

## Plausibility and the comprehension of text

Alison Black, Paul Freeman and P. N. Johnson-Laird

---

Implausible versions of stories were constructed by a panel of judges who filled in sentence frames from the originals from which all but the referring phrases had been deleted. In one condition each judge completed a sentence frame independently, resulting in a highly implausible story; in another condition each judge saw the two previous completed sentence frames before filling in a sentence frame, resulting in a mildly implausible story. Both versions had the same referential structure as the original. Two experiments showed that the more implausible a text, the harder it was to understand and remember. The results confirm the importance of plausibility in the construction of a mental model of a text.

---

What makes discourse hard to understand? Obviously individual sentences may be difficult to comprehend because they contain recondite words, or have a complicated syntax, or express abstract and involuted ideas. But what else over and above the difficulty of individual sentences affects comprehension? The process of combining the propositions conveyed by separate assertions is a constructive one in which inferences provide information not explicitly stated in the text (e.g. Kintsch, 1974; Bransford & McCarrell, 1975) and such inferences are evidently easier to make given a relevant title for a text (e.g. Dooling & Lachman, 1971) or a relevant picture (e.g. Bransford & Johnson, 1972). These effects operate on a number of distinguishable factors, and we will examine two that have been identified as crucial to the comprehension of discourse: referential coherence and plausibility. Kintsch & van Dijk (1978) argued that referential coherence is the single most important factor for the cohesion of text; Schank (1975) claimed that causal coherence is the most important factor; and recently Keenan *et al.* (1984) have also argued that the emphasis on referential coherence is mistaken. One aim of this paper is to show that both sides in the dispute are wrong—or, equivalently, that both are, in part, right. There are, of course, other factors that may affect the comprehension of discourse, e.g. its rhetorical significance for the reader, and its abstract structure if certain story grammarians are to be believed (e.g. Rumelhart, 1975; Mandler & Johnson, 1977, 1980; Thorndyke, 1977). The trouble is that it is extremely difficult, if not impossible, to manipulate these aspects of a text without also affecting its referential coherence and plausibility.

A text is referentially coherent if each sentence refers to something that is referred to in at least one other sentence. Some theorists have treated co-reference as dependent on the repetition of noun phrases (e.g. Kintsch, 1974, p. 358; Kintsch & van Dijk, 1978), or as a matter of the substitution of one noun phrase for another (e.g. Halliday & Hasan, 1976). An alternative view, however, is that readers are attempting to construct a mental model of the discourse and that it is therefore a radical mistake to treat referential coherence as depending solely on the superficial form of texts: it depends on whether sentences make reference, either explicitly or implicitly, to entities in the mental model of the previous discourse (see Johnson-Laird, 1980, 1983; Johnson-Laird & Garnham, 1980). Referential coherence is readily grasped if a noun phrase is repeated in the discourse (Haviland & Clark, 1974; Kintsch *et al.*, 1975; Manelis & Yekovich, 1976), or if an inference linking two noun phrases can be made (Garrod & Sanford, 1977; Sanford & Garrod, 1981, chapter 5). When the sentences of a text are jumbled up in a random order, then a story becomes difficult to understand and to remember (Thorndyke, 1977). Scrambling the

order of sentences disrupts referential coherence, but if the noun phrases are modified so as to make clear what they are referring to, then there is a significant increase in comprehensibility and memorability (Garnham *et al.*, 1982). Likewise, if the order of three sentences describing a simple spatial layout is manipulated, then the crucial factor for both reading times and memory is whether or not the sequence allows each co-referential item (after the first sentence) to be linked to an entity already represented in the mental model of the discourse (Ehrlich & Johnson-Laird, 1982). Referential coherence can accordingly have an effect on comprehension independently from other factors.

Plausibility is a more problematical notion than referential coherence, but we can illustrate it by describing an experiment carried out by Johnson-Laird & Shapiro (see Wason & Johnson-Laird, 1972, p. 76 ff.). Consider the following two sentences:

It is raining. The girl is soaked.

When subjects had to describe a truth table for such assertions, they often included an explicit causal link in their description, whereas they included far fewer causal links with control pairs such as:

It is raining. The girl is prejudiced.

These results suggested that:

[Subjects] are always ready to leave the logical requirements of the task behind and to try to establish some meaningful connection between events. If the events cannot be directly related, the best that can be done is to establish some temporal connection between them. But where the events do seem to be related to one another, the subjects will spend a considerable amount of time attempting to derive a satisfactory causal model of them (Wason & Johnson-Laird, 1972, pp. 81–82).

Meaningful connections are clearly established by plausible inferences based on knowledge, and these inferences are extremely rapid, presumably because they do not call for any attempt to establish their validity. In fact, they are seldom valid but merely plausible, i.e. likely to yield true conclusions though they do not follow necessarily from the text.

Subsequent studies of plausibility have shown that causally related sentences are read faster and rated as more comprehensible than unrelated sentences (Haberlandt & Bingham, 1978; Keenan *et al.*, 1984), that the second sentence from causally related pairs is better remembered than from pairs that are merely temporally related (Black & Bern, 1981; but cf. Keenan *et al.*, 1984), and that subjects tend to make false alarms for explicit statements of causal relation in recognition tests (Thorndyke, 1976; Bower *et al.*, 1979) — a phenomenon obviously predicted by the fact that Johnson-Laird and Shapiro's subjects misdescribed truth tables in a causal way even when all the information was in front of them. Likewise, subjects remember the gist of texts describing determinate spatial relations very much better than texts that differ in only one word but that fail to describe a determinate relation (Mani & Johnson-Laird, 1982). Plausibility of certain sorts can evidently have an effect on comprehension independently from other factors.

The major problems with plausibility are theoretical and analytical. There are clearly many different sorts of plausible relations amongst sentences and, although various theorists have attempted to classify them (e.g. Halliday & Hasan, 1976; Schank & Abelson, 1977; van Dijk, 1977), there is as yet no comprehensive taxonomy. There is no existing classification procedure that automatically specifies all the relations that link the propositions of a text. The difficulty arises because plausibility depends on inferences based on knowledge, and there are many varieties of knowledge that may be used in understanding discourse. Furthermore, as Brown & Yule (1983, p. 195) have argued,

theorists often confuse the explicit expression of relations in a text with the underlying semantic relation that is expressed. We have already pointed out a similar confusion between referential coherence and a mere relation between the explicit referring phrases in a text. In fact, the crucial factor for both reference and plausibility is whether the appropriate relation can be established – by inference, if need be, not its explicit realization in the text.

Perhaps the severest problem in the analysis of plausibility concerns the subtleties of that most common relation between events – causation. Van Dijk (1977), who pioneered the analysis of textual relations, followed philosophers in distinguishing causes from reasons. According to his analysis, causation occurs where one event is sufficient for the occurrence of another (e.g. 'Because it did not rain, the soil dried out'), whereas a reason is a case where the knowledge of one event is sufficient for the occurrence of another (e.g. 'Because it did not rain, we irrigated our fields'). Schank and his colleagues (Schank & Abelson, 1977; Schank & Riesbeck, 1981) recognize a richer set of causal relations. They add to the basic set of simple physical causes and reasons three further notions: enabling, preventing and psychological causation in which someone is made aware of a change of state. To make matters still more complex, Miller & Johnson-Laird (1976, Section 6.3) argue that there are at least two varieties of each of the main causal relations (causing, allowing and preventing) depending on whether a sufficient or a necessary and sufficient relation is at stake.

Experiments manipulating causation have lumped all the different varieties into one, and have often conflated causes of events with reasons for actions. Their materials have invariably been devised according to the experimenters' own intuitions and validated by pilot studies or subjects' ratings (e.g. Haberlandt & Bingham, 1978; Black & Bern, 1981; Keenan *et al.*, 1984). It is therefore unclear to what extent the results are genuinely representative of the types of plausible relations that actually occur in texts.

Our goal was twofold. First, we wanted to advance the analysis of plausibility by attempting to discover more about the actual relations that are inferred from texts. Second, we wanted to develop a simple procedure for manipulating plausibility that did not depend on our intuitions and that would allow us to determine its effects when referential coherence was held as constant as possible. Our basic assumption was that when subjects are asked to make up stories, they use their knowledge to create texts that are plausible. If members of a panel of subjects each contributes merely one sentence to a story, then its plausibility can be manipulated by varying the number of previous sentences that the subjects are allowed to see before they contribute their own sentence. The method is plainly a variation on the traditional procedure for generating statistical approximations to English (Miller, 1951) except that our subjects each contribute a complete sentence rather than a single word. Texts generated according to our method do indeed approximate to normal depending on the size of the prior context that the subjects are allowed to read (Johnson-Laird, 1983, chapter 14). Such texts, however, also vary in their referential coherence. Our problem was to control this factor and our solution was to combine the method of generating approximations with a Cloze procedure. We took existing stories and deleted from them all the verbs and relational terms leaving behind only a skeleton of referring phrases for each sentence. Our subjects then generated the approximations by each completing just one sentence skeleton from the story. In order to vary the degree of plausibility we manipulated the number of prior completed sentences that the subjects were allowed to read before completing their own sentence. We predicted that the greater the implausibility of a text, the harder it would be for separate groups of subjects to understand and to remember.

## Experiment 1

### Method

*Design.* A panel of subjects constructed two variants of each of three folk-tales. One highly implausible variant (a 'first-order' approximation) was constructed by asking the subjects to complete separate sentences from the stories from which the verbs and other relational terms had been excised. Another less implausible variant (a 'third-order' approximation) was constructed in the same way except that the subjects read the two previous completed sentences before they filled in the gaps in the current sentence. In the experiment proper, a separate group of subjects acted as their own controls and read three different stories that varied in their plausibility: one was an original story, one was a third-order approximation and one was a first-order approximation. The materials were rotated across the subjects in order to ensure that each story was presented equally often in each of its three versions. The order of presentation of the three sorts of texts was counterbalanced across the subjects.

*Materials.* Three versions of folk-tales were constructed from stories in the literature: 'The old farmer and his stubborn animals' (adapted from Thorndyke, 1977), 'The magpie and the fox' (adapted from O'Sullivan, 1966) and 'The wise turtle' (adapted from Greenway, 1965). All three stories are similar in complexity and are examples of what Thorndyke calls 'single goal, single protagonist' narratives, and therefore belong to the same genre of stories. To construct the first-order approximations, each member of a panel of subjects was given a single sentence frame from each story, from which the verbs, relational terms and other such adverbials had been deleted. For example, the following frame: '...the farmer...the donkey...the shed', was derived from the sentence: 'One evening the farmer wanted to put the donkey into its shed'. The subject had to complete the frame by choosing terms that were appropriate. We call these approximations 'first-order' because the choice of words presumably reflects merely what relations subjects are likely to entertain about the various entities that are referred to in the individual sentences. The third-order approximations were obtained in a similar way except that the subjects were presented with two completed sentences followed by the sentence frame that they were to fill in, e.g. 'There once was an old farmer. He had a stubborn donkey. ...the farmer...the donkey...the shed.' The choice of terms in this case should reflect the context created by the two previous sentences as well as the referents in the third sentence.

The three original stories differed in length (223, 202 and 168 words, respectively). Since the approximations differed slightly in length from the originals, a small number of adjectives were added to the approximations to minimize the differences.

*Procedure.* The subjects were tested individually in a quiet room. They were told that they would be given three texts to read and that after each of them they would be tested on the material that it contained. When they had studied the first text for a period of two minutes, they were given a two-minute letter cancellation task to induce some forgetting. They were then told to write down as quickly and as accurately as possible all that they could remember of the text. The experiment continued in the same way for the two remaining texts. Finally, the subjects filled in a questionnaire about the task. The first question called for them to rate on a nine-point scale how easy the three texts had been to understand. The sixth question called for them to rate on a nine-point scale how surprising they had found the three texts. The remaining questions were designed to elicit systematic introspections about how the subjects had carried out the task, whether they had used imagery, etc.

*Subjects.* Twenty-six university students served to create the experimental texts. A further 18 students (11 women and 7 men) carried out the experiment proper. They were all unpaid volunteers.

### Results

We used two methods to score the subjects' recall protocols. First, we counted the length of a protocol and expressed it as a percentage of the length of the presented text. [Garnham *et al.* (1982) noted that this gross measure yielded an index that correlated with other more refined scoring procedures.] Second, we divided each passage up into clauses (main, subordinate, infinitival and relative), scored the protocol for the number of such items for

**Table 1.** Mean length of recall protocol (words) and mean number of idea units recalled in Expt 1, expressed as percentages of the total for the presented texts (each entry is summed over 18 subjects and three stories)

	Version		
	Original	Third-order	First-order
Recall protocol length (%)	75	68	58
Idea units recalled (%)	73	65	56

which the gist had been recalled, and finally expressed this score as a percentage of the number of such units in the text. Since we were concerned only with the recall of the gist of clauses, we refer to this measure as the number of 'idea units' that were recalled. Table 1 presents the lengths of the recall protocols expressed as percentages of the length of the presented texts, and the percentages of idea units that were recalled, for the three sorts of text. There was a reliable decline in memory for the texts as they declined in plausibility, and the trend was significant both for the percentage length of the protocols (Page's  $L = 235$ ,  $P < 0.01$ ; see Page, 1963) and for the percentage recall of idea units (Page's  $L = 238$ ,  $P < 0.01$ ). Analyses of variance confirmed these results and established that there was no significant effect of practice.

The three original stories were all rated as easy to understand (the mean ratings were 8.5 on the nine-point scale, and there was no significant difference between them). The mean ratings declined reliably as a function of implausibility (8.5 for the originals, 5.4 for the third-order approximations and 2.3 for the first-order approximations; Page's  $L = 237$ ,  $P < 0.01$ ). Conversely, the ratings of how surprising the subjects had found the texts increased as plausibility declined (2.1 for the originals, 4.6 for the third-order approximations and 6.8 for the first-order approximations; Page's  $L = 239$ ,  $P < 0.01$ ).

### *Discussion*

When a text is modified so that the sequence of events it describes is less plausible according to our procedure, it becomes more surprising, which provides an independent validation that the procedure does indeed manipulate the extent to which the text can be construed on the basis of general knowledge. The results of the experiment confirmed our prediction that such texts would be harder to understand and to remember; they also bear out our view that plausibility is not an all-or-none characteristic of a text, but rather one that can exist in varying degrees. The degrees of plausibility were achieved operationally without postulating or attempting to manipulate the different 'types' of plausibility discussed in the introduction. In their answers to the questionnaire, the subjects reported the use of a variety of strategies in trying to remember the texts. Some used imagery, others did not; some tried to impose a structure on the texts, others did not; some deliberately sought to ignore certain sentences; others did not. The trends in the data emerged reliably despite this lack of a uniform approach to the task.

Although the results are convincing, we had some minor worries about the materials of the experiment. Some of the sentence frames used in constructing the approximations contained pronouns, and our panel of subjects did not always make the correct assumption about their referents. Hence, there were some small deviations in reference from one version of a text to another: all the texts were referentially coherent, but they did sometimes differ slightly in their referential structure. Since the stories were folk-tales, the original versions were themselves often rather implausible. All three of them, for example,

featured animals that talked. The original versions also differed in length. We therefore decided to try to replicate our findings in an experiment that corrected these shortcomings.

## Experiment 2

### *Method*

*Design.* The design was very similar to the previous experiment. A panel of subjects constructed first-order and third-order variants of three stories by completing separate sentence frames from which verbs and other relational terms had been excised. In the experiment proper, a separate group of subjects acted as their own controls and read an original story, a third-order variant and a first-order variant. The materials were rotated across the subjects to ensure that each story was presented equally often in each of its three versions. The order of presentation of the texts was counterbalanced over the subjects.

*Materials.* We wrote three stories based on ordinary everyday experiences, each 150 words in length. The stories had clearly defined protagonists, events and outcomes, and they were of a similar degree of complexity. A group of subjects then generated the first-order and third-order approximations by completing sentence frames from which all but the noun phrases and pronouns had been deleted. If a frame contained any ambiguous words or pronouns, then their meanings and referents were spelt out, e.g. '...the youth...the package...him (refers to an old man)...he (refers to the youth)...bank (of a river)'. These sentence frames were then given to subjects to complete in the same way as in the previous experiment. They were instructed to add appropriate verbs, adverbs, adjectives or function words (but not to refer any to new entities) until they had constructed what they felt to be a satisfactory sentence in English.

Each subject created either three first-order sentences (one from each story) or three third-order sentences (one from each story); they were told that the sentences were totally unrelated. The resulting approximations were all shorter than the original stories. We accordingly added a few adjectives and adverbs to them in order to produce texts that were all 150 words long (examples are reproduced in the Appendix). With hindsight, the addition of such material may have altered the rhetorical significance of some passages, e.g. the third-order approximation given in the Appendix may have been altered by the addition of 'however' and 'so'. An alternative procedure might have been to force the judges who constructed the passages to use a fixed number of words; this procedure, however, would have made the task of producing completions more difficult. The maximum number of words added to any text was 14.

*Procedure.* The experiment was administered to two groups of nine subjects. Each subject was given a printed booklet containing the texts and a blank answer booklet, and the allocation of the materials to subjects was randomized. As before, the experiment comprised three cycles of a three-phase procedure: the subjects read a text for two minutes, they carried out a letter cancellation task for two minutes, and then they wrote down as much as they could recall from the text. The instructions were read out to each group at the beginning of the session. The subjects were told that they would be given three texts to read, each followed by a letter cancellation task and then by a comprehension test. The precise nature of this test, which again called for an attempt at verbatim recall, was not described until after the first letter cancellation task. The subjects were then asked to write down as accurately as possible all that they could remember of the text they had just read. Finally, when the subjects had completed the three recall tests, they were asked to put the passages into their rank order of comprehensibility.

*Subjects.* Thirty-two students and staff at the Applied Psychology Unit generated the experimental texts from the original stories. Eighteen women (ranging in age from 25 to 40 years) from the subject panel of the Applied Psychology Unit were paid £2 for participating in the experiment proper.

### *Results*

We again used two main measures of recall: the length of a protocol expressed as a percentage of the length of the presented text, and the percentage of idea units that was

**Table 2.** Mean length of recall protocol (words) and mean number of idea units recalled in Expt 2, expressed as percentages of the total for the presented texts (each entry is summed over 18 subjects and three stories)

	Version		
	Original	Third-order	First-order
Recall protocol length (%)	77	69	58
Idea units recalled (%)	76	64	55

recalled from the text. Table 2 shows these two percentages for the three sorts of text. There was, as we expected, a reliable trend in both measures: the more implausible a text, the poorer it was recalled (Page's  $L = 241.5$ ,  $z = 4.25$ ,  $P < 0.001$ , for percentage lengths; Page's  $L = 339$ ,  $z = 20.5$ ,  $P < 0.001$ , for idea units). Analyses of variance confirmed these findings, and established that there was no reliable difference between the two groups to whom we had administered the experiment.

Seventeen out of the 18 subjects ranked the texts in the predicted order for ease of comprehension: original stories easier than third-order approximations, which in turn were easier than first-order approximations (sign test,  $P = 0.0001$ ).

We computed the rank correlation between the order in which the idea units were recalled in each protocol and their order in the presented text. The overall values of Kendall's tau were as follows: 0.99 for the original stories, 0.98 for the third-order approximations, and 0.92 for the first-order approximations; and each of these correlations was very reliable ( $z = 19.75$ ,  $17.88$ ,  $16.45$  respectively,  $P < 0.001$ ). Yet there was a significant decline in the correlations as the plausibility of the texts diminished (Page's  $L = 233.5$ ,  $z = 2.92$ ,  $P < 0.05$ ). We also scored the protocols for the degree to which consecutive idea units in the texts were recalled with other material interpolated between them. The percentages of such interruptions were as follows: 24 per cent for the original stories, 29 per cent for the third-order approximations, and 36 per cent for the first-order approximations. This trend towards more disrupted protocols as the texts declined in plausibility was highly reliable (Page's  $L = 235.5$ ,  $z = 3.25$ ,  $P < 0.001$ ).

### General discussion

The experiments corroborated our predictions that subjects would find implausible texts harder to understand and to remember than plausible texts even where the referential structure of the texts was identical. They also confirmed our expectations about the quality of the recall protocols. As plausibility diminishes, there is a smaller correlation between the order of the ideas in a protocol and their order in the text, and a greater tendency for what were originally adjacent ideas to be disrupted by extraneous matter, whether invented by the reader or interpolated from elsewhere in the story. This apparent degradation of the original text could reflect the operation of a text or story schema but, in our view, it is probably caused by subjects' attempts to reorganize the material so that it makes better sense. The absence of plausible constraints also places fewer restrictions on readers who attempt to reconstruct the text from fragmentary memories of the original (Bartlett, 1932).

We used an operational technique for manipulating plausibility for two reasons. First, previous studies had established that materials selected on the basis of experimenters' intuitions (and validated by independent judges) can have an effect on comprehension. These experiments had examined limited subsets of causal, temporal and spatial relations.

We wanted to show that a general and easily replicable manipulation of plausibility also had effects on comprehension when referential coherence was held as constant as possible. Second, there is no existing taxonomy of plausibility and certainly no theory that allows this aspect of texts to be manipulated in an *a priori* way. An independent goal, though one that does not directly concern us here, was to advance towards a classificatory system by discovering the sorts of plausible relations that were violated in our first-order and third-order approximations.

Our technique for manipulating plausibility operates in a general way without concern for different types of plausible relation. It is based on the assumption that the hallmark of plausibility is the ability of readers to establish the appropriate connections between the various events and states of affairs that are described in the text. The potential relations between propositions can be analysed by way of a series of dichotomies. The first dichotomy is whether the text as a whole is relatively plausible or not. If it is, then the second dichotomy for any particular proposition is whether or not it is related to any other proposition expressed by the text. If it is isolated from the rest of an otherwise plausible discourse, then the information it conveys is hardly likely to be integrated into the model of the discourse, and the proposition is not likely to be remembered (Bransford & McCarrell, 1975). If a proposition is plausibly related to others, then the third dichotomy is whether or not it is relatively predictable from the previous discourse. Predictable propositions tend not to be remembered very well if allowance is made for their reconstructability, and conversely propositions that are surprising given their context tend to be well remembered (Graesser *et al.*, 1979).

To revert to the first dichotomy, what happens if a text as a whole is relatively implausible? The answer is that it is also less predictable. The different versions of the stories constructed using our technique differ in the extent to which they were rated as surprising, and texts were rated as more surprising as the order of approximation declined. Evidently, surprise alone does not guarantee memorability. What seems to be required is a surprising event within an otherwise predictable sequence: when everything is surprising, it is harder to remember anything (cf. Brewer & Lichtenstein, 1981).

Let us return to the question that we raised in the introduction, namely, which factor is more important for the cohesion and comprehension of discourse – is it referential coherence as, for example, Kintsch & van Dijk (1978) suggested or plausible knowledge-based relations as, for example, Schank (1975) suggested? In principle, the two factors are entirely distinct. Referential coherence is a property of discourse, whereas most aspects of plausibility are properties of events and states of affairs themselves. In real life, most sequences of events are plausible, but some can, of course, be aberrant and, in literature and other imaginative discourse, sequences can be highly bizarre. Yet, a sequence of events as opposed to a description of the sequence can in no sense be said to be referentially incoherent, since this property arises only when reference is made to the events by some means of communication. In practice, however, the two factors are much harder to disentangle because they are both often established only by inferences based on knowledge. We can illustrate this point by considering the following text (cited by Keenan *et al.*, 1984):

Things were getting very tense. Suddenly John punched George and knocked him out. Mary started screaming. I ran to the phone and called the police. Kathy ran for the doctor.

The text is coherent even though none of the five sentences makes explicit reference to any individuals or entities in common. According to the theory of mental models (Johnson-Laird, 1983), the crucial factor is that it is nonetheless a straightforward matter to establish referential coherence – and plausible relations – between the sentences. The



reference of 'things' in the first sentence is at first unclear, since it could refer to a physical system or to a social situation. The second sentence establishes that it refers to a social situation in which at least John and George are participants. A plausible relation between the state of affairs described in the two sentences is readily inferred: tension between John and George has led to John punching George. Although the third sentence, 'Mary started screaming', makes no explicit reference to anything previously mentioned, it is natural to infer that Mary is another participant in the situation, and this conclusion follows from the conversational convention that remarks should be relevant to what has gone before (Grice, 1975; Sperber & Wilson, 1982) and from the plausible inference that Mary started screaming because she saw John punch George. Thus, a plausible relation between events described in two sentences also helps to create a relation between their referents, i.e. Mary, on the one hand, is related to John and George, on the other hand, because she witnessed the former punch the latter. Similar inferences from the remainder of the text establish that the speaker and Kathy are participants in the social situation, and that there are plausible relations between their respective actions and the punch described in the second sentence. A model of the discourse accordingly represents five individuals interrelated in a social situation that revolves around a fight. Each new referent can be related to the situation introduced in the first sentence and thereafter only implicitly referred to since there is a convention that it is not necessary to introduce referents or establish relations between them if these are readily inferable.

Studies of plausible relations should obviously control for referential coherence, but it is easy to overlook the effects of implicit references based on inferences. For example, Keenan *et al.* compared pairs of causally related sentences, such as:

Joey's big brother punched him again and again  
The next day his body was covered with bruises

with indirectly related sentences, such as:

Joey's crazy mother became furiously angry with him  
The next day his body was covered with bruises.

Although the second sentence in both pairs is co-referential with the first sentence, the actual referential structure differs from one pair of sentences to another, and the degree of co-reference also varies. It is plausible to infer that Joey's brother punched him on the body. Hence, there is a strong referential tie in the first pair, because 'his body' in the second sentence is readily linked to the first sentence. The referential link in the second example is weaker, since there has been no implicit prior reference to Joey's body. Keenan *et al.* write that often highly related causes so strongly suggest the event described in a subsequent sentence that subjects may expect it before the sentence is presented. Such expectations, however, may also operate referentially in the way that we have just described.

What this analysis brings out is that the debate about referential coherence and plausibility is akin to the argument about nature and nurture: both factors are important because they interact. It is wrong to regard texts that lack explicit co-reference as establishing that referential coherence is irrelevant to cohesion (*pace* Keenan *et al.*, 1984). They may be referentially coherent as a result of inferences. Such inferences are often vital to establish co-reference, the membership of an individual in a class referred to by a generic label and the relation between the explicit referents of one sentence and those of another. Inferences that establish a plausible relation between the events described in different sentences often aid referential coherence, and vice versa.

The phenomena can best be understood in the light of current theories of comprehension

in which the information drawn from the literal meaning of the sentences of a text is a precursor to the construction of a richer model of its significance – the mental models of Johnson-Laird (1980, 1983) and Garnham *et al.* (1982), or the episodic models of van Dijk & Kintsch (1983). The surface sentences of the text cannot alone establish its significance, because the events they describe need to be interrelated. Readers construct models based on the text, but they rely continuously on their knowledge of the world, and often on their knowledge of the specific situation referred to in the text in order to establish general patterns of temporal, spatial, causal, intentional and deontic relations (e.g. Schank & Abelson, 1977; Black & Bower, 1980). The trouble with a text that is wholly implausible is that the contents of its sentences do not readily allow inferences to be made to establish the appropriate relations between the events they describe. No sensible coherent model of the text can be constructed. For example, it is very difficult to connect the following pair of sentences from one of our first-order stories;

He struggled with a youth who killed him.  
Then they stretched out indolently in the sun on the bank.

The context establishes that the pronoun 'they' in the second sentence refers to the man and the youth, and this sentence obviously attributes a voluntary action to both of them. Readers may search their model of the text for alternative antecedents to the pronoun, but such a search will be fruitless and disrupt comprehension. The only way to begin to make sense of the sentence is to imagine that the story has suddenly taken a supernatural turn, but nothing subsequently confirms this interpretation.

Previous studies have shown that referential coherence influences interpretation where plausibility is held constant (e.g. Ehrlich & Johnson-Laird, 1982). The influence of plausibility has similarly been demonstrated many times. But these studies have established either general effects without controlling for referential coherence (e.g. Bransford & Johnson, 1972), or highly specific effects involving restricted sets of plausible relations (e.g. Keenan *et al.*, 1984). We have been able to demonstrate general effects across passages in which referential structure is held as constant as possible.

Not all discourse, however, is referentially coherent and plausible: poetry abounds in language that is neither, although it is often difficult to understand and its interpretation may be more akin to problem solving than comprehension (Empson, 1961). Furthermore, coherence and plausibility are not the only factors affecting the interpretation of discourse. Discourse can be logically consistent or inconsistent; it can conform to or contravene formal principles governing certain highly constrained genres – the sorts of principle that originally inspired the notion of story grammars; and it can be rhetorically significant, i.e. revealing of its author's communicative intentions, to a greater or lesser degree, since some experimental evidence suggests that readers interpret texts differently depending on their attitude to the author (Ciechanowicz, 1983) or towards the content of the text (Carraher, 1984). Nevertheless, such responses to the logic, to the genre or to the author's intentions are likely to rest upon a fundamental ability to construct plausible and coherent representations of discourse.

## References

- Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.
- Black, J. B. & Bern, H. (1981). Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior*, 20, 267–275.
- Black, J. B. & Bower, G. H. (1980). Story understanding and problem solving. *Poetics*, 9, 233–250.
- Bower, G. H., Black, J. B. & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, 11, 177–220.
- Bransford, J. D. & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some behavioral investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 11, 717–726.

- Bransford, J. D. & McCarrell, N. S. (1975). A sketch of a cognitive approach to comprehension: Some thoughts about what it means to comprehend. In W. B. Weimar & D. S. Palermo (eds), *Cognition and the Symbolic Processes*. Hillsdale, NJ: Erlbaum.
- Brewer, W. F. & Lichtenstein, E. H. (1981). Event schemes, story schemes and story grammars. In J. Long & A. Baddeley (eds), *Attention and Performance*, IX. Hillsdale, NJ: Erlbaum.
- Brown, G. & Yule, G. (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.
- Carraher, D. W. (1984). The critical reading of news. Paper presented at the Seventh World Congress of Applied Linguistics, Brussels.
- Ciechanowicz, A. (1983). Effect of author's and subject's gender on perception of author and text. *Polish Psychological Bulletin*, 14, 107-112.
- Dooling, D. J. & Lachman, R. (1971). Effects of comprehension on retention of prose. *Journal of Experimental Psychology*, 88, 216-222.
- Ehrlich, K. & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior*, 21, 296-306.
- Empson, W. (1961). *Seven Types of Ambiguity*. London: Penguin.
- Garnham, A., Oakhill, J. & Johnson-Laird, P. N. (1982). Referential continuity and the coherence of discourse. *Cognition*, 11, 29-46.
- Garrod, S. & Sanford, A. (1977). Interpreting anaphoric relations. *Journal of Verbal Learning and Verbal Behavior*, 16, 77-90.
- Graesser, A. C., Gordon, S. E. & Sawyer, J. D. (1979). Recognition memory for typical and atypical actions in scripted activities: Tests of a script pointer + tag hypothesis. *Journal of Verbal Learning and Verbal Behavior*, 18, 319-332.
- Greenway, J. (1965). *The Primitive Reader*. Hatsboro, PA: Folklore Associates.
- Grice, P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (eds), *Syntax and Semantics*, vol. 3, *Speech Acts*. New York: Academic Press.
- Haberlandt, K. & Bingham, G. (1978). Verbs contribute to the coherence of brief narratives: Reading related and unrelated sentence triples. *Journal of Verbal Learning and Verbal Behavior*, 17, 419-425.
- Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Haviland, S. E. & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 13, 512-521.
- Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, 4, 72-115.
- Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge: Cambridge University Press.
- Johnson-Laird, P. N. & Garnham, A. (1980). Descriptions and discourse models. *Linguistics and Philosophy*, 3, 371-393.
- Keenan, J. M., Baillet, S. D. & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, 23, 115-126.
- Kintsch, W. (1974). *The Representation of Meaning in Memory*. Hillsdale, NJ: Erlbaum.
- Kintsch, W., Kozminsky, E., Stretby, W. J., McKoon, G. & Keenan, J. M. (1975). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior*, 14, 196-214.
- Kintsch, W. & van Dijk, T. A. (1978). Towards a model of text comprehension. *Psychological Review*, 85, 363-394.
- Mandler, J. M. & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9, 111-191.
- Mandler, J. M. & Johnson, N. S. (1980). On throwing out the baby with the bathwater: A reply to Black and Wilensky's evaluation of story grammars. *Cognitive Science*, 4, 305-312.
- Manelis, L. & Yekovich, F. R. (1976). Repetitions of propositional arguments in sentences. *Journal of Verbal Learning and Verbal Behavior*, 15, 301-312.
- Mani, K. & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory and Cognition*, 10, 181-187.
- Miller, G. A. (1951). *Language and Communication*. New York: McGraw-Hill.
- Miller, G. A. & Johnson-Laird, P. N. (1976). *Language and Perception*. Cambridge: Cambridge University Press.
- O'Sullivan, S. (1966). *Folktales of Ireland*. Chicago: Chicago University Press.
- Page, E. B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of American Statistical Association*, 58, 216-230.
- Rumelhart, D. E. (1975). Notes on a schema for stories. In D. G. Bobrow & A. Collins (eds), *Representation and Understanding*. New York: Academic Press.
- Sanford, A. J. & Garrod, S. C. (1981). *Understanding Written Language*. Chichester: Wiley.
- Schank, R. C. (1975). The structure of episodes in memory. In D. G. Bobrow & A. Collins (eds), *Representation and Understanding: Studies in Cognitive Science*. New York: Academic Press.
- Schank, R. C. & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Erlbaum.
- Schank, R. C. & Riesbeck, C. K. (eds) (1981). *Inside Computer Understanding: Five Programs Plus Miniatures*. Hillsdale, NJ: Erlbaum.
- Sperber, D. & Wilson, D. (1982). Mutual knowledge and relevance in theories of comprehension. In N. V. Smith (ed.), *Mutual Knowledge*. London: Academic Press.
- Thorndyke, P. W. (1976). The role of inferences in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior*, 15, 436-446.

62 Alison Black, Paul Freeman and P. N. Johnson-Laird

- Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 9, 77-110.
- Van Dijk T. A. (1977). *Text and Context: Explorations in the Semantics and Pragmatics of Discourse*. London: Longman.
- Van Dijk, T. A. & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.
- Wason, P. C. & Johnson-Laird, P. N. (1972). *Psychology of Reasoning*. London: Batsford.

*Received 17 December 1984; revised version received 16 April 1985*

Requests for reprints should be addressed to Alison Black, MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK.  
P. N. Johnson-Laird is also at the above address.  
Paul Freeman is at Scicon Ltd, London.

### Appendix

#### **An example of an original story used in Expt 2, together with its third-order and first-order approximations**

##### *Original*

An old man stood on a bridge overlooking a fast-flowing river. Suddenly, he noticed a package lodged in the reeds beneath him. Seized with curiosity, he tried to climb down to the package but he was unable to negotiate the bank alone. He called to a passing youth to help him. Together they descended the bank. When they reached the package, the old man opened it. He gave a cry of delight when he found it was full of banknotes.

The youth demanded that they share this windfall. The old man refused. Then the youth snatched the package from him and ran off up the bank.

As he approached the top of the bank, the youth tripped on a small stone. The package fell from his hands towards the river. Its contents scattered into the water. Sadly, the old man and the youth watched their new-found fortune disappear downstream.

##### *Third-order approximation*

An old man stood on a bridge overlooking a fast-flowing river. Suddenly, he noticed a package lodged in the reeds beneath him. Full of curiosity, he retrieved the package and he clambered up the bank rapidly. He called over a youth who knew him. Thoughtfully, they sat down on the bank. Although they did not own the package, the old man wanted to keep it. He gave an emotional cry when he heard it contained banknotes.

The youth suggested that they keep their windfall. The old man however was not persuaded. Generously the youth left the package with him whilst he examined the bank.

When he finally got to the bank, the youth shouted that he could only see a stone. The package fell into his hands then rolled towards the river. Its contents fell into the water. So, the old man told the youth to seek his fortune elsewhere.

##### *First-order approximation*

An old man fell off a bridge into a river. Furtively he dragged a package out of the reeds towards him. Out of curiosity, he opened the package and he hid it beneath the bank in safety. He struggled with a youth who killed him. Then they stretched out indolently in the sun on the bank. When they saw the package, the old man grabbed it. He uttered a cry when he realized it held banknotes.

The youth suggested they spend the windfall. The old man became most cantankerous. Sensibly the youth gave the package to him while he climbed down the bank.

When he reached the top of the bank, the youth threw down a stone. The package unravelled in his hands and fell into the river. Its contents were soon rotted by the water. Sagely, the old man and the youth agreed there was a fortune in store.