the Cinderella story do not bother to say what happens to the evil stepmother and stepsisters after Cinderella gets married. Those events are off the central causal chain and, no matter how they are resolved, they do not affect the central causal chain. As a result, if non-central events are explicitly included in the story, they are not remembered as well as more causally central elements (Fletcher, 1986; Fletcher & Bloom, 1988; Fletcher et al., 1990).

Second, causal connections between propositions in stories also provide a basis for maintaining information in an active state between processing cycles. According to the leading edge strategy (Kintsch, 1988, 1994), comprehenders connect propositions to one another based on argument overlap. So, when they are choosing propositions to maintain in working memory, they choose the most recently encountered propositions. Other accounts appeal to the importance of causal connections in narrative (e.g., Fletcher, 1986). According to the current state strategy, the discourse processing system looks for recently encountered propositions that have antecedents but no consequences. Antecedents are events in the preceding story that caused the proposition in question to happen. So, in the Timmy mini-story, being punched is the antecedent of having bruises; and having bruises is the consequence of being punched. In the Timmy mini-story, Timmy’s body was covered in bruises is both the most recent proposition and the end of the causal chain. So, both the leading edge strategy and the current state strategy predict that this proposition will remain in working memory as a continuation of the Timmy story is being processed. In longer and more complex narratives, the most recently encountered proposition is not necessarily at the end of the causal chain. So, experimentally, it is possible to manipulate whether the final clause or sentence in a story meets the criteria of having antecedents but no consequences in the preceding text. According to the leading edge strategy, this will not matter, so long as a new proposition shares an argument with the most recently encountered proposition. According to the current state strategy, processing times should increase if the most recently encountered proposition does not provide a cause for an incoming, new proposition. When extended narrative texts are analyzed for their causal structure, and when causal structure and argument overlap are used to predict the amount of time it takes people to process and interpret different portions of the text, the current state strategy does a better job of predicting how long it takes people to read different parts of the text. In other words, causal structure, rather than argument overlap, appears to exert a bigger influence on how long it takes people to integrate different bits of narrative texts.

Further evidence against the leading edge strategy, according to which comprehenders link sentences based on associative links, comes from a reading time and memory study conducted by Wolfe and his colleagues (Wolfe, Magliano, & Larsen, 2005). They showed that that semantic (associative) relationships between sentences influenced processing time and memory only when the sentences did not share a clear causal relationship. The current state strategy also does a better job than the leading edge strategy in predicting how people will choose to continue a story (van den Broek, Linzie, Fletcher, & Marsolek, 2000). When given a choice between a more distant proposition that is at the end of the causal chain, and a more recent proposition that is not, speakers base story continuations on the more distant, but more causally central proposition.

The causal structure of a narrative affects how long it takes people to process incoming text, and this is probably because causal structure influences how easy it is to integrate adjacent elements of text. This factor influences the amount of time it takes to process incoming text, but it also affects the likelihood that comprehenders will remember how two pieces of text go together. This has been demonstrated in experiments that systematically varied how causally related adjacent elements of texts were. For example, consider the Timmy mini-story again (Keenan, et al., 1984; see also Duffy, Shinjo, & Myers, 1990; Myers, Shinjo, & Duffy, 1987).
Timmy's brother punched him again and again. The next day his body was covered in bruises. (High coherence)

These two sentences are highly coherent because they have a strong causal relationship. Timmy getting punched caused him to have bruises. It is possible to manipulate this relationship by replacing the first proposition. For instance, you could have:

Timmy was riding his bicycle. The next day his body was covered in bruises. (Medium coherence)

These two events do not have the same tight causal relationship as the preceding version, but it is not too hard to connect them together. Riding a bicycle in and of itself does not cause people to have bruises. So the two sentences can not be integrated as easily as the version where the first sentence provides the cause for the second (as defined by Hegel, see above). Most people when they read this second version infer a plausible bridging event, such as Timmy fell off his bicycle. Falling off one's bicycle is a reasonable consequence of riding a bicycle, and it provides a reasonable cause for having bruises. So, while it takes longer to read this second mini-story (because you need to take time to make the bridging inference), comprehenders can build an integrated representation of the two explicitly mentioned events by inferring the presence of an unmentioned third event. Now, consider this third version of the Timmy story.

Timmy went to his neighbor's house to play. The next day his body was covered in bruises. (Low coherence)

In this version, it is more difficult to figure out how the two events go together, and there is no clear path to a bridging inference. There are many events that could connect the two sentences, and none of them seems very highly likely or very extremely plausible. Were they playing tackle football? Did he get in a fight with his neighbor? Were they practicing tae kwan do with no pads? Did they get in a car accident? In the low coherence version, it will probably take comprehenders a long time to come up with a plausible way to relate the two events, they are likely to have lower confidence in the accuracy of their inferences, and they may fail to bridge the gap.

In fact, when participants are asked to read the second sentence in each pair (The next day his body was covered in bruises), their reading times are a straight line function of causal relatedness. The target sentence is read fastest in the highly coherent version, slower in the medium coherence version, and slowest of all in the low coherence version (Keenan et al., 1984). These differences in reading time likely reflect inferencing processes. No inference is needed in the highly coherent version, an obvious inference is available in the medium coherence version, and it takes a long time to come up with one of many non-obvious inferences in the low coherence version (and some additional time might be necessary to choose which inference to include in the integrated discourse representation). While coherence has a straightforward effect on processing time, it has a less straightforward effect on memory. If the first sentence is given to participants as a retrieval cue, and their task is to remember the target sentence (The next day his body was covered in bruises), the best recall is not in the highly coherent version, it is actually in the medium coherence version of the Timmy story. Across the three conditions, high coherence, medium coherence, low coherence, memory performance creates an upside-down-U-shaped pattern. Both the high and low coherence versions of the story lead to lower levels of recall than the medium coherence version.

There are likely two factors at work that determine recall performance. First, comprehenders may fail to build an integrated discourse representation at all in the low
coherence version. That is, they fail to discover a connection between the two events during on-line interpretive processing, and so, if they store both events in long-term memory, they store them as independent, unrelated events. So, activating one event at test does not lead to any increased activation of the other event at test, because there is no connection between them in long-term memory.

The difference in recall performance between the high and medium coherence condition is most likely a function of depth of processing. Depth of processing refers to the fact that the more mental effort we put into processing a stimulus, the stronger our mental representation and memory of that stimulus will be (Craik & Tulving, 1975). Ironically, because the highly coherent pair of sentences does not require much effort to encode, it does not leave much of an impression in memory. By contrast, the moderately coherent pair requires a bit more effort in the form of a bridging inference, but the extra time and effort taken to connect the two sentences and the fact that we engage in a more active process to make the connection both contribute to a stronger integrated representation, and this leads to superior memory performance. In fact, it is possible to have people do more active processing on the highly coherent pairs, by asking them to mentally elaborate on the explicitly stated information. In Susan Duffy and colleagues’ (1990) experiment, they made sure that some of their research subjects would elaborate on all of the sentence pairs by instructing them to write down a sentence that could “come between” the two critical sentences (Duffy et al., 1990, p. 30). The other subjects just read the sentences, as in the previous studies. Participants who read and elaborated on pairs of sentences recalled all of the target sentences equally well, whether the original sentence pair had been in the high, moderate, or low coherence condition. Further, in the elaboration condition, recall was just as good after 24 hours as it was if subjects were tested immediately after they read the sentences. This study shows that the way comprehenders approach the text has a big effect on how much they can remember later on. If they use a “deeper” encoding strategy, such as elaboration, memory is much stronger and lasts much longer, and this can overcome features of the text that make it either too coherent or too incoherent to leave a strong memory trace.

The Role of General World Knowledge in Discourse Processing

All current models of discourse processing and interpretation place great emphasis on the role that general world knowledge plays in the construction of coherent representations of texts. What form does this knowledge take and how is it applied as texts are being processed? General world knowledge can affect discourse processing in at least three distinct ways. First, we have knowledge about the way stories are typically structured, and these expectations affect the way we process and remember texts. Second, general world knowledge provides the information we need to make inferences that keep our representations of texts coherent. Third, general world knowledge affects the form and content of the situation models we build to represent what texts are about. Let’s consider each of these ideas in turn.

First, because we are storytelling monkeys, we have had a great deal of exposure to narrative. But because different cultures put together narratives in different ways, our expectations for what kinds of events will occur in narratives, and how those events will be expressed, is heavily dependent on the kind of culture we are raised in. For example, Sir Frederic Bartlett (1932/1995) is still widely quoted in psycholinguistics because he
showed that memory for texts depends largely upon the expectations and knowledge that comprehenders bring to the task as they are listening to or reading stories. In his landmark study, Bartlett had people read the story below, "War of the Ghosts," and then engage in a variety of re-production tasks.

One night two young men from Egulac went down to the river to hunt seals and while they were there it became foggy and calm. Then they heard war-cries, and they thought: "Maybe this is a war-party." They escaped to the shore, and hid behind a log. Now canoes came up, and they heard the noise of paddles, and saw one canoe coming up to them. There were five men in the canoe, and they said:

"What do you think? We wish to take you along. We are going up the river to make war on the people."

One of the young men said, "I have no arrows."

"Arrows are in the canoe," they said.

"I will not go along. I might be killed. My relatives do not know where I have gone. But you," he said, turning to the other, "may go with them."

So one of the young men went, but the other returned home.

And the warriors went on up the river to a town on the other side of Kalama. The people came down to the water and they began to fight, and many were killed. But presently the young man heard one of the warriors say, "Quick, let us go home: that Indian has been hit." Now he thought: "Oh, they are ghosts." He did not feel sick, but they said he had been shot.

So the canoes went back to Egulac and the young man went ashore to his house and made a fire. And he told everybody and said: "Behold I accompanied the ghosts, and we went to fight. Many of our fellows were killed, and many of those who attacked us were killed. They said I was hit, and I did not feel sick."

He told it all, and then he became quiet. When the sun rose he fell down. Something black came out of his mouth. His face became contorted. The people jumped up and cried.

He was dead.

(Bartlett, 1932/1995)

One task involved a group of people each telling the story from memory to one other person in the group, who would then tell the story to a third person, and so on. The point of the experiment was to find out how close the final version of the story was to the original. Bartlett found that his subjects consistently changed the story as they retold it, in particular, "by adding statements about the characters’ thoughts, motivations, intentions, and feelings" (Johnston, 2001, p. 355). Violating people’s expectations about story structure influences their ability to remember the verbatim form of the story and elements of meaning, but it also affects how difficult texts are to process and interpret. When story constituents are moved away from the normal locations in stories, reading times slow down, both at the place where the moved constituent is taken from and the new place where it is inserted (Mandler & Goodman, 1982).

Mandler and Johnson (1977) developed an artificial intelligence approach to representing the contents of stories. They reasoned that stories have an internal structure the same way that phrases and sentences have an internal structure. When comprehenders encounter a new story, they use their knowledge of the typical story components and their relations to encode information from the new story. In Western narrative, we expect stories to begin with descriptions of characters and settings; and we expect to encounter a series of episodes that plays out in a temporal order roughly from oldest to most recent,
and we expect the episodes to be coherent by virtue of having a causal structure—we expect later episodes to be directly or indirectly caused by the events taking place earlier in the story. As a result, people’s memories for stories will be biased in the direction of the typical story structure. When stories actually conform to the typical style, this bias does not lead to substantial differences between the actual story and people’s memory for the story. But when the story differs substantially from the normal form (in that culture), as the “War of the Ghosts” tale does, then the way people remember the story will be much different than what the actual text of the story would dictate. This is because, when people read stories such as the “War of the Ghosts”, which do not conform to the usual narrative form, they impose the typical narrative structure on the text itself, and so they find causal connections, and they infer particular motives for characters in the story, where none are explicitly stated.

Story grammars represent one kind of schema—a structured, pre-existing package of knowledge related to a particular domain (story form, in this case)—but other kinds of schemas also play an important role in discourse comprehension. Stories frequently refer to common experiences, and speakers and writers depend upon comprehenders having knowledge about these common experiences, so that the comprehenders can “fill in the gaps” between pieces of information that are explicitly stated in the text itself. These knowledge schemas, thus, play an important role in inference generation (Schank, 1972; Schank & Abelson, 1977). For example, if we are reading an episode about going to a restaurant, the author is not likely to mention typically occurring objects or characters, like waiters, cooks, tables, and silverware, unless these objects are critical to moving the plot forward or developing the characters. In fact, there is a fair amount of consistency between different individuals in the content and organization of their verbal reports about common events, like going to the doctor or eating at a restaurant (Bower, Black, & Turner, 1979). As a result, if you read

Susan and Bill went into the restaurant, sat down, and ordered lunch.

This would make sense, because your schema for going to a restaurant contains the information that restaurants have chairs (to sit on) and that you can get food there, including lunch. Further, you would experience very little difficulty integrating the following

Susan dropped her fork.

even though fork was not explicitly mentioned in the context. This is because your pre-existing knowledge schema for the typical objects and events in restaurants includes the information that when people sit down at a restaurant, they sit down at a table, and the table has silverware on it. Schemas also include knowledge about typical events that take place in a given context. And so we would not be surprised to read

They paid the check and left.

even if the story did not explicitly mention that Susan and Bill finished eating and their waiter brought them a bill. If the specific restaurant in the story violated an aspect of the comprehender’s restaurant schema, that would likely be mentioned by the writer (e.g., Oddly, there was no silverware on the table). In a similar way, schematic knowledge helps us appreciate when something interesting or unusual happens in the story. So, comprehenders would slow down and pay more attention if our restaurant story continued:

When Susan bent down to pick up her fork, she saw a bomb ticking away under the next table.
If people did not deploy schematic knowledge while comprehending stories, writers would have to spend pages and pages filling in basic facts, storytelling would be much less efficient, and it would be more difficult to highlight the unusual (which is really the point of telling stories—what speakers and writers really want to do is convey new information).

Although world knowledge is normally described in a propositional or fact-based way, comprehenders bring more than that to the table when they are understanding discourse. In particular, people have moral and ethical beliefs that are somewhat independent of their storehouse of factual knowledge (although the facts we learn certainly do influence our moral and ethical judgments). It turns out that our moral and ethical beliefs do affect the way we interpret texts, and those effects can be observed even in the brain wave activity that occurs when people read or listen to texts (van Berkum, in press; van Berkum, Holleman, Nieuwland, Otten, & Murre, 2008).

**Building Situation Models**

Some accounts of discourse processing adopt the idea that texts can be treated as sets of instructions. The instructions tell the language processing system how to build a situation model. So, to understand how texts are processed, we need to understand what the instructions are, and how comprehenders use these instructions to build situation models. According to the *structure mapping and focus theory* (Sanford & Garrod, 1981, 1998, 2008), texts provide instructions that lead to the automatic activation of situation-specific background knowledge, and they lead comprehenders to focus their attention on specific parts of this background knowledge. One surprising aspect of this kind of account is that, if contextually inappropriate information is included in the verbatim form of the text, but that information is not in the focus of attention, readers will often fail to notice that there is anything strange about the texts. This phenomenon is frequently referred to as the Moses illusion, because of studies involving sentences like this (Barton & Sanford, 1993; Bredart & Modolo, 1988; Erickson & Mattson, 1981; Hannon & Daneman, 2001; van Oostendorp & de Mul, 1990):

> How many animals of each type did Moses take on the ark?

Of course, if you are familiar with the “great flood” story from the Bible, you know that it was Noah, and not Moses at all, who put the animals on the boat. But a high proportion of people who read questions like the preceding one fail to note the anomaly, and they go ahead and give the answer “two.”

Similar effects occur in longer narratives as well. Barton and Sanford (1993) presented large groups of people with paragraphs about a plane crash, like this one:

> There was a tourist flight travelling from Vienna to Barcelona. On the last leg of the journey, it developed engine trouble. Over the Pyrenees, the pilot started to lose control. The plane eventually crashed right on the border. Wreckage was equally strewn in France and Spain. The authorities were trying to decide where to bury the survivors. What is the solution to the problem?

Many of their research subjects wrote solutions like:

> They should be buried in their home countries.
Of course, it’s possible that participants who wrote answers like this really didn’t notice that the paragraph used the word *survivors* rather than the word *deceased* in the critical final sentence. However, some people wrote answers like this:

The *survivors should be buried* where their relatives wish.

These kinds of findings indicate that, when participants start reading a story about a plane crash, they activate situationally relevant information (consistent with schema theory; Schank, 1977; Schank & Abelson, 1977), and they use that activated background knowledge to assign reference to subsequently encountered pieces of the text. When they get to the word *survivors*, they map that word to the activated portion of the background knowledge that corresponds to the concept *deceased or victims*, even though the stored lexical meaning of *survivor* is opposite to that of the schematically supplied concept *deceased*. On other words, the situation model has overpowered the lexical level of representation (in fact, comprehenders may have simply bypassed the lexicon and mapped the word *survivor* directly to an already activated portion of their situation models).

The Moses illusion shows that the situation model has the power to override semantic information tied to individual words. But when does this process take place, and how do lexical and contextual information interact to produce meaning? There is considerable evidence that discourse information places immediate constraints on interpretation, such that the “normal” meaning of individual words is never activated if the text is sufficiently constraining (as the Moses illusion texts seem to be) (Camblin, Gordon, & Swaab, 2007; Hess, Foss, & Carroll, 1995; Ledoux, Camblin, Swaab, & Gordon, 2006; van Berkum, Zwis overhead, Hagoort, & Brown, 2003). Normally you would read the sentence *The peanut was in love* much slower than the sentence *The peanut was salted*, and your brain wave activity would show a bigger N400 effect in response to *love* than to *salted*. However, if the discourse context introduces a kind of cartoon scenario where the peanut is portrayed as an animate, sentient being, the N400 effects are reversed. So, the discourse context has overridden the normal features associated with peanuts and replaced them with situation-specific features (Nieuwland & van Berkum, 2006).

Of course, some of the people in the Moses illusion studies do notice that things like burying survivors would be strange, and subsequent studies have shown that at least two factors contribute to the likelihood that an individual comprehender will experience the Moses illusion. First, if the anomalous word shares aspects of meaning with the intended word, the likelihood of experiencing a Moses illusion increases. For example, Moses and Noah are pretty close in meaning in many people’s understanding of the terms—they are both older, male, bearded, serious Old Testament characters. When more distinctive characters are introduced into the scenario—Adam, for example—the strength of the Moses illusion is greatly reduced (van Oostendorp & de Mul, 1990). In terms of online processing, eye-tracking data show that people notice distinctively anomalous intruders—such as Adam in the Noah scenario—but less distinctively anomalous intruders such as Moses lead to no initial difficulty in the situation model building process. Later on, processing slows down a little. This slow-down probably reflects the mental processes used to assign a new, extended meaning to the intruding word (Stewart, Pickering, & Sturt, 2004).

Another way to reduce the Moses illusion and to make it more likely that comprehenders will detect the anomaly is to use linguistic cues to focus attention on the intruding item. Syntactic structures such as *clefts* (like 16) and *there-insertions* (like 17) offer ways to do this.

(16) It was Moses who took two of each kind of animal on the Ark.
(17) There was a guy called Moses who took two of each kind of animal on the Ark.
When attention is focused on Moses using these kinds of grammatical cues, subjects are more likely to notice that he does not fit in with the great flood scenario, and they are less likely to experience the Moses illusion.

Focusing constructions have effects on other aspects of situation model construction, as they help to regulate the degree of activation that is assigned to different parts of the situation model. Structure mapping and focus theory adopts a view of situation model construction under which characters in stories are represented by tokens (mental placeholders in the model), and other information in the stories is then mapped onto activated tokens. If a token has been focused, it is easier to map information onto that token. So, for example, it will be easier to map new information onto the token for mayor in sentence (18) than in sentence (19) (from Birch & Garnsey, 1995, p. 289; Birch, Albrecht, & Myers, 2000):

(18) It was the mayor who refused to answer a reporter’s question.
(19) The mayor refused to answer a reporter’s question.

Because the it-cleft is a more marked structure (it deviates from the norm, is less frequent, and has explicit cues that differentiate it from the norm), the concept mayor is more accessible after (18) than after (19), even though mayor is the first-encountered character in both cases, and is in the prominent subject syntactic position in (19) and the less prominent object position in the main clause in (18). This difference in the information structure of the two sentences leads to differences in recognition and recall. Recall and recognition are better when a character is focused (mayor in (18)), than when it is not (mayor in (19) (Cutler & Fodor, 1979; Singer, 1976). People also spend more time looking at focused than unfocused parts of texts (Zimmer & Engelkamp, 1981), suggesting that they are making special efforts to encode focused concepts (which is also consistent with the structure-building process of laying a foundation). In recognition probe experiments, people respond faster to target words when those target words refer to focused tokens than unfocused tokens (Birch & Garnsey, 1995; Birch et al., 2000). All of these findings suggest that focused parts of text enjoy higher than normal levels of activation in the comprehender’s situation model.

Languages provide other ways besides syntactic position to signal comprehenders to boost the activation of specific parts of their situation models. English has cues that signal comprehenders that particular concepts will be referred to in the future. These cues are called cataphors, and elements in a discourse that are cataphorically marked receive higher degrees of activation and are more resistant to having their activations reduced by other elements of the comprehender’s situation model (Gernsbacher & Jescheniak, 1995; Jescheniak, 2000). For example, speakers can use loudness—spoken stress—to mark parts of the discourse that should be kept more available than normal. When the loudness of nouns in a spoken discourse was manipulated (as in the verbal equivalent of Susan needed to buy an ASHTRAY versus Susan needed to buy an ashtray), people respond to the probe word ashtray faster after they hear a sentence where ashtray is louder than when it is spoken at a lower volume. People also mark elements of a discourse that they will continue to talk about using the indefinite article this (even children do this; Wright & Givón, 1987). The indefinite this has a similar effect to spoken stress on the activation of concepts in a comprehender’s situation model. When a concept is introduced with the indefinite this (as in Susan needed to buy this ashtray versus Susan needed to buy an ashtray), people respond to the probe word ashtray faster than if the concept were introduced with the more commonly encountered indefinite article a/an (Gernsbacher & Shroyer, 1989). Further, when concepts are introduced with these cataphoric devices, they resist being deactivated or suppressed when other concepts are introduced into the discourse. Normally, if speakers introduce a new topic, previously encountered information becomes less active or accessible. So, if Susan saw an ashtray was followed by and then she found an end table, normally the accessibility of
ashtray would decrease after end table was introduced into the discourse. So, reaction time to the probe word ashtray generally increases after comprehenders process the words end table. However, when ashtray is marked by spoken stress or the indefinite this, its activation level remains high even after comprehenders encountered end table, and so reaction times to the probe word ashtray remain fast.

Inferencing: Memory-Based Account of Discourse Processing: Minimalist vs. Constructionist Inferencing

Memory-based text processing accounts appeal to general memory processes to predict and explain how comprehenders will react to texts as they read them and how the texts will be remembered over the long term. General working memory functions lead to primacy and recency effects (Deese & Kaufman, 1957). Almost totally regardless of the kinds of stimuli people are exposed to, they remember stimuli at or near the beginning of the group better than later-occurring information (the primacy effect); and they remember stimuli at or near the end of the group better than information in the middle (the recency effect). Similar effects occur in discourse processing and memory. People remember the first character in a text better than characters that come later (which could reflect a kind of primacy effect); and, immediately after they finish processing a piece of text, recently encountered parts of the text are remembered better than earlier occurring parts of the text (a kind of recency effect) (e.g., Gernsbacher et al., 1989).

According to the memory-based approach, texts activate information from long-term memory by a process of resonance (as in Doug Hintzman’s Minerva computational model of memory processing; Hintzman, 2001). Resonance activates information from long-term memory depending on how closely related the information conveyed by the text matches or is associated with the information stored in long-term memory. As Gerrig and McKoon (1998, p. 69) indicate, “the degree to which specific information in memory will be evoked depends on the strength of the association between the cue in short-term memory and the information in long-term memory.” Further, knowledge activation occurs automatically (we can’t control it) and the activation process is very dumb. It does not select information based on relevance, or interest, it just activates whatever information has an association of any kind to the bit of text that is being processed at the moment. As Kintsch notes (Kintsch, et al., 1990, p. 136),

Comprehension is simulated as a production system, the rules of which operate at various levels: some build propositions from the linguistic information provided by the text; some generate macropropositions; some retrieve knowledge from the comprehender’s long-term memory that is related to the text, thus serving as mechanisms for elaboration and inference. All these rules share one general characteristic: they are weak, “dumb” rules that do not always achieve the desired results. In addition to what should have been constructed, these rules generate redundant, useless, and even contradictory material.

The memory-based approach to discourse processing contrasts with other more “top-down” approaches, which assume that a reader’s goals and “search for meaning” play a more active role in the construction of mental representations from texts (e.g., Singer et al., 1994).

One of the potential benefits of a fast, dumb knowledge-activation process is that when concepts, characters, or objects are explicitly mentioned in texts, they will already have had
their activation and accessibility boosted by virtue of their associations with previously processed parts of the text (Gerrig & McKoon, 1998). This notion of readiness suggests that the fast, automatic activation of associated knowledge will speed discourse comprehension by simplifying the task of figuring out what the explicitly mentioned information refers to. Another benefit of memory-based text processing is that using a fast, dumb, associative mechanism limits the number of inferences that the reader will draw from any given text. Given the huge store of knowledge that comprehenders bring to the text interpretation process, an unconstrained inference process would result in massive numbers of inferences, most of which would not be relevant to either the author's intent or the comprehender's goals in reading the text. By limiting activation to only the information in long-term memory most closely associated with the set of propositions active in working memory, the information overload that unlimited inference entails can be avoided. An additional claim is that information explicitly provided by texts can activate associated information from long-term memory, even if the text does not directly refer to the associated information (McKoon, Gerrig, & Greene, 1996; see also Deese, 1959; Roediger & McDermott, 2000).

In the memory-based approach, information becomes activated when it is associated with the current state of the working memory system, and information decays or has its activation reduced unless it is refreshed by new information conveyed by the text. Information explicitly conveyed by a text can activate information from long-term memory because the explicitly conveyed information has a pre-existing association to information in long-term memory (i.e., words have standard meanings that can be accessed from the lexicon during discourse processing) and the text itself can set up new associations that are then stored as part of the long-term memory representation of the discourse. Evidence for this latter type of association comes from experiments on anaphoric reference. If Susan and Jane discuss a third character, Ted, an episode that brings Susan and Jane back together will increase Ted's activation level and make him more accessible (McKoon, et al., 1996; Gerrig & McKoon, 1998). As a result, subjects in experiments will respond to the word Ted faster and will be more capable of figuring out the connection between a pronoun and Ted. This reactivation of the third character happens even when the text only alludes to an episode involving him or her. The text does not have to explicitly mention the third character. Thus, reading this story has created a new set of associations in long-term memory between the characters that interacted, even indirectly, in the story.

One of the areas where memory-based text processing and other approaches diverge is in the area of inference generation. Specifically, accounts differ as to which kinds of inferences are drawn naturally by comprehenders as they are in the act of interpreting texts. The memory-based position says that very few inferences are drawn during the actual process of interpretation, and those inferences that are drawn are constructed by automated mental processes. This notion of minimal inference says that inferences will be drawn under only two limited conditions. First, inferences will be drawn if they are necessary to establish cohesion between adjacent parts of the text (two sentences, say). Second, inferences will be drawn if the inferences are based on “quickly and easily available” information (McKoon & Ratcliff, 1992, p. 441). So, in cases like those in the Keenan et al. “Timmy” experiments, causal inferences will be drawn because they are necessary for the information in the two sentences to be integrated into a coherent whole. But other kinds of inferences may not be drawn (see the box on p. 222 for some of the more common types of inferences). For example, if you read The delicate vase fell off the high shelf, and the text does not explicitly state what happened to the vase, you might infer that the vase broke. Similarly, if you read the sentence The woman stirred her coffee, you might infer that she used a spoon (rather than a fork or her fingers). According to the minimal inference hypothesis, none of these inferences are drawn, and so the associated information, broke for the causal inference and spoon for the instrument inference, would not become activated when you read The delicate
vase fell off the high shelf and The woman stirred her coffee. You would make the inference Timmy fell in the bridging inference case, because that event (or one very much like it) is needed to tie the two sentences together.

A large number of experiments have tried to test exactly when people make different kinds of inferences when they read or listen to stories (see Zwaan, 2006, for a review). The general consensus is that bridging inferences are routinely drawn, elaborative inferences are rarely drawn, and causal and instrument inferences are drawn under very limited circumstances. For example, instrument inferences are drawn quickly if the instrument has been explicitly introduced previously in the discourse and the context selects very strongly for one particular instrument. Most people stir coffee with a spoon almost all of the time they stir coffee, so that context is highly constraining. Consistent with the minimal inference hypothesis, comprehenders appear to avoid drawing some inferences that they might reasonably draw.

This does not mean that the minimal inference hypothesis makes the correct prediction all of the time, however. For example, according to the minimal inference position, inferences will be drawn only if the information needed to draw the inference is readily available or the information is necessary to establish cohesion between adjacent elements in the text (that is to establish local coherence). However, some evidence suggests that comprehenders make inferences when the text is locally coherent—each adjacent element has a clear relation to the preceding and following elements—and the information needed for the inference is distant in the surface form, and so should not be quickly and easily available (Long, Golding, & Graesser, 1992; Singer et al., 1992; Singer et al., 1994).

Take a moment to read the following story (Goal inference story from Singer, 1993, in Singer et al., 1994, p. 432):

Valerie left early for the birthday party. She checked the contents of her purse. She backed out of the driveway. She headed north on the freeway. She exited at Antelope drive. She spent an hour shopping at the mall.

[The control condition starts with the sentence: Valerie left the birthday party early.]

Because the story refers to Valerie in each sentence, the story is locally coherent throughout—each sentence can be related to the preceding and following sentences because of argument
overlap. Under these conditions, no inference is necessary to establish coherence. Nonetheless, when people read the final sentence *Valerie spent an hour shopping at the mall*, they very quickly verify the statement *Birthday parties involve presents* (and they do so faster than in the control condition, which makes the birthday party irrelevant to the rest of the story). Thus, it appears as though subjects in this study inferred Valerie’s motive for shopping (i.e., she was looking for a birthday present). This outcome is problematic for the minimal inference position in two ways. First, no inferences should be drawn when the text is locally coherent. Second, by the time people reach the critical final sentence, the information necessary to make the goal inference (that there’s a birthday party) should have been long gone from working memory.

There also appear to be individual differences in the extent to which comprehenders incorporate inferred information into their discourse representations, although knowledge activation processes appear to be fairly uniform across groups of better and poorer comprehenders (Long & Chong, 2001; Long, Oppy, & Seely, 1994; 1997). If minimal inference is taken as a universal inference-generation mechanism, it does not explain why some people make inferences while others do not. Because the empirical record provides partial support for both minimalist and constructionist positions, some authors advocate hybrid accounts that factor in both passive, dumb knowledge activation processes and more strategic, top-down inference generation processes (e.g., Long & Lea, 2005; van Den Broek, Rapp, & Kendeou, 2005).

The Neural Basis of Discourse Comprehension

Although scientific investigation of discourse processing is still in its early stages, considerable progress has been made in the last decade in understanding how the brain responds to connected discourse (see Ferstl, 2007; Ferstl et al., 2008, for reviews). In particular, the advent of brain imaging techniques like PET and fMRI has allowed researchers to investigate the links between brain activity and text properties in new ways, and this has led to new insights about how the brain is organized to process discourse.

Language scientists have known for a long time that more coherent text produces different brain wave activity than less coherent texts (Kutas, van Petten, & Kluender, 2006). For example, the initial words of sentences processed in isolation produce larger negative voltage at the scalp (reflected by the N400 ERP wave form) than the same words appearing as part of a connected narrative (Van Petten, 1995). Whether a word makes sense in the context provided by preceding text also modulates the size of the N400 wave form (van Berkum et al., 2003). Violations of background knowledge, whether acquired in the experimental session or brought in via general experience, are also reflected in brain wave activity as indexed by the N400 (Fischler, Childers, Achariyapaopan, & Perry, 1985; Hagoort, Hald, Bastaiaansen, & Peterson, 2004; Van Berkum, Hagoort, & Brown, 1999). Notably, these N400 effects really are caused by the fit between the currently focused text and the linguistic and general knowledge context, and not by low-level word-to-word associations (Otten & Van Berkum, 2007). However, other ERP components may also be sensitive to changes in discourse coherence or plausibility. Using ERP methods, Petra Burkhardt tested pairs of sentences that can be easily mapped together on the basis of lexical (word–word) associations (*Yesterday a Ph.D. student was shot downtown. The press reported that the pistol was probably from army stocks*; Burkhardt, 2007, p. 1852) versus sentences that are more difficult to connect (*Yesterday a Ph.D. student was killed … the pistol…; Yesterday a Ph.D. student was*...
found dead ... the pistol...). Burkhardt’s study showed that the P600 component of the ERP signal increased as the difficulty of the coherence relationship increased.\textsuperscript{13}

Some of the early work in brain imaging of discourse processing tried to find out how textual cohesion and coherence affects the brain’s response to written text. One way to make a text more or less cohesive and coherent is to manipulate the kinds of articles that appear in the text. Generally, the\textit{ indefinite article} “a”/“an” introduces a new topic into the discourse, and so it indicates that there is a break between the new piece of text and what has come before. By contrast, the\textit{ definite article} “the” indicates that the following noun has already been introduced in the discourse context. As such, the definite article tells the comprehender that the new information is closely related to the preceding text, and, rather than preparing for a new concept, the comprehender should search the discourse representation and map the following noun onto a previously introduced referent. Thus, by manipulating whether an article is indefinite or definite, the author can make a text seem more or less cohesive or coherent (see the examples in Table 5.1). Notice how the right-hand definite article version seems more coherent. Part of this is because the definite article makes it easier to connect new sentences to previous ones (e.g., it’s easier to map \textit{the grandchild} and \textit{the little boy} together than \textit{a grandchild} and \textit{a little boy}).

David Robertson and his colleagues were interested in finding out whether the brain responds differently to less coherent and more coherent text, so they manipulated the presence of definite articles (Robertson et al., 2000). So, they had people read sets of sentences with and without definite articles and used fMRI to assess which parts of the brain responded more strongly to the different kinds of stimuli.

Figure 5.7 shows that, compared to the definite article condition, the indefinite article condition produced more brain activity in the right hemisphere, but not in the left hemisphere. The black bars show that there were greater increases in blood flow to right-hemisphere regions during processing of the incoherent texts than during processing of the coherent texts. While the effect was in the same direction in the left hemisphere (the white bars), the difference between the two conditions was not statistically significant. Previously, the left hemisphere had been considered the dominant or sole contributor to processing the meaning of language input. This study was one of the first to indicate that right-hemisphere regions play a role in establishing coherence. In particular, the right hemisphere is far more activated when texts lack the cues that normally help comprehenders figure out how different parts of the text go together.\textsuperscript{14}

Other brain imaging experiments have attempted to determine what role different parts of the brain play in establishing causal coherence in texts. Some of these studies have

<table>
<thead>
<tr>
<th>Indefinite article condition</th>
<th>Definite article condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A grandmother sat at a table.</td>
<td>The grandmother sat at the table.</td>
</tr>
<tr>
<td>A child played in a backyard.</td>
<td>The child played in the backyard.</td>
</tr>
<tr>
<td>A mother talked on a telephone.</td>
<td>The mother talked on the telephone.</td>
</tr>
<tr>
<td>A husband drove a tractor.</td>
<td>The husband drove the tractor.</td>
</tr>
<tr>
<td>A grandchild walked up to a door.</td>
<td>The grandchild walked up to the door.</td>
</tr>
<tr>
<td>A little boy pouted and acted bored.</td>
<td>The little boy pouted and acted bored.</td>
</tr>
</tbody>
</table>
capitalized on a coherence manipulation that produced robust reaction time and memory effects in behavioral studies (Duffy et al., 1990; Keenan et al., 1984; Myers et al., 1987). Rob Mason and Marcel Just performed an fMRI study on sentence pairs like those in the “Timmy” experiments. To analyze their fMRI data, Mason and Just divided the brain into large regions (as shown in Figure 5.8; only the left hemisphere is shown, but Mason and Just analyzed data from both hemispheres). They divided each of these large regions into a set of small cube-shaped regions called voxels. Subjects had their brains scanned as they read highly related (“Timmy’s brother punched him … his body was covered in bruises.”), moderately related (“Timmy rode his bike … bruises”), or distantly related pairs of sentences (“Timmy went to the neighbors’ … bruises”). To see how the brain responded in each of these conditions, Mason and Just found voxels that had more blood flow in response to people reading the sentences as compared to when people just looked at a fixation cross on the computer screen. Figure 5.9 shows the outcome of this analysis.
To estimate how hard different parts of the brain were working in the different conditions (low, medium, and high coherence), they counted up the number of voxels that were activated (had greater blood flow during the sentence processing task than during the resting look-at-a-fixation-cross task). The leftmost bars in Figure 5.9 show that, in the left hemisphere, many voxels were activated by the sentence reading task, but there were no differences between the low, medium, and high coherence conditions. The next set of bars shows the right hemisphere’s response to the sentence reading task. Here, the most voxels were activated by the medium-coherence sentence pair (*Timmy was riding his bike … The next day his body was covered in bruises*). Fewer voxels were activated by the low and high coherence pairs. The rightmost bars show that a frontal part of the brain (dorsolateral prefrontal cortex, of DLPFC) had greater activation in the medium and low coherence conditions than in the high coherence condition. This part of the brain is thought to be involved when working memory resources are brought to bear on an information processing task, and so activation differences here probably reflect that working memory is needed to make bridging inferences (although the attempt to make a bridging inference may fail in the low coherence condition). The main thing to take away from Figure 5.9 is that, while the right hemisphere does respond differently to different degrees of coherence, the left hemisphere apparently does not (at least, there is no indication in this experiment that the left hemisphere cares whether pairs of sentences are related or not). Thus, one might conclude that, while the left hemisphere figures out what sentences mean, the right hemisphere is responsible for establishing coherence between sentences, or perhaps that the right hemisphere is responsible for drawing inferences.

The conclusion that the right hemisphere is responsible for drawing inferences is consistent with some recent research on the phenomenon of insight. Insight experiences happen when people are working on a problem, and just before they work out the solution, they get the feeling that they know what the solution is (these experiences are sometimes described as “Aha!” moments; Bowden & Beeman, 2003). As panel B in Plate 9 shows, the
The neural basis of discourse comprehension

The left hemisphere does not respond differently when people experience an “Aha!” moment of insight and when they do not (Jung-Beeman et al., 2004; see also Beeman & Bowden, 2000; Kounios et al., 2006). But, as Panel C shows, the right hemisphere shows greater neural activity when people have an insight that leads them to solve a problem.

A similar pattern of processing may take place during the generation of inferences. To generate an inference, comprehenders take the information from two adjacent parts of the text and use their background knowledge to come up with a way to connect them. One way to accomplish this, especially when there is no direct overlap between the arguments in two adjacent pieces of text, is to rely on more distant semantic relationships, which would be more likely to be activated in the right hemisphere than the left (Beeman, 1993; Beeman et al., 1994).

Both behavioral and neuroimaging data support the idea that the right hemisphere plays a role in inference generation. In behavioral experiments, semantic priming for target words related to the solution of an insight problem gets bigger as subjects’ “Aha!” feelings increase (Bowden & Beeman, 2003). Also, predictive inferences appear to engage the right hemisphere more than the left (Beeman, Bowden, & Gernsbacher, 2000). Take, for instance, the sentence The space shuttle sat on the ground, waiting for the signal. This sentence might lead you to predict that there will be information forthcoming about the shuttle taking off. So, reading the sentence and making a predictive inference could lead to facilitated processing of the target word launch (Beeman et al., 2000, p. 311; Duffy, 1986). To see whether people draw this kind of inference, Mark Beeman and his colleagues presented target words like launch in the right visual hemifield. When this is done, the target word is processed first by the left hemisphere. The experiment showed that people processed launch no faster than an unrelated control word when the left hemisphere had the first shot at it. However, priming was observed when the word launch was presented in the left hemifield and was processed first by the right hemisphere. Similar methods have been used to test the maintenance of topic information (Faust, Barak, & Chiarello, 2006). When topic is manipulated by either having two sentences that refer to the same topic or referring to different topics, priming is observed for left visual field/right-hemisphere targets whether the target word is related to the first or second topic. However, priming in the RVF/left hemisphere is observed only when the target sentence is related to the most recently encountered meaning. These results suggest that the two hemispheres respond to topic information in different ways. The left hemisphere has more specific activation that delays more rapidly, and the right hemisphere has more diffuse activation that lasts longer (consistent with the coarse coding hypothesis, Beeman et al., 1994).

Neuroimaging experiments provide additional support. In one fMRI experiment investigating inference generation during the processing of fairly normal stories, greater activation was observed in the right hemisphere, in the right superior temporal lobe to be more exact, when the text supplied information that strongly implied a particular inference (Virtue, Haberman, Clancy, Parrish, & Beeman, 2006). Plate 10 shows areas of the right hemisphere that were more activated when the text implied, but did not explicitly state, a particular event. In that case, and given the degree of textual support for the inference, subjects were likely to infer the “missing” event. When processing of coherent stories is compared to processing of lists of unrelated sentences, substantial differences in activation are observed in the right hemisphere, which suggests that right-hemisphere regions play a role in establishing coherence (Vogeley et al., 2001). Finally, major differences in right-hemisphere activity happen when the ability to map text to a global theme is manipulated. Recall that in the Bransford and Johnson (1972) experiment, participants had a much easier time recalling story elements when elements of the story could be related to a specific scenario (like a young man serenading a young woman by using a set of floating speakers). Similar memory results are obtained when a title is used as the functional equivalent of the
picture in the Bransford and Johnson paradigm. In one such study using neural imaging (fMRI) as the dependent measure, participants read paragraphs like this (St. George, Kutas, Martinez, & Sereno, 1999, p. 1318):

This is very rewarding but tends to be quite expensive even if you own all that you need. The outfit does not really matter. One can get seriously injured without proper instruction even if it comes more naturally to some people than others. Some don’t like the smell or the lack of control …

This paragraph makes a lot more sense if you know that the title is “Horseback Riding,” because now referring expressions like outfit and smell can be tied to specific concepts and the entire paragraph can be related to a single, consistent theme. Plate 11 shows the results of an fMRI experiment. In the experiment, people read paragraphs like this one either without a title (where it would be hard to make sense) or with a title (where you could build a much more coherent discourse representation). Plate 12 is a little bit confusing, because the right side of the brain is shown on the left side of the graph. So, first find the right hemisphere. Then compare the amount of activation (the red part) when there is a title (the left-hand picture) versus when there is no title (the right-hand picture). Now look at the right-hand side of Plate 12, which shows the right-hemisphere response when there was a title (top) and when there was no title (bottom). Both figures show that the left-hemisphere activity is about the same whether there is a title or not, but the right hemisphere shows greater activity when the passage has no title and less activity when it does have a title. One way to explain these results is to suppose that the right hemisphere plays a special role in establishing textual coherence by mapping the different parts of the passage onto a central theme. When the passage lacks a title, this kind of processing is more difficult, and so the right hemisphere works harder.13

Although the right hemisphere does appear to play a role in inferencing and the establishment of coherence relations between different parts of texts, it would be a mistake to think that all inferencing takes place in the right hemisphere or that the right hemisphere is involved in every kind of inference. One of the earliest neuroimaging studies of auditory discourse processing showed that connected discourse led to greater activation in both the right and left temporal poles (the very frontmost part of the temporal lobes; Mazoyer et al., 1993). When topics are changed across sentence pairs (Do you believe in angels? Yes, I like to go to camp) versus kept the same (Do you believe in angels? Yes, I have my own special angel), both right- and left-hemisphere regions are activated, although relatively greater activity is observed in the right hemisphere, in both adults and children (Caplan & Dapretto, 2001; Dapretto, Lee, & Caplan, 2005). Additionally, data from patients with damage to the left prefrontal cortex show that they have difficulty establishing coherence relations between adjacent parts of texts, and they also had trouble drawing inferences that would help to maintain coherence (Zalla, Phipps, & Grafman, 2002; see also Ferstl, Guthke, & von Cramon, 2002).

The left hemisphere also appears to play a role in processing stories that involve character’s thoughts, beliefs, and emotions. Stories that require people to infer the mental responses of the characters involved in the story are called theory of mind stories. When people read such stories, they draw conclusions about characters’ thoughts and feelings (Gernsbacher et al., 1992; Gernsbacher et al., 1998). Processing of theory of mind stories has been compared to processing of stories that call for inferences about physical, but not mental, events. A theory of mind story might talk about character A getting character B fired (in which case, you might infer that character A responded by feeling guilty). A physical story might talk about a delicate vase falling from a high shelf (in which case, you might infer that the vase broke). When processing of theory of mind stories is compared to processing of physical stories, both of which require inferences, but different kinds of inferences, greater activation is seen in two parts of the left hemisphere (Fletcher et al.,
Summary and Conclusions

This chapter has reviewed three prominent and complementary accounts of discourse processing: construction integration, the structure building framework, and the event indexing models. Each of them makes a unique contribution to our understanding of how people process and interpret narratives. The goal of processing narratives is to build a mental model of the situation described in the real or imaginary world that the story is
about. To accomplish this, information that is associated with the explicitly stated contents of the story is combined with the comprehender’s general world knowledge. This process involves a high-fidelity, but short-lived representation of what the text actually says (the surface model), a more abstract representation that captures the propositions conveyed by the text (the text-base), and a long-lived situation model that incorporates inferences that the comprehender generates herself from the verbatim information in the text and her own store of world knowledge. Inferences, especially causal inferences, play an important role in filling in the gaps in stories when two adjacent elements of the text cannot be readily integrated. Finally, although neurophysiological (ERP) methods are well established in language science, sophisticated neuroimaging techniques have only recently been brought to bear to help us figure out how discourse interpretation processes are implemented in the brain. These newer imaging techniques have already revealed that right-hemisphere structures participate in discourse comprehension in ways that were unknown in the recent past. However, the available data do not support a clear division of labor between the two cerebral hemispheres. Discourse processing and interpretation rely on distributed networks of cooperating neural systems in both hemispheres.

**TEST YOURSELF**

1. Describe Kintsch’s construction–integration account of discourse processing. What kinds of representations are involved? How are they related to one another? How are they built or activated?

2. What are *propositions* and what do they contribute to discourse comprehension? What is the relationship between surface form and propositions? What is the relationship between propositions and situation models? What evidence supports the psychological reality of propositions?

3. What do comprehenders remember after they read a story? Describe an experiment showing that some representations are more durable than others.

4. Describe Gernsbacher’s structure building framework. In what ways does it resemble construction–integration? In what ways does it differ? Describe experiments that support the existence of mapping, shifting, enhancement, and suppression and explain how each process contributes to discourse comprehension.

5. What does Zwaan’s event indexing model say about discourse processing? What kinds of information do comprehenders put in their mental models? What kinds of evidence support claims made by the event indexing model?

6. What does the causal chain hypothesis say about discourse comprehension? What happens when two adjacent parts of a text do not have an obvious causal connection?

7. Describe different kinds of inferences. What role do inferences play in discourse comprehension? When and how do comprehenders draw inferences?

8. Describe the structure mapping and focus account. How does it relate to the Moses illusion?

9. Which parts of the brain participate in discourse processing? What functions do the left and right hemispheres undertake?