Cohesion-Establishment in Text Comprehension: A Case-Study of Anomaly Detection

Stephen B. Barton

Thesis presented for the degree of

Doctor of Philosophy

Department of Psychology

Faculty of Social Sciences

University of Glasgow

October 1991

For Mum, Dad and Gillian.

Abstract

This thesis develops the view that establishing a cohesive mental representation from a text demands elaboration by inference. Such cohesion-based inferential activity is necessarily incomplete, since there is no limit to information potentially relevant to coherence tests. Hence, cohesion is only more or less complete, possibly varying as a function of textual and contextual factors.

Text-based anomalies which pass undetected during reading act as signals for shallow underlying cohesion. The rate of detection for any particular anomaly is an index of processing depth. The detection rates for several manipulations of one anomaly are investigated. Cross condition comparison of the obtained rates allows evaluation of manipulation effects.

Over a series of eight experiments, considerable non-detection is observed, supporting the above views. The results suggest that the contribution any word makes to overall coherence is not a simple function of its own properties. Rather, these interact with more global constraints such as the background interpretative scenario and the pragmatic status of the message. **These factors have a controlling effect on on-line processing**.

Since pragmatic status, for example, may only emerge gradually over message presentation, selective influences on prior information can be delayed, even across sentence boundaries. These effects suggest that the time-course of full interpretation may be longer than has previously been thought to be the case. Implications for inference measurement and models of text processing are discussed.

List of Contents

List of Tables

Chapter Six

Chapter Seven

Chapter Eight

Chapter Nine

Chapter Ten

List of Illustrations

Acknowledgements

Thanks firstly to Professor Tony Sanford for inspiring the background to the work and for his loyal support, advice and encouragement throughout. He has been more than any student could have hoped for.

The Centre for Cognitive Science, University of Edinburgh, were of considerable help in the initial work for the thesis, and I am indebted in particular to Professor Keith Stenning for his co-operativeness when my needs seemed at their greatest. Similarly, I extend my thanks to the Department of Psychology, University of Glasgow, for their co-operation and interest throughout.

Many other people have been willing listeners to my thoughts as the research progressed. In particular, thanks are due to Anne Anderson, Chuck Clifton, Kenny Coventry, Simon Garrod, John Gillies, Erica McAteer and Linda Moxey, all of whom provided helpful comments at important junctures.

Hugh Pyper and David Morrison, as well as being close friends, have been a fundamental part of my whole postgraduate experience. The time would have been less fulfilling, and far less enjoyable, without them.

Many other friends have been very supportive. Their caring was vital at the times when my enthusiasm faded, and my journey has been far richer for the time spent with them. In particular, I must thank Douglas Bowie, Marion Corcoran, John Elliott, Peter Gallagher, Neil Hamilton, Lindsey Irwin, Clive Jachnik, Dorothy Jennings, Mike and Veronica Kinane, Margaret Lambie, Joe Lappin, Alasdair Mackenzie, Sally Menzies, Niven Richardson, Philip Sheldrake and Ronald Weitzman.

No experimental psychology would be possible without willing subjects. I am grateful for the co-operation of everyone who took part. Most entered into the spirit of the research, and many offered interesting thoughts that were greatly appreciated.

The Economic and Social Research Council supported me with a studentship for the duration of the research. Their help has been invaluable.

Lastly, many thanks to my family who have provided me with much-needed support in many different ways, always cheerfully and willingly.

Part One

Chapter One

An Introduction to Central Issues in Text-Driven Inductive Inference

1.1 Introduction

The purpose of this first chapter is to introduce major issues that have emerged in the study of text-driven inference. The aim is not to present an exhaustive account of all relevant psychological literature. Rather, some research is selected in order to illustrate particular themes. Chapter 2 deals more closely with the issue of processing completeness which is under direct investigation in this thesis. This topic can only be properly tackled in light of the information to be presented in the present chapter.

After illustrating the importance of general knowledge in text comprehension, we address various levels at which this knowledge is used. These levels are presented as different forms of inferential activity. The final section of the chapter will view these forms not as distinct goals of independent component processes, but as levels of representation fulfilled as the language processor produces thought from text.

1.2 The role of background knowledge in text comprehension

a: Insights from the study of memory for text

The importance of background knowledge in the comprehension of discourse was the pioneering discovery of Frederick Bartlett (1932). Although the focus of his work was centred principally upon the study of memory, it was no accident that through such study he uncovered principles equally fundamental for theories of language comprehension. His basic strategy was to test subject's memory for stories that had been read to them. He observed the accuracy of the subject's recall protocols taken at various time intervals following the hearing of the story. There were two basic findings. First, subjects were successful at capturing the general point or 'gist' of the story, but were unable to recall every detail of the text. Second, and perhaps more obviously relevant to the study of inference, embellishments or elaborations would appear in the protocols that were not part of the original 'text', but which were consistent with the 'gist' that had been extracted from it. These elaborations found their source in the part of the listener's general knowledge relevant to the topic of the text.

It is these recalled embellishments that are of import for theories of text processing, because their occurrence could be the result of inferential activity occurring during the text processing phase, not necessarily at the time of recall. Elaboration at recall has been described as 'deferred inferential activity' (Garnham, 1982) and is more likely to result because of memory limitations and reconstructive strategies: the inference is deferred until it is deemed necessary for some purpose instigated after initial processing. Such activity does not come under direct study in this thesis. What is under study are the factors controlling inferential activity during comprehension, in particular concerning the completeness of processing of semantic information. To place this topic within the broader context of text comprehension, it is appropriate to introduce themes and issues which have emerged in this field of study, principally over the last twenty-five years.

We have already observed, then, that discourse scientists have to pay a healthy respect to the conditions under which evidence for inferential activity is obtained, since elaborations to existing representations (or representations coming into existence) can, *a priori*, occur at any time. This understood, Bartlett's work clearly reduces the credibility of any notion that text representations are complete, memorised copies of the texts from which they arise. Although such states of affairs can occur (for example in the memorisation of poetry or some prayers: cf. Rubin, 1977), we would rightly be inclined to speak of such activity as a particular type of learning and considerably removed from the normal business of reading and understanding a piece of text.

Later psycholinguistic work supports the view that surface information from a text is lost soon after comprehension. Sachs (1967) established through observation of sentence recognition errors that memory for syntactic constructions of sentences is lost quite rapidly after processing. Clearly, the processor is interested in transforming the text into some representation which is built with respect to the semantic specifications of the input, as Sachs herself asserts,

'Even when the meaning of a sentence was remembered, formal properties that were not necessary for that meaning were forgotten very quickly. The results suggest that the original form of the sentence is stored only for the short term necessary for comprehension to occur...Thus the memory of the meaning is not dependent on memory of the original form of the sentence.'

Further to his experimental observations, Bartlett was the first psycholinguistic researcher to emphasise the importance of the mapping that is made between the text and the agent's knowledge in the comprehension process. Parts of a discourse were thought to be understood with respect to such general knowledge, and this knowledge could be highly culture-specific and stereotyped. Bartlett's term for any particular packet of knowledge was a 'schema'.

After Bartlett, the study of text comprehension, and text-driven inference in particular, diminished for a gap of almost forty years. The work that was completed in the intervening period focused almost exclusively on factors affecting recall and memorisation of word strings, often in response to increasing developments in the field of linguistics (Tulving, 1962; Marks and Miller, 1964; Bobrow and Bower, 1969). At one level, this work was similar to Bartlett's. It studied the relationship between the presentation of items of information and the corresponding memorisation of them. However, the new work focused almost exclusively on word-string and digit configurations, and consequently there are question-marks over the relevance of such to the comprehension of discourse in general. The principal limitation is that memorisation at that level could move subjects to adopt particular processing and memory strategies. The processing machinery may become configured in order to handle the characteristics of a particular input. This could give rise to performance charactersitics at odds with those of normal reading. We have already witnessed a bias against such levels of retention in the work of Sachs. For example, there is considerable literature reporting on 'chunking' memorisation techniques, where relationships between inputs are stored with respect to some existing meaningful material in general knowledge in order that each input (e.g. digit or syllable string) can be retrieved <u>as originally presented</u> (c.f. Chase and Ericsson, 1981). Such memorisation strategies are also facilitated by the deliberate clustering of groups of inputs together, this presumably helping encoding into working memory (c.f. Miller, 1956; Ryan, 1969). No evidence from discourse science suggests that this is normal processing activity for reading texts.

Later research returned to the more global issue of comprehension of texts per se, and developed on the theme of the role of general knowledge initiated by Bartlett. We shall observe the extent to which general knowledge has come to be viewed as vital not only for inferences that embellish the textual content, but also for the resolution of ambiguities and the maintenance of cohesion within the text itself.

b: Titles and perspectives

Dooling and Lachman (1971), like Bartlett, were interested in memory for prose. Their manipulations involved varying the comprehensibility of texts through the inclusion or otherwise of informative titles for the texts, and they tested memory for discourse of subjects under both free recall and recognition conditions. Under both conditions subjects were able to recall more effectively where their comprehension had been aided by the inclusion of an informative title.

For our purposes, the finding of interest is that appropriate titles appear to facilitate the comprehension process, and this is reflected in the accuracy of the memory trace for the passage. Interestingly, the inclusion of a title appears to be most effective where it is provided <u>prior</u> to the main body of the text (Dooling and Mullett, 1973), which suggests that a title helps the reader to select the most appropriate knowledge to be used in the interpretation of the discourse <u>that is to follow</u>. The title effects some control over the reader's expectations of the text's likely content. Naturally, titles can be more or less informative depending upon how well that content is signalled. It has emerged, however, that comprehension of texts with less metaphorical content than those used by Dooling and Mullett can be facilitated by titles given after the main body of the text (e.g. Bransford and Johnson, 1973).

This gives rise to a series of issues involving the knowledge selected to be used in on-line comprehension, the resulting memory representation, and the flexibility of that representation to be reconstructed from a different perspective or in light of new relevant knowledge. These issues are best introduced by describing work on memory for discourse by Anderson and Pichert (1978). Their experiment involved asking subjects to read passages of text and in doing so adopt a particular perspective on it. For example, in a text about a house, subjects had to imagine that they were reading from the perspective of a burglar or, in another condition, a housebuyer. As one might expect, the adopted perspective of the subject effected the details of potential interest in the text, and consequently in memory protocols from the 'burglars', information pertinent to burglars would predominate, and a similar pattern emerged for the 'housebuyers'. This supports the general view that the knowledge brought to the text at the time of comprehension is highly influential in effecting the form of representation that will result from it. In this case, anticipation of the purpose to which knowledge of the text could be put ('this knowledge could be used in a future burglary') has a natural consequence for the importance of particular parts of the text, and presumably has a controlling effect on the attention paid to them.

The extreme opposite interpretation of these results would have it that the salient features relevant to the reader-perspective manifest as an outcome of reconstructive activity at the time of the memory test. We have already alluded to such possible processes. This view would propose that adopting a perspective becomes possible subsequent to the comprehension of the text material, but not during it. Of course, this is an extreme view introduced principally for purposes of illustration. It seems beyond doubt that an adopted perspective can effect processing at comprehension. However, Anderson and Pichert explored the intriguing possibility that subjects instructed to read from one perspective could change to another at the time of recall. Hence, those subjects asked to read from the point of view of a burglar, were then asked to recall from the point of view of a housebuyer. They found that subjects could recall information relevant to the second perspective, and the results suggested that the adoption of the new perspective opened up access to stored information that was not accessible via the initial point of view.

This raises interesting issues concerning the specificity of the encoding (or comprehension) process and the subsequent accessibility of information from the resultant memory representation. We need to make recourse to the work of Tulving (1968), who drew a distinction between the <u>availibility</u> and <u>accessibility</u> of stored information. With reference to the Anderson and Pichert work, there are subjects who change perspective at recall and remember details not found in the protocols of those who retain their encoding perspective. These subjects have that information <u>accessed</u> by searching through the memory in a different way. However, that the same information is just as likely to be available in the minds of those not prompted to alter perspectives. So a new perspective could effectively reconstruct the existing memory representation. The control of the accessibility and availability of particular references arising from texts, or suggested by them, is likely to have important implications for understanding the sort of inferential activity that becomes activated during comprehension. This issue shall be discussed in greater depth in subsequent sections.

By similar reasoning to that of the perspective changes, knowledge accessed after the comprehension of the text could be used to re-organise the existing representation, and could make the prior text more comprehensible. Bransford and Johnson (1973) have a particularily striking example of this. They give the following sentence (1).

(1) The haystack was important because the cloth ripped

This is difficult to comprehend at any depth without the cue word 'parachute'. Here the cue word serves to bring to mind situations in which parachutes are used, and from the instantiation of this knowledge, the cloth ripping can be taken to refer to a parachute failure. The importance of the haystack is that it provides a semi-comfortable landing site for the unfortunate parachutist.

Bransford and Johnson describe the process as follows:

'Here the word parachute does not simply specify a reading for cloth. It also sets up conditions for realizing the relations between the cloth and the haystack, namely that the parachute was above the haystack when it ripped. (This example) illustrates that the contributions that an individual must often make in order to comprehend include more than a specification of appropriate individual referents. He must generate appropriate relations among

entities as well.'

The cue word accesses knowledge that brings to mind a context or situation which allows the description to become meaningful. In so doing, the text can be accurately viewed as a partial description of a possible or actual state of affairs. Constructing a representation of these states of affairs is the business of text comprehension. Also crucial is the mapping activity that takes the references in the text and allocates them to roles that exist in the reader's general knowledge of these situations. With the benefit of the cue word, interpretation of sentence (1), results in the 'cloth' standing for part of the parachute in the discourse model.

c: Schema-based theories of comprehension

It is this organised general knowledge of situations that Bartlett first termed 'schemata'. We can think of these knowledge-structures as retaining both necessary and variable specifications of information that have been true of the world in the past and are anticipated as having predictive power in the future. In effect, different knowledge-structures represent different situations, properties, objects and individuals that are part of the cognitive agent's world. These structures can be utilised in understanding and organising incoming descriptions that make reference to the same properties of the world that the 'schemas' represent.

Throughout the 1970's, there were various theories proposed suggesting the importance of the selected knowledge-structure in the interpretation of linguistic information. These have been given different names by different researchers who varied their emphasis and applicability to other forms of information processing. Minsky (1975) introduced his general 'frame' concept, Rumelhart (1975) re-introduced the 'schema' idea in accounting for narrative structure, and with more specific applicability to text comprehension, Schank and Abelson (1977) introduced their 'script' concept, and Sanford and Garrod (1981) the 'scenario'. The intention here is not to give detailed description and evaluation of the respective theories, but rather to introduce and develop the importance of world knowledge in the interpretation of language. Some of the work of these researchers will be described more closely in relation to later topics.

The 'haystack' example (1) of Bransford and Johnson illustrates the advantages of using general knowledge in the interpretation of a single, short sentence, but such knowledge is just as necessary to interpret passages of discourse or whole texts. This can be illustrated by returning to the notion of the informative title first studied by Dooling and Lachman. Consider a reader processing the title given in (2).

(2) **'Eating out'**

Under the general schema-based view of comprehension, this title serves to select knowledge from the reader's memory to do with eating out in restaurants etc. This particular schema will contain specifications of things that are always true of restaurants (e.g. they sell food, contain staff etc.), to those that have variable specifications but can be expected to be true of restaurants in general, or even of only some restaurants (e.g. role information; waiter, customer, manager: procedural information; customer enters, shown to seat by waiter, given menu, etc.: physical information; normally housed in a building etc.).

This knowledge-structure (made famous as an example by Schank and Abelson, 1977) allows text-based descriptions to be mapped as references onto the variable slots in the structure. Hence, references to some named individual reading the menu can be understood by mapping that individual onto the schema as fulfilling a 'customer-role'. Or a reference to another individual handing out a menu could be understood by mapping that individual as fulfilling a 'waiter-role'.

For present purposes, the important point to be made is that the mapping process from input to knowledge-structure can be described as inferential activity, since information not inherent in the text is becoming incorporated in the comprehension process and therebye the resultant discourse model. In simple terms, readers have to infer that a particular character is a waiter or customer on the basis of what they already know about the normal activities of such people in those situations.

In support of the general schema view, there is evidence from Bower, Black and Turner (1979) that memory recognition errors are prone to occur for events not explicitly mentioned in texts, but which are presumably incorporated from an episode-schema as part of the overall text representation. The incorporation of these events result from inference that preserves the relations between the described episodes. An example is given in (3).

(3) **'Eating out'**

Mary went into her favourite fish restaurant for some haddock.
She paid by cheque.

(Possible recognition confusion: e.g. 'Mary ate haddock')

They also found that events which are abnormal with respect to the expected episode-structure are well remembered in subjects protocols. This suggests that such events achieve salience in the reader's mind by deviating from the schema-based expectations. We shall analyse in some depth the status of the memory recognition paradigm to try and establish how far it is sensitive to the conditions under which inferences are actually made. Whether or not inferences are deferred subsequent to comprehension, their incorporation presupposes an existing source of relevant knowledge that can provide plausible and usable elaborations on the basis of what is given. The Bower, Black and Turner data support the view that schema-type knowledge structures serve this function. Mapping text onto schematic structures also provides a possible mechanism for the establishment of 'gist' representations, first discovered by Bartlett. Episode-schemata will house the fixed plan-units that serve the main goal of the whole event, but incidental details will be less likely to manifest irrespective of their consistency with the main structure. This level of generality is characteristic of gist representations.

The implication of world knowledge in active, on-line processing licences inferential constructions of different forms. All of this activity has a common theme: it effects an elaboration upon what is given in the text. Different forms of elaboration will fulfil different ends, and some of these forms will be presented over the next two sections. The following section 1.3 introduces inferences that serve to establish the coherence of a text, and also inferences that go beyond coherence establishment into coherence embellishment (or elaboration). Section 1.4 will discuss work carried out on inferential activity that functions to establish cohesion from a text. The final section 1.5 provides an overall discussion of the issues raised.

1.3 World coherence

a: Coherence: establishment and embellishment

We have observed that the mapping process between text-based references and schematic role-slots renders the text a partial description of a possible or actual state of affairs. It is partial in the sense that there is always room for the adoption of new perspectives or further levels of description that could be true of the world being described. So the relation between the text and the world is never complete. It follows from this that there is a potentially infinite set of inferences that can elaborate upon any text, even if some of these are unlikely to be true of the world. Under some conditions it will be of benefit for a reader to draw some inferences on the basis of what is written in a text. The fundamental issue for an understanding of inferential control is to discover the conditions under which inferences will be drawn.

A text's incompleteness does not entail incoherence. The form of coherence being referred to in this section concerns the potential of a text to **drive a representation that interprets text-based references as descriptions of the same states of affairs in the world**. Segment (4) is an example of a text which, by itself, fails to acheive this level of coherence.

(4) *The angry husband threw the delicate vase against the wall.*

(4') *It cost a lot of money to replace.*

There is a content discrepancy between the two sentences. For the second sentence to make meaningful reference to the same state of affairs signalled by the first sentence, an inference has to be drawn to account for the need of the vase to be replaced. The most plausible evaluation is that the vase broke after making contact with the wall. The inference helps to establish a coherent representation from the text. The claim is not that the text is incoherent per se, but rather that inferred relations are required to acheive its coherent potential. These inferences relate to information about the world being described.

It is possible that a reader would want to infer information, even where the inference was not a condition required to establish the coherence of the text (in the sense of coherence being used here). Segment (5) is an example of this.

(5) *The angry husband threw the delicate vase against the wall.*
(5') *He had been feeling impatient for some time.*

The information that the husband had been feeling impatient for some time does not force the reader to infer that the vase broke. This inference is not necesssary to establish the comprehensibility of the text as a unit. However, the reader may which to embellish her model of the 'world' with this information anyway, and the likelihood of it being true of the world may be assessed on the basis of the other information given in the text. These types of inference which, so to speak, fill out the picture (but are not necessary conditions for having a picture in the first place) have been referred to in the psycholinguistic literature as 'elaborative inferences'. This is not a good term for them because all inferential activity makes some kind of elaboration, and there is therefore an underspecification of what is really being meant. In the present thesis, this form of inference will be referred to as **coherence embellishment**.

It is possible that a committed embellishment will become necessary for coherence purposes later in interpretation. For example, if the embellishment inference 'the vase broke' is drawn after reading (5) and (5'), and sentence (4') is read thereafter, the conditions required for coherence are effectively the same as in example (4).

(5) *The angry husband threw the delicate vase against the wall.*
(5') *He had been feeling impatient for some time.*
(4') *It cost a lot of money to replace.*

Following through this rationale, it is said that such inferential activity is made forward of when it is needed. If the inference is only drawn under coherence establishment conditions then it is described as a backward inference, since the relation is composed subsequent to the identification of the required information. Opinions differ with respect to the prominence of forward and backward inferential activity.

Another distinction frequently made by some discourse scientists is between necessary and elaborative inferential activity. For example, McKoon and Ratcliff (1990) speak of elaborative inferences as those which

> *'go beyond what is actually **required** to connect the explicitly stated ideas in a text'.* (my emphasis)

Necessary inferences are taken to be those that do support such connections. One can see the motivation behind the distinction. It is similar to the one being made between establishment and embellishment. However, it has already been stated why the concept of an elaborative inference is less than informative for taxonomic purposes: connecting explicitly stated ideas does not preclude elaboration. Further, conditions are only necessary with respect to the ends they fulfil. Assuming that connecting explicitly stated ideas is a necessary function of a language processor, and that necessary inferences will always be drawn, risks having a taxonomy that is prescriptive rather than descriptive. The main body of this thesis will investigate how complete such 'necessary' activity is.

Interestingly, more work has been carried out on embellishment activity than basic establishment processes, probably because the latter has been more closely equated with issues surrounding text cohesion (which shall be discussed in section 1.4). To maintain continuity in topic presentation, the embellishment inferences will be discussed first.

b: Some research on embellishment inference

We have observed that embellishment inferences add on extra information to an existing coherent representation, and that such additions will be potentially effective where they are made plausible or predicted by the content of the text. Note again the importance of relevant background knowledge in providing information, cued by the text, which makes available likely outcomes or high-plausibility co-occurances to embellish the representation. Research on embellishment inference will be presented in two broad sections, the first concerning causal inferences, and the second verb-instrument inferences.

{1} Causal inference

Causal inference is a major research area in its own right, not restricted to text-based information processing (e.g. Hilton, 1988). However, the term is used here to refer to a class of embellishment inferences which arise from predictions of a causal nature, most typically concerning the probable consequences or outcomes of described events.

Johnson, Bransford and Solomon (1973) conducted an experiment where subjects were read texts which strongly implied certain outcome. An example is given in segment (6).

(6) *It was late at night when the phone rang and a voice gave a frantic cry. The spy threw the secret document into the fireplace just in time since 30 seconds longer would have been too late.*

This text clearly does not demand the inference that the document was burned in the fireplace for purposes of coherence. However, knowledge of the type of situation being described might lead one to assume that such an outcome is highly probable. What Johnson (et al) tested was whether or not subjects would falsely recognise passages, with the plausible consequence explicitly stated, as having been those passages that had been read to them. The corresponding test passage for (6) would be (7).

(7) *It was late at night when the phone rang and a voice gave a frantic cry. The spy burned the secret document just in time since 30 seconds later would have been too late.*

In comparison with control groups who heard similar original passages (but without the implication of the predicted outcome), significantly more false recognitions were recorded.

Thorndyke (1976) utilised a different procedure in his study of discourse inference. He had subjects read passages, and at certain points in the discourse they were asked to write down statements which they <u>believed were likely to be true but were not explicitly stated in the text</u>. This is an inference generation procedure. The texts in question were longer than in the Johnson, Bransford and Solomon study (mean length 20 sentences) and by manipulating the content of particular sentences of the text across conditions, Thorndyke was able to gauge the effect such variation had on the resulting inference generation patterns. His often-quoted example is given below in sentences (8) to (13).

(8) *The hamburger chain owner was afraid his love for french fries would ruin his marriage.*

Although Thorndyke does not explain the inferential activity in this way, early in this text a relation, the underlying cause of which is not stated, is explicitly introduced (why should the owner have such a fear?). It would not be surprising if the comprehension system were to be interested in resolving this implicit question, and as such, potential embellishment inferences may be generated. Thorndyke gives three possibilities, as in (9), (10) and (11).

(9) *The hamburger chain owner got his french fries for free.*

(10) *The hamburger chain owner's wife didn't like french fries.*

(11) *The hamburger chain owner was very fat.*

Later in the text, one of two actions of the owner is introduced (depending upon the condition of the experiment). These are given in segments (12) and (13).

(12) *The hamburger chain owner decided to join weight-watchers in order to save his marriage.*

(13) *The hamburger chain owner decided to see a marriage counsellor in order to save his marriage.*

Under Thorndyke's rationale, including information about the owner seeing a counsellor should have no effect on the plausibility of any of the possible inferences in (9), (10) or (11). But including information about the owner joining weight-watchers should preferentially support inference (11), that he was very fat, over inference (10), that his wife didn't like french-fries. A visit to weight-watchers would suggest that it is the owner's fatness that is the problem rather than anything else. The plausibility of inference (9), that he got his french-fries free of charge, should remain unaffected.

The frequency of subjects references to the potential inferences supported this view as did subjects own plausibility ratings of these possibilities where they were made. Further, Thorndyke followed up this study by carrying out a false recognition procedure (as used by Johnson, Bransford and Solomon) on the inference statements of interest, and a similar pattern of 'preferences' was received to the inference generation procedure.

The results of these experiments would strongly point to the view that embellishment inferences relating to probable outcomes are prevalent in text comprehension. There are, however, problems in correctly interpreting the findings. The general problem in all false recognition procedures is that assumptions have to be made concerning the relation between any subjects memory for a particular text <u>at the time of the recognition decision</u>, and the memory representation that resulted from comprehension activity <u>at the time of text processing</u>. This distinction borrows from Garnham's (1982) idea of deferred inferential activity, already mentioned in section 1.2.

We have little understanding of the extent to which models of texts become decayed following comprehension. This presupposes that decay will occur, of course, but it seems consistent with all other evidence presented thus far that without continued attention and re-processing, our models for specific discourse segments become fragmented, or incorporated within some larger representation (e.g. your memory for this chapter thus far). Various factors could effect this process. In particular, the information load present upon the cognitive agent in the post-processing phase would appear to be important. In the experiment mentioned by Johnson, Bransford and Solomon, the inferred-outcome materials were part of a compound experiment also testing for the occurrence of other inference types (some to be introduced later), which forced subjects to listen to a total of 20 unrelated discourse segments, of comparable length to the example given in (7), prior to the test sentences being presented for recognition decisions.

Because of the extent of the memory load in such cases, it is possible that the level of detail present for any particular text representation will be lost in the service of allocating resources to processing of other texts or the maintenance of other representations. What may remain of the memory for any particular text could be an approximation to the original model. The fragmentation process may result in a memory-trace that amounts to a simple pointer to the more stable general knowledge used in the interpretation. So, for example, at some point the original memory for (7) may be reduced to 'having something to do with spies'.

Hence, false recognitions could be the artifact of some criterion-acceptance of a pattern-match between a newly processed input (the test segment) and the possibly fragmented (and hence <u>more</u> generalised) memory for the original discourse. It is the latter of these that could contain situation-specific information that would detail probable outcomes. The possibility exists, of course, that the incorporation of these plausible causal relations has occurred at comprehension. Clearly, however, **the false recognition paradigm is not a sensitive enough inferential probe to make discriminations at this level**.

Thorndyke's sentence generation paradigm can be called to question on similar grounds. We have already noted the importance of discerning the extent to which evidence for inferential activity can be due to reconstructive procedures incorporated subsequent to initial processing. Inference generations could be reflecting subject's efforts to make a selection of plausible inferences as a direct consequence of being given a problem solving task of this nature. Under these conditions, the relevant search domain would consist of information (or pointers to information) about likely inferences on the basis of the text. The dilemma for this form of investigation is that there is no guarantee that the search domain is not an <u>adaptation</u> of the current text model to fulfil the ends imposed by the generation procedure. Hence, **the inference generation procedure may only be of value as a general index of inferential plausibility and not of on-line activity.**

This insight sheds light on some limitations of Thorndyke's overall approach. Where he asserts that the inferences (9), (10) and (11) are all drawn in the comprehension process, this underestimates the importance of two points. The first of these is that the infinite number of inferences that can be drawn on the basis of any text can lead to an inferential explosion. Clearly, not every inference that has any degree of plausibility is drawn: the system has to be constrained. In Thorndyke's work no indication is given to the grounds on which the generation of such inferences are constrained, nor is it indicated that the plausibility of such inferences will be highly inter-related. To return to example (8), the hamburger chain owner's wife may have a dislike for french fries, (10), because of the fatness they are assumed to have caused in her husband, (12). <u>An increase in the plausibility of one inference does not necessarily entail a decrease in another</u>.

The second, and related, point is that it would often appear to be disadvantageous to construct fully-blown inferences in on-line processing for information that exists at a given time only as a <u>possible embellishment</u> to the text. What (9), (10) and (11) represent are possible conditions which could resolve the implicit question triggered by the original statement (8). Indeed, one expectation that a reader may have after reading sentence (8) is that further information would be provided in the text which would explicitly introduce or implicitly indicate the unknown relation. This occurs in Thorndyke's materials.

A theory of inference is required which does not presuppose that such activity occurs either in a complete form or not at all. Where we can have sympathy with Thorndyke's line of thought is in recognising that it is the build-up of information in the text that contributes to the plausibility, or otherwise, of possibilities that may become inferred. In harmony with Sanford (1990a), it is better to think of inferential control as a process where the intention of the writer, manifested in the text, signals to the processor the forms of knowledge that are relevant and should be easily accessible, perhaps at some point to become fully instantiated inferences. The concluding section of this chapter will question what is meant by a fully instantiated inference.

Recent research has attempted to utilise more sensitive measures of embellishment activity. Singer and Ferreira (1983) used a timed multiple-question procedure, where the time taken to respond successfully to questions about a text was taken to be an index of the likelihood that the answer to that question had been inferred during text comprehension. Their texts typically contained eleven sentences, some of which invited embellishment inferences relating to probable outcomes. Segment (14) gives an example text.

(14) *Bob the spy read a report by the fire.*
A rock flew through the windowpane.
Bob read a note attached to the rock.
He quickly threw his report in the fire.
The ashes floated up the chimney.
Next he called the airline.
He placed the coded sugar cube in water.
He poured the clear liquid into the drain.
Bob left and flew to a tropical resort.
He sat all the next day in the sun.
But Bob knew he was not safe here.

Notice that the fifth sentence here forces backward coherence maintenance activity to be related to the content of the previous sentence. Consequently, answering a question such as 'Did Bob burn the report?' would access the inferable coherence information that this must have been the case for ashes to result in the chimney. Such answers occured reliably faster than the corresponding set to a text identical except for the inclusion of the coherence-forcing fifth sentence. The latter case only provides conditions for embellishment activity, where the burnt nature of the report is only a possible consequence of it being thrown in the fire.

Singer and Ferreira concluded from this that people draw backward coherence inferences more reliably than forward embellishment ones. However, both McKoon and Ratcliff (1986) and Potts, Keenan and Golding (1988) have pointed out limitations of the methodology used. Since the question probes occur <u>after</u> reading the text, it is possible that the question initiates a checking procedure which tests the compatibility of the text with a question response. If the question was 'Did Bob burn the report?', then the more compatible such an outcome is with the text model, then the faster an affirmative answer can be computed. Again, this would be a form of deferred inferential activity. It could be that the latency response facilitation for the backward coherence inferences occurs only because the extra sentence in this condition contains information that facilitates backwards integration of the test material. Hence, inferential activity could be occurring to the same or different degrees (or not at all) in the establishment and embellishment conditions, and the timed multiple-question procedure would be insensitive to such variation.

Other inferential probes have been used in the investigation of embellishment inference. McKoon and Ratcliff (1986) presented a group of experiments comparing probe-measures for texts that either predicted or were neutral with respect to various outcomes. An example of one of their predicting texts is shown in segment (15), with the corresponding non-predicting control sentence in segment (16).

(15) *The greedy teenager sat down at the table, put her napkin on her lap, and picked up her fork.*

(Predicted outcome that the teenager will eat: test word EAT).

(16) *The teenager was greedy when choosing bridal tableware; she wanted napkins to match her knives, forks and spoons.*

(Test word EAT)

Their first probe, using a recognition decision paradigm, involved subjects deciding whether or not a word presented subsequent to reading a text had occurred in the text. The speed and accuracy with which such decisions are made can provide information concerning the inferential status of the test word. McKoon and Ratcliff found that correct responses to the test word were slower when the event described by the test word was predicted than when it was not. They acknowledge that this effect could be due to various factors. It could be that extra time is required to disambiguate inferred information from the memory of the explicit text, confusion resulting because the text model has already primed the concepts related to the test word before it is presented. Or it could be (comparable with sentence recognition effects) that there is no prior activation, and the increased latency results from a compatibility between the test word and the prior text. Hence, subjects take longer to make decisions about the actual content of the text. Because of these contrasting interpretations of the latency effect, McKoon and Ratcliff recognise that this procedure does not offer any conclusions concerning when inferential activity occurs, but rather is potentially useful in giving an indication of the information that would be easily accessible in normal reading. We shall return to this point in the final section of this chapter.

McKoon and Ratcliff also carried out surprise cued-recall tests, where following a delay of five minutes after reading, the relevant test words were presented to help subjects try and remember the content of the texts. Those test words that predicted the outcome were found to be more effective memory cues for the texts. This method is also limited by the possibility that it is backward integration that effects the memory facilitation. We shall look more closely at cued-recall techniques in the following sub-section on verb-instrument inference.

Finally, McKoon and Ratcliff used what they call a 'speeded item recognition' procedure to test for the presence of embellishment inferences in relation to predicted outcomes. This paradigm used a prime word following a text and then a target test word for sentence recognition decisions. The prime word could be neutral ('ready') or could come from one of the texts that had been read. The hypothesis again was that if the predicted outcome had been inferred during comprehension then the correct ('no') response should be slow and/or inaccurate. The main finding was that response times were slower for predicting texts, but only when the prime word was taken from the text. Where this procedure was adapted to force responses within set time limits (hence reducing the likelihood of the adoption of strategies) there were no differences in response times, but there were significantly more errors for the predicting than the control texts, again where the prime word is taken from the text.

McKoon and Ratcliff's conclusion from this finding is that there is a 'minimal encoding' of the outcome perhaps in a general form that the incoming test word can make contact with, thus accounting for the increased errors for predictive texts. The text-based prime is necessary to facilitate that contact. The notion of minimal encoding seems to be consistent with the view that drawing inferences is not an all-or-nothing process.

Potts, Keenan and Golding (1988) have raised objections to McKoon and Ratcliff's methodology. While accepting that the speeded item procedure prevents time-consuming strategies, recognition decisions still demand a compatibility check from the test word to the text representation. Putting a constraint on the time available does not change the nature of the task, and it remains insensitive to the time-course dimension. Potts et al argue that such comparative probes are necessarily limited in this way, and that measures such as lexical decision (deciding whether a presented letter-string is a word) and word naming (the time taken to voice the name of a presented word) are more likely to indicate the presence of inferential activity occurring prior to the onset of the probe. Using these techniques on similar materials to McKoon and Ratcliff, they conclude that response times to word naming probes are least likely to be susceptible to backward compatibility activity. They only find evidence for inferential activity under coherence establishment conditions and not under possible embellishment conditions.

We can see from the selection of studies presented in this section the importance of the properties of the inference-probe in assessing the conditions under which inferential activity occurs. The problems found in discerning the claims of the above researchers are rediscovered in the work to be described on verb-instrument inferences.

{2} Verb-instrument inference

Verb-instrument inference can again be illustrated by the work of Johnson, Bransford and Solomon (1973). The rationale behind an instrument inference is that upon the description of an activity, the instrument used to effect that activity is inferred. Johnson, Bransford and Solomon give the example in (17).

(17) *John was trying to fix the birdhouse. He was pounding the nail when his father came out to watch him and to help him with the work.*

This activity implies that John was using a hammer, since hammers are the normal things used in pounding nails while fixing birdhouses. Mapping the verb 'pound' into the scenario for such situations makes available the instrument information. Johnson et al tested subjects on such materials using the false recognition procedure. They found that subjects were more likely to falsely recognise test sentences with the instrument explicitly stated following materials where the instrument had been implied. This suggests the incorporation of the instrument, when implied, in the text representation. This evidence is subject to the same limitations as those studies previously mentioned.

The cued-recall paradigm mentioned in relation to Thorndyke's (1976) work has also been used in the study of instrument inference. Paris and Landauer (1976) found that adult readers show facilitated recall of sentences cued by implied verb instruments. One of their examples is given in (18).

(18) *The teacher cut into the juicy steak.*

(Cue word; knife).

We already know that the cue word could be having a positive effect on recall without actually being inferred at comprehension. It could merely be assisting in accessing existing representations with which it can make contact. Evidence for this reconstructive view comes from Corbett and Dosher (1978). They showed that highly plausible verb instrument cues could facilitate the recall of sentences where a less plausible instrument had been explicitly stated. An example is given in (19).

(19) *The athlete cut out an article with a razorblade for his friend.*

(Cue word; scissors).

Further, subjects were still able to successfully recall the instrument that had been explicitly stated, strongly suggesting that the cue-word was not part of the representation developed at comprehension.

The recognition decision paradigm has also been used to study instrument inference. McKoon and Ratcliff (1981) compared decision times for test words that were highly predicted instruments with those that were compatible with prior verbs but not predicted by them. Faster response times were found for the predicted instruments which the researchers interpreted as evidence for the instantiation of those concepts during comprehension. However, this paradigm still has the limitation of being insensitive to the possibility of backward integration.

Sanford and Garrod (1981; Garrod and Sanford, 1983) found evidence for the forward instantiation of verb instruments using a reading-time paradigm. In this paradigm the time taken by subjects to read successive sentences of a text is recorded, subjects controlling the onset of presentation of each sentence. Sanford and Garrod discovered that there was no time disadvantage in reading a sentence containing an instrument that was only implied by a verb in a previous sentence, over the time taken when a previous sentence had the instrument explicitly stated. One of Sanford and Garrod's examples is given in (20) and (21).

(20) *Mary put the baby's clothes on.*
(20') *The clothes* were made of pink wool.*
[Explicit mention]

(21) *Mary dressed the baby.*
(21') *The clothes* were made of pink wool.*
[Implied occurrence]

In apparent conflict with these findings, Singer (1981) found a resolution time disadvantage for implied antecedents. The solution to this anomaly came form the work of Cotter (1984). She replicated both of the Sanford and Garrod and Singer studies using the same materials as the original studies, and obtained the same confounding results for each of the material sets. Her conclusion was that the time differences between the materials reflected real variability in the extent to which different verbs implied objects and instruments as part of their meaning. With the Sanford and Garrod materials, the instruments of the verbs were more central to the meaning of the verb. To return to the example in (20) and (21), for a particular sense of the verb 'dress', clothing entities are inferred as part of the associated schematic representation by being part of what it means to dress. In other usages, for example in dressing a wound or a salad, different instruments and objects will be central to the meaning of the desrcription. Singer's results were derived from materials where the instruments were less central to the verb's meaning, and these do not appear to be inferred as readily. So at least for some verbs used in some contexts there is evidence that co-occuring entities become inferred into the discourse model in some way.

The question arises as to the sort of forward inferential activity that is occuring in such decompositions. One view would be that upon instantiation of the verb 'dress', default values for 'clothes' become inferred and accepted as part of the overall representation, almost as if their existence had been asserted by the writer. Sanford and Garrod's view is rather different. They write,

'...The representation could contain an empty slot indicating where (the clothes) would fit into the whole conceptualisation if mentioned, but no more than this. The slot can be thought of as embodying a series of tests which examine the new text for concepts which could act as slot instantiations; clothing-like entities will all fit the bill, probably with 'clothing' itself being the first thing checked for, but other things, such as specific instances of clothing, being equally acceptable.'

This view would assert that descriptions of instruments subsequent to the verb decomposition would be mapped efficiently into the verb-instrument slot since that slot would already be open and receptive to information consistent with its specification. However, the coherence-preserving advantage of the verb-instrument slots may not extend to more complex inferential resolutions. Sanford and Garrod rightly question the extent of this advantage for implied entities. They write,

> *'...this result raises a further interesting problem. Is it really the case that it makes no difference whether a writer states the antecedent explicitly or merely opts to leave it implied?'* P.104

Indeed, in another experiment Garrod and Sanford (1983) found conditions under which certain instrument descriptions were resolved more slowly to antecedents that were only implied. An example is given below. Sentence (23) is resolved to (22) more quickly than it is to (24).

(22) *Mary put the baby's clothes on.*
(23) *The material was made of pink wool.*

(24) *Mary dressed the baby.*
(23) *The material was made of pink wool.*

The explanation for this time difference must lie in the inferential processes required to identify a relation between the specification of the opened instrument-slot for 'dress', and the consequent description 'material'. Clearly there is a limit to the information related to 'clothing' that can be retained in the relevant slot specification of the 'dress' verb schema. We shall see that this level of analysis is equally relevant to issues emerging in the following section on cohesion-establishment. In attempting to establish cohesion, the processor preserves the co-referential status of descriptions within the text. This in turn will serve to fulfil the higher goal of establishing coherence. Cohesion, then, would appear to be a pre-requisite of coherence, and for expository purposes some work with implications for the current section will be left to the section dealing directly with cohesion.

c: Coherence establishment

We have already observed some work (e.g. Singer and Ferreira, 1983; Potts, Keenan and Golding, 1988) which contrasted measures of inferential activity for coherence establishment and embellishment conditions. Both these sets of researchers presented evidence suggesting that activity necessary for coherence was more likely to occur than activity introducing embellishments. Arguments have been presented on the limitations of the paradigms used.

All the research presented thus far points strongly to the view that establishing coherence is a priority for the component processes of comprehension. Further evidence from Beeman and Gernsbacher (1990) suggests that this is the case. These researchers measured lexical decision times for words presented after the comprehension of a text. They found that faster lexical decisions were made for words that represented concepts which would be inferred to preserve coherence in the text. Further, these lexical decisions were slower when the disrepancy in the world brought about by the references in the text occurred over an episodic boundary. The possible effects of episodic shifts in narrative texts will be assessed in the main body of this thesis.

Other conditions demanding coherence maintenance have been investigated, where the relation described in the text is an abnormal one. Bransford and McCarrell (1975) adopted a false recognition paradigm in studying sentences such as the following in (25), (26) and (27).

(25) *Bill is able to come to the party tonight because his car broke down.*
(26) *John missed the bus because he knew he would have to walk to school.*
(27) *The floor was dirty because Sally used the mop.*

They discovered that subjects were more prone to make false positive recognitions of these sentences than with sentences which included information that justified the abnormal relation. They interpret this finding as supporting the view that in the comprehension of such sentences,

> *'people assume an antecedent condition that explains or justifies the relations between the two phrases.'*
> (Bransford and Johnson, 1973, P.389.)

The following section deals more directly with cohesion establishment in text comprehension.

1.4 Cohesion-establishment

a: Reference resolution

The inferential activity discussed thus far operates at the level of world coherence. A satisfactory degree of text cohesion has to be established in order to maintain and elaborate upon coherent representations of the world. One aspect of text processing that demands cohesion activity concerns the identification of co-referential items within a text. Examples are given in (28) and (29).

(28) *The boy ate his dinner.*

(28') *He was due to go out later.*

(29) *We checked the picnic supplies.*
(29') *The beer was warm.*

In (28), the singular masculine pronoun 'he' is co-referential with 'the boy' introduced in the previous sentence. Hence, the processor has to establish a relation between these descriptions such that in the resultant discourse model they make reference to the same individual. In this example the pronoun is said to be an anaphor of the original noun-phrase. Likewise in (29), a similar resolution is required for cohesion between the definite noun-phrase 'the beer' and the previously mentioned 'picnic supplies'. Anaphoric relations can also hold for indefinite noun-phrase anaphors as in (30).

(30) *John was anxious about the examination.*
(30') *His mother entered the room and found a nervous wreck.*

Such reference resolution clearly requires the involvement of general knowledge. It is through being able to identify a semantic relation between an anaphor and its antecedent that anaphoric resolutions can be established. So the processor has to know that 'he' entails singular and male properties and can then search for those properties in previously asserted references. Similarly, it has to know that 'beer' can form part of the class of things called 'picnic supplies' in order that a co-referential relationship can be established between these objects.

Interest in reference resolution and cohesion-establishment found its source in the linguistic work of Halliday (1967). Halliday emphasised the importance of taking into account an implicit contract that speakers and listeners undertake in their communication. Their utterances should contain signals that differentiate information that is assumed to be mutually known (and hence 'given') from information that the speaker believes to be new to the listener. Although his ideas were couched within the terms of spoken communication, the principles behind the 'Given-New' contract remain valid for written discourse. There should be syntactic markers which signal to a reader that she should take certain information as given, and therefore be able to search for an appropriate antecedent in memory to resolve the present reference. This theory was developed and tested by Haviland and Clark (1974), and segment (29) is taken from one of their examples. It is the definite article in 'the beer' that signals the presence of information already given which will help resolve the reference of the present description. This cue tells the processor to search back in working memory for an appropriate antecedent. Hence, by a process of bridged, backward inferencing the cohesion of the text is established.

On the basis that the construction of backward inferences should be a time-consuming activity, Haviland and Clark compared the time taken to comprehend sentence pairs as in segment (29) where there is an indirect antecedent for the definite noun-phrase, with pairs such as segment (31) where there exists a direct repeat expression.

(31) *We got some beer out the trunk.*
(31') *The beer was warm.*

They found that comprehension times for the direct pairs were reliably faster, and ruled out any explanation due to simple repetition of the noun across the sentence pairs. Comprehension time data from Garrod and Sanford (1977, 1978) appeared to support the Given-New theory of Haviland and Clark. They introduced more specific investigations of the relation between the antecedent and anaphor in such vignettes. Category exemplars rated either high or low in typicality of class membership were used as co-referential noun-phrases in sentence pairs. The exemplar was tested both as an anaphor and as an antecedent to the category description. Examples of low and high conjoint frequency pairs (the typicality relation between any exemplar and its class) are given in (32) and (33).

(32) *A bus came trundling down the hill.*
(32) *A pedestrian was killed by the vehicle.*
[High conjoint frequency pair]

(33) *A tank came trundling round the hill.*
(33') *A pedestrian was killed by the vehicle.*
[Low conjoint frequency pair]

Garrod and Sanford found that the higher the conjoint frequency, the shorter the comprehension time would be. This result was irrespective of which description fulfilled the antecedent role, and is highly consistent with the truth verification times received for class membership statements (Wilkins, 1971; Rosch, 1973). The result is also consistent with the view that in resolving anaphoric references in texts, backward searches for previously introduced entities are carried out, and these can be facilitated, or otherwise, depending upon the accessibility of the relations exisiting between the antecedent and consequent descriptions.

Although this result is not inconsistent with the Given-New formulation, Sanford and Garrod (1981) posited that its weakness is in under-representing the inferential activity found to occur in cohesion-establishment. Garrod and Sanford tested similar sentence pairs to (32) and (33), this time with a non co-referential, but semantically related, noun-phrase in the second sentence. An example is given in (34).

(34) *A tank/bus came trundling down the hill.*
(34') *It nearly smashed into a vehicle.*

The Given-New hypothesis would predict that there would be no effect due to conjoint frequency found in the comprehension times for these sentence pairs. This is because there is no definite article in the second sentence to cue a search back for previously mentioned, semantically related entities. In fact, Garrod and Sanford found a small effect due to conjoint frequency for these materials. This suggests that semantic relations can be (even partially) computed between descriptions even where no cohesion-establishing procedure (predicted by Given-New) is forcing such a test. Such a finding points to a significant degree of automaticity and autonomy in the establishment of reference resolution. As Sanford (1990b) writes

'The result thus suggests an autonomous matching process based upon pre-existing and text-independent semantic relations between tank/bus and vehicle.'

These observations support the view that the establishment of reference resolution is a 'hard' constraint in the processing system. Further, such resolution appears to occur quite rapidly as an anaphor is read, as suggested in the processing framework of Sanford and Garrod (1981).

The same authors, Garrod and Sanford (1985), found further supporting evidence for this view when employing a spelling error detection task on various materials. Verbs with spelling mistakes were detected faster when they were appropriate to the immediate prior context. This context was constituted by an anaphoric reference to a previously mentioned character. Text-based knowledge of the character effected the appropriateness of the verbs that were to follow. The enhanced latency effect must have derived from the force of the anaphor to be resolved quickly to the previous mention, thus shaping the plausibility of action descriptions for the individual. Because the mis-spelled verbs occurred directly after the anaphors in the text, the resolution and shaping activity must have occurred rapidly.

Some of the material already introduced concerning coherence activity has presupposed successful cohesion establishment. We observed in sentence pairs (20) and (21) a contrast between the implication of a verb's object and its explicit mention, and this was introduced in the discussion of the embellishment of verb instruments.

(20) *Mary put the baby's clothes on.*
(20') *The clothes were made of pink wool.*
[Explicit mention]

(21) *Mary dressed the baby.*
(21') *The clothes were made of pink wool.*
[Implied occurence]

Discussion of the work of Sanford and Garrod (1980, 1981), Singer (1981), and Cotter (1984) revealed that reading times for sentences containing implied objects would be of comparable latency with sentences containing repeat expressions, providing that the object was part of the meaning of the verb. This result is also of relevance to cohesion activity, since it involves anaphoric reference resolution. To this point reference resolution and bridging inferences have been described in terms of backward activity, i.e. in terms of searches that are instigated in the event of a description believed to be referring to a previously mentioned object. Here the findings suggest that forward activity can occur in support of subsequent cohesion demands.

Recognising the need for discernment of the inferential status of explicitly and implicitly introduced entities, Sanford and Garrod (1981) developed a framework for a theory of text comprehension that is underpinned by their notion of focus. In general, focus mechanisms exercise control over the attention that is paid to different parts of a text or the memory representation arising from it. A good example from their work concerns a distinction between explicit and implicit focus partitions of an agent's memory. Explicit focus is a memory trace of the things that have been explicitly introduced in the text, while implicit focus contains objects only implied by the text. Sanford and Garrod claim that pronominalised anaphoric relations can only have antecedents that are objects in explicit focus, i.e. that have been mentioned explicitly in the text. This can be illustrated by pronominalising the definite noun-phrase in sentence (20') (becoming (34)) following either sentence (20) or (21), with the following effect.

(20) *Mary put the baby's clothes on.*
(34) *They were made of pink wool.*

(21) *Mary dressed the baby.*
(34) *They were made of pink wool.*

Although the second of these vignettes is a possible discourse construction, it does not display a normal, felicitous pronoun usage, and is difficult to comprehend. This is reflected in reliably longer reading times for this type of construction (c.f. Garrod and Sanford, 1983). A backward inference has to be constructed to bring 'clothing entities' more into the foreground. Pronouns, then, generally have antecedents that have been explicitly mentioned in the text, and searches for these antecedents are carried out in the explicit focus part of the memory system.

b: Reference-to-situation mappings

We can cite some other evidence from Sanford and Garrod (1981) which gives an indication of the relationship between the knowledge-structures used in linguistic interpretation and contents of implicit focus. This work returns to the idea of the facilitating effect of informative titles. Sanford and Garrod's framework posits that incoming descriptions are mapped onto knowledge-structures or 'scenarios' where they are understood by fulfilling roles within the structure. These scenarios contain information specific to situations suggested by a theme or title. Part of the problem of comprehension involves selecting an appropriate scenario to act as a background structure. Titles are one means by which this can be achieved, and as a consequence it is possible to empirically test the effectiveness of titles by using the reading time paradigm. An example of such an experiment is taken from Sanford and Garrod (1981).

Appropriate scenario

Title: **In court**

Fred was being questioned *(by a lawyer)*.
He had been accused of murder.
The lawyer was trying to prove his innocence.

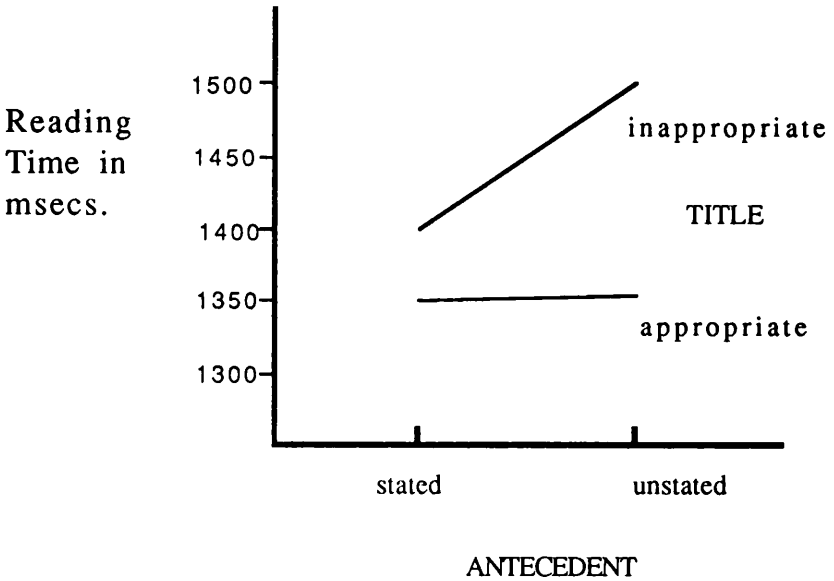
Inappropriate scenario

Title: **Telling a lie**

Fred was being questioned *(by a lawyer)*.
He couldn't tell the truth.
The lawyer was trying to prove his innocence.

The rationale behind this experiment is that with the selection of an appropriate scenario about courtroom scenes, references to lawyers should be easily interpretable since they are part of what is expected in such situations. With an inappropriate scenario (in this case to do with telling lies) references to lawyers will be interpretable and consistent with the overall theme, but will not be predicted. References to 'the lawyer' in the final sentence take significantly longer than in the appropriate scenario condition where 'the lawyer' has no explicit antecedent in the text. A graph of the final sentence reading times is given in Figure 1(a).

FIGURE 1(a): Reading times for target sentences which refer to either a stated or unstated antecedent, following appropriate or inappropriate titles. (Adapted from Sanford and Garrod, 1981)



These results support the view that information implied by prior descriptions becomes available through the scenario selection process. Things that are referred to can be resolved more easily because they are already expected and have been given priveleged status in working memory. It is in this sense that the background scenario becomes part of the implicit focus partition during comprehension. Sanford and Garrod describe this process of focusing on relevant information in implicit memory as extending the domain of reference. The set of implied individuals, roles and objects contained within a scenario constitute the extended domain of reference, and it is in this sense that these objects come into focus when the scenario is active.

We have then examples of conditions of interpretation where there appears to be forward activity that facilitates subsequent reference resolution and scenario mappings. Evidence for this sort of activity has been found at the level of verb decomposition and at a more general level in relation to implied objects in selected interpretive knowledge-structures. These findings raise a breadth of issues for understanding inferential mechanisms. It is not clear what sort of inferential status we can give to implied objects. Are all implications inferences? Nor is it clear what effect forward facilitation has on the comprehension machinery, despite the fact that we are able to observe consequences of these effects in reading time measures.

It is difficult, and undesirable, to present the topic of cohesion without encroaching on the ground of coherence activity, and vica versa. This is no accident. What it means for the processor to take a linguistic input and find a way of fitting its components together in a cohesive way, is intimately related to establishing a world-coherent representation. This parallel again emerges with the cohesion-demanding properties of verbs.

c: Verb Schemata

Verbs seem particularily important in the establishment of cohesive representations. As well as giving information about actions and events they implicate a rich source of other things. We have already devoted a section to verb-instrument inference in 1.3(b), and much of its content was relevant to the sub-section on reference resolution in this section. As we recognised that the plausible instrument of a verb could be inferred, so other 'slots' or 'cases' can become implied by an action description. For example, as an _instrument_ slot for the verb 'pound' can be taken by 'hammer', so an _agent_ slot (who did the pounding?), or an _object_ slot (who or what was pounded?) can be established. Similarly, other types of information can be slotted in and be viewed as satsifying a case-specification of the verb's structure (c.f. Fillmore

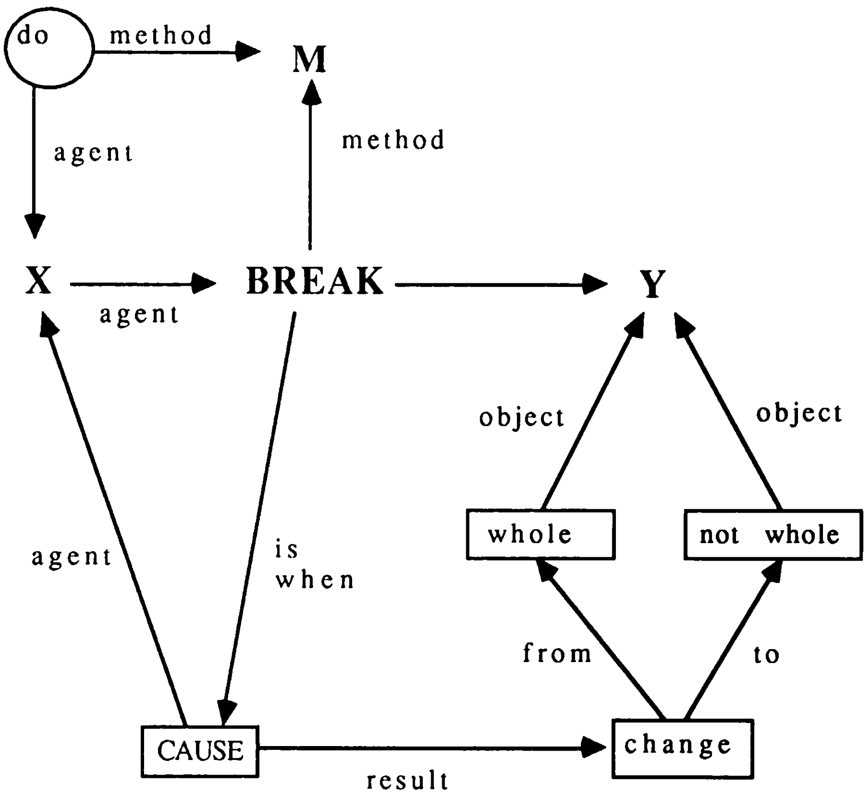
1968). A selection of other possible slots is adapted from Sanford and Garrod (1981), presented in Table 1.1.

TABLE 1.1: Possible case assignments to an action description. (Adapted from Sanford and Garrod, 1981)

Location:	The place where the event takes place. Often two different locations are involved, one at the start of the event and one at the conclusion. These are identified as *to* and *from* locations.
Purpose:	Identified the purpose of the event: e.g. Simon went to the river *to catch fish*.
Quality:	A descriptor, one that modifies a concept: e.g. The budgerigar was *yellow*.
Recipient:	The person who is the receiver of the effect of the action on the object. e.g. Mary cooked the dinner for *Simon*.
Time:	When an event takes place: e.g. The salmon were jumping *yesterday*.

The background scenario or schema acting as a context for a verb will provide the source for plausible slot fillers. This way of thinking about how information from a text will be fitted together led to the expanded idea of a verb-schema (Rumelhart and Ortony, 1977). Here the schematicisation of a verb's meaning necessarily implicates its relation to the sorts of things that fill its slots. These are effectively semantic relations, and a possible schema for the verb 'break' is given in Figure 1(b).

FIGURE 1(b): A possible schematicisation for the verb 'break'. (Adapted from Rumelhart and Ortony, 1977)



In this shematicisation the lexical item 'break' accesses a network of relations composed of the case specifications (X, Y, M) and particular configurations of primitives (e.g. CAUSE, DO). The case or slot variables can take values constrained by the nature of the primitives (e.g. the object, Y, has to be brittle). This 'break' schema will be accessed by any form of the verb (e.g. first person singular etc.), and the form used together with the rest of the message will provide the information for establishing a cohesive set of mappings. Some examples are given below.

(35) 'I broke the window'

Agent(s): 'I'
Object (brittle): the window
Method: unknown
Instrument: unknown

(36) The tile was broken with the heavy hammer by John and Mary.

Agent(s): John and Mary
Object (brittle): the tile
Method: unknown (plausible inference: hammering)
Instrument: the heavy hammer

Under this perspective, cohesion establishment demands finding modifying noun-phrases (or prepositional phrases and adjectival phrases) which satisfy the primitive specifications of the verb. This process is not, however, remote from that of filtering the input through elicited background knowledge. The evidence points to the view that contextual information is available and active in support of such cohesion establishment (e.g. the results of Sanford and Garrod (1981), Figure 1(a)). How these active influences interact is unclear, as are the constraints which operate on the time-course of much of this activity. For example, if verbs appear to be such cornerstones of maintaining cohesion, what effect on processing activity is acheived by varying the point at which the verb appears in the sentence? Is the significance of other information worked out prior to the onset of the verb, or is such processing so dependent on action-based information that it is put on hold until the appropriate verb-schema is accessed? Perhaps neither extreme is the norm, and if so what would it mean for partial information processing of a verb-preceeding modifier to take place? Some of these issues surrounding partial activity are developed in the final section of this chapter.

1.5 Preliminary conclusions: levels of inferential activity

Over the last three sections we have looked at various research techniques applied to various types of inferential activity. Throughout there have been themes common to all these types (e.g. the importance of background knowledge, inferential measurement sensitivity etc.). Although typologies of the sort presented are necessary to simplify complex research problems, it is clear that cohesion, coherence and embellishment activity are all highly inter-dependent processes. At a functional level, they each perform a different kind of elaboration, but in no sense is it to be imagined that there are distinct modules of the language processor producing three disparate representations from the same message.

The crudest taxonomy available distinguishes between embellishment inference as it is classically viewed (see section 1.3(b)), and cohesion/coherence establishment. The focus of this thesis is on the control factors effecting activity in the latter. However, this does not render irrelevant much of what is imagined to happen in embellishment inference. For example, we have already seen that embellishments can become necessary for coherence purposes depending on the content of subsequent parts of a message. And although some researchers argue for the non-existence of forward activity (e.g. Corbett and Dosher, 1978), we have seen at both the verb decomposition level (sections 1.3(b) and 1.4(a)) and the scenario-mapping level (section 1.4(b)) the shaping of implicit information made accessible for anticipated comprehension demands. Of course, this accessibility-shaping is not part of what is generally thought to constitute an embellishment inference, but it must be a real part of forward, elaborative inferential activity. Even if this activity only amounts to the setting up of pointers to general knowledge that is likely to be relevant, it is processing as basic as this the control relations of which little is known.

Perhaps this reflects an emergent investigative strategy in discourse science to detect inferences assumed to be clean and easily describable propositions (e.g. Kintsch; 1974, 1988), rather than to detect inferential activity which will probably be messy (even statistical) and more cumbersome to describe adequately. The essentially parallel connectionist architectures which have emerged in the 1980's may lend themselves more to such types of representation than do classical von Neuman forms of computation. An assessment of this possibility will be made in Chapter 2.

Much of the speculation surrounding accessibility-shaping and partial forms of inference presupposes more than one level of representation within the mind of a reader. Since we imagine that inferences can be drawn in the classical 'inference-engine' fashion (c.f. Minsky, 1988), partial inference (whatever that turns out to be) must be a dumber kind of animal that doesn't represent the imagined world in such a complete way. Or at the very least it is a dumber kind of process which supports the derivation of conclusions at a later stage. Arguments over the number of levels that exist in a representation are almost destined to become meaningless, since there will be an infinite (or finite but very large) number of levels of description for any embodied process (from partical physics upwards). The following taxonomy of the sorts of levels of activity that could be thought to obtain in text comprehension should be viewed with these comments in mind.

The four levels of inferential activity to be described are intended to be conveyed as aspects or perspectives on what may turn out to be a single discourse model (or mental model; c.f. Johnson-Laird, 1983) that is established during comprehension. As such, no strong claims are being made about the status of the levels. Rather, the argument is that the distinct types of information in each level should be derivable from a model of a text, and at the current stage of enquiry into these topics it could be of value to assert these differences.

1: Semantic Model:

(c.f. Miller, 1979) This level holds everything that has not been ruled out by what has already been stated. This will be describable as an infinite set of propositions, and these propositions will be the set of possible inferences that could elaborate on what has been given. These possible inferences will not make a logically consistent set (since both 'p' and 'not p' could be true given that neither has been previously asserted).

2: Plausibility Filter:

The possibilities derivable from the semantic model will not all be equally likely to be true given the information provided in the text. Hence the need for a level of representation that can derive the plausibility of any member of the semantic model in co-occurance with a text. It seems to be a fair assumption that only plausible possibilities will become inferences. Further, if 'p' becomes highly plausible then 'not p' will necessarily become implausible (though technically still possible). Hence, a degree of logical consistency seems to fold out of the plausibility evaluation process, though exactly how this would be incorporated in a process model is unclear.

3: Thematic Focus and Inferential Fields:

Psychological investigation has tended to work under the assumption that plausibility relations are the predominating control factor of inferential activity (e.g. McKoon and Ratcliff, 1986). However, it is possible that highly plausible sets of inferences will not be drawn because they are not central enough to the topic or theme that the writer has been developing (assuming a well-written text). Hence the need to remain open to the notion that focus-based controls can exercise veto on potential inferences with 'established' high levels of plausibility. Perhaps discourse science needs to develop the idea of an **inferential field** to capture the importance of maintaining relevance in elaboration. Some preliminary work suggests that this may be a fruitful direction to follow (Walker and Meyer, 1980), and the case study presented in the main body of this thesis will return to this issue.

4: Discourse Model:

This is the final set of tokens representing the world, derived directly from the text-content together with embellishment inferences. At this level of coherence the representation must be logically consistent (i.e. cannot simultaneously hold the belief that 'p' and 'not p'). The embellishment inferences are presumably expressable as propositions.

It is not meant to be implied that in the construction of an inference a whole set of possibilities in the semantic model are systematically placed through each inferential level. This would neither be practical nor desirable. Rather the focus-control devices employed by a writer should effectively constrain the attention of the processor to the topics being emphasised at all levels of comprehension activity. So what the reader is thinking about, the inferential field she is exploring, is 'moved around' by the emerging content of the text and significance of the message.

Although the four levels are given principally for the purposes of exposition, they do shed light on three crucial problems. First, what is the status of an inference at the level of the discourse model? If it commands the same token representation as explicit information, and is expressable in propositional form, then what sort of record is there that it is an inference at all? It could be that in many cases there will be no record, or that the record of what has been inferred and what has been explicitly presented will be lost over time. (There is no reason to suppose, of course, that individuals will not differ considerably in their inferential skill, and so effective reading may demand a level of ability in remembering the explicit from the implicit (c.f. Gernsbacher, Varner and Faust, 1990)). In general, inferred information appears to be more easily revisable, suggesting that inferred parts of a coherent representation are 'hedged' in some way (c.f. Lakoff, 1972). Perhaps implicit information is represented with an attenuated strength of belief.

This leads into the second problem. If some information, encoded in a discourse model, and expressable as a proposition, is given a lower 'belief-status' than explicit information, then why should such information be thought of as being part of the discourse model in the first place? Surely it becomes equivelant to all the other focussed and plausible possibilities derivable at the 'inferential fields' level? Is it possible, then, that the idea of a discourse model of this kind can be effectively ruled out? Perhaps all that happens in standard text processing is the establishment of the focus-sensitive inferential fields, and information can be drawn out of that source when it is deemed necessary for particular purposes (for example see 1.3(b) for comments on the inference generation procedure). Whatever the right answer turns out to be, commencing an inference detection study under the assumption that it is extra propositions that are to be detected is clearly a limiting research framework.

The third problem presents itself. If the search for inference is not restricted to elaboration at the proposition level, then what detection techniques are going to be sensitive to partial or minimal inferences, and how are these to be characterised? Further, what sort of process model will be flexible enough to capture degrees of completeness in partially established inferential activity, and to what extent would this framework need to revise long-held assumptions about the atomic nature of cognitive operations?

Chapter Two

An Assessment of Contemporary Models of Language Processing in View of Incompleteness in the Machinery of Cohesion-Establishment

2.1 Introduction

Chapter 1 introduced prominent issues that have arisen in the field of discourse inference. The preliminary conclusions regarding these issues point towards the need for a better understanding of partial or minimal inferential activity. The necessary development is due to empirical research largely involving embellishment-type inferences. The present chapter introduces phenomena concerning cohesion-type inferential activity, which suggest that partial inferencing is fundamental here too. With this in mind, an assessment is made of the contribution present models of language processing can make to such an endeavour. The discussion contrasts the potential contributions of the classical (symbolic) computational linguistic framework, present research in (sub-symbolic) connectionist architectures, as well as relevant insights from experimental cognitive psychology.

2.2 Incompleteness in cohesion-establishment: failures to notice anomalies in written discourse

The preliminary conclusions made in Chapter 1 hinged upon the view that a better understanding of partial inferential activity is necessary for richer models of inference in language comprehension. This view emerged from observations concerning properties of inference, largely of the embellishment type. This section aims to introduce a number of phenomena which strongly support the view that such partial or minimal activity must be fundamental to cohesion-establishing mechanisms also. Consider the following examples.

(1) This book fills a much needed gap.

(2) How many animals of each kind did Moses take on the ark?

(3) Can a man marry his widow's sister?

(4) How much soil is there in a hole that is two metres deep, six metres long and three metres wide?

(5) When an aircraft crashes where should the survivors be buried?

The interesting property shared by these sentences is that they are often interpreted in a way which renders the resultant discourse model _inconsistent_ with the written message. The common interpretation of (1) is that the book in question was much needed in order to fill a gap, but the written message does not give this information. In fact, the qualification 'much needed' modifies the gap, not the book! The question (2) has been researched by Erickson and Mattson (1981). Across various manipulations, they discovered a surprisingly large number of readers who failed to notice the discrepancy between their narrative-specific knowledge concerning _Noah's_ ark, and the _Moses_ character included in the question. This went unnoticed despite the fact that subjects were warned that some of the questions presented may be anomalous.

Sentence (3) is similar to (1) since a property asserted of one object or individual is inappropriately bound to another in the text representation. The anomaly in this case is that if a man has a widow then he is already dead! The dominant interpretation is that it is his wife who has died, thus making relevant the mention of the wife's sister. If anything, the mention of the sister evokes knowledge of kinship/marital constraints which seem to lend the processor towards establishing an innappropriate interpretation. Similarly in (4) the inclusion of the depth, length and width parameters seems to direct attention away from the basic anomalous 'stem' ('How much soil is there in a hole?'). Of course, holes are by their very nature empty and therefore could not contain soil. The dominant interpretation seems more consistent with the question 'How much soil could you get in a hole....?' Sentence (5) is also intriguing since often readers fail to notice that the question asks about burying survivors and not about burying dead people!

This selection of examples indicates that readers can be tricked into making mis-interpretations of messages that are deliberately constructed in order that this may happen. They show that the human language processor can be configured to make mis-mappings from an input string to a background model. Hence, under some conditions, incompleteness can occur in the processes underlying mapping which support cohesion. Whatever the developments in successive models of language processing, mis-interpretation and incompleteness from infelicitous input is something that will have to be accounted for.

It would be folly to dismiss these phenomena as mere tricks that are of no relevance to normal communication. It is the author's view (see also Erickson and Mattson, 1981; St.John and McClelland, 1990) that such tricks can reveal partial inferential activity which is just as inherent in consistent and normal reading, but which does not seem to manifest other than in error production. In observing the processor at fault, some aspects of its standard activity may appear. The underlying assumption is this. There is no reason to suppose that separate forms of processing will be carried out on felicitous and infelicitous input, since the above examples indicate that there are not always exhaustive checks made on whether a message is felicitously constructed in the first instance. So the incomplete or partial activity evident from mis-interpretation will be as fundamental to the establishment of appropriate mappings. This argument will be expanded at a later point.

The phenomena in question also remind us that message construction and interpretation can be a far messier business than logically motivated models have suggested. Indeed, there is every reason to suppose that anomalies and contradictions occur quite frequently in dialogue and discourse. Recent compendiums of the so-called 'Colemanballs' support this view. These list some of the anomalous, ambiguous and meaningless things (usually well-known) people have said. The statements are often highly amusing (an interesting issue in its own right). A selection is included in Table 2.1 under five categories which are intended to be informative for the present discussion. The categories are not mutually exclusive or absolute.

TABLE 2.1: A selection of statements taken from the 'Colemanballs' compendium (1984, 1986, 1988, 1990).

1: Insufficient relevance or informativeness:

'Oh and that's a brilliant shot. The odd thing is his mum's not very keen on snooker.'
'The late start is due to the time.'

2: Mixed metaphor:

'Its obvious these Russian swimmers are determined to do well on American soil.'
'Celtic manager David Hay still has a fresh pair of legs up his sleeve.'
'...and she finally tastes the sweet smell of success.'
'Then there was that dark horse with the golden arm, Mudassar Nazar.'
'He's been burning the midnight oil at both ends.'

3: Providing obvious or known information:

'He's 31 this year. Last year he was 30.'
'I'm a firm believer that if you score one goal the other team have to score two to win.'

4: Implication of strange prior conditions or expectations:

'...and somewhat surprisingly Cambridge have won the toss.'
'Hodge scored for Forest after only 22 seconds, totally against the run of play.'
'Attached to the bottom corner is a rope of finite length.'

5: Contradictory parts of the same message:

'I can only see it going one way, that's my way. How it's actually going to go I can't really say.'
'...and Magri has to do well against this unknown Mexican who comes from a famous family of five boxing brothers.'
'I'm not going to make it a target but it is something to aim for.'
'Even when you're dead you shouldn't let yourself lie down and be buried.'
'Tombay's hopes, which were nil before, are absolutely zero now.'

Although these errors concern language production, they are obviously related to comprehension, in that the errors are not always obvious. The mapping processes mediating between messages at the level of thought and their manifestation in language appear to be susceptible to fuzziness or bias in some way. With the 'Colemanballs' examples one can often recognise what was intended to be communicated, despite the ambiguous or sloppy construction of the speech act (Searle, 1969). *The point* can can get through. It may be that the mirror image of this observation is manifest in the phenomena of sentences (1) to (5). Here, anticipation or expectation of the significance of the message seems to have an effect on the (on-line?) constraints satisfaction of cohesion-establishment. The desire to establish a meaningful model seems to overide a systematic and disciplined analysis of the input signal. There are two other related examples of this 'contextual' influence on completeness in the psycholinguistic literature.

The first (Ehrlich and Loridant, 1990), concerns the resolution of anaphoric relations. In a text written in the French language about economic resources in Canada, reference was made to some of the effects of <u>wealth</u>, and a subsequent (anomalous) anaphoric reference used the term <u>poverty</u>. The great majority of the subjects failed to detect this, and had interpreted poverty as standing for wealth. In this instance, the constraints imposed by the context seem to promote reference expectations which can be satisfied by a minimal quota of evidence.

The second interesting example (Wason and Reich, 1979), concerns enigmatic sentences such as the following.

(6) No head injury is too trivial to be ignored.

The dominant interpretation of this sentence is that no matter how trivial a head injury is it should not be ignored. In fact, this sentence is of the same form as (7).

(7) No missile is too small to be banned.

The dominant and correct interpretation here is that *however small a missile it should still be banned*. Following through the same logic for (6), the conclusion should be that however trivial a head injury, it ought to be ignored! The same semantic form exists for the two sentences (No WUG is too DAX to be ZONGED). So one would expect that a processor applying structure-sensitive operations would detect the non-pragmatic implicature in (6), by the same method that it would reliably interpret the pragmatic implicature in (7). In this case the pragmatic status of the message seems to interact in a complex way with the large number of negatives in the sentence. **The result suggests that some kind of contextually-driven expectation of what the message will be has a biasing effect on the type of cohesion that is satisfied for it.**

From these findings, cohesion patterns seem to be strongly influenced by efforts to resolve the significance of the whole message. The influence of relevant background knowledge seems to make available sets of expectations to this end. Exactly what these expectations do is not clear, nor is it clear how particular inputs are tested against such expectations, or how rigidly they are applied. Similarly, there is no available data measuring the time-course of such expectation activity.

Since much of psycholinguistic thought has been positively influenced by computational linguistics (and the symbolic processing model of mind in general), a close inspection of such classical approaches seems neccessary to investigate how they can handle the phenomena being discussed. This will be undertaken in the next section.

The few papers which have emerged on the topic of incompleteness have served the purpose of drawing attention to the existence of such phenomena (Wason and Reich, 1979; Erickson and Mattson, 1981; Ehrlich and Loridant, 1990). However, these phenomena have not undergone sufficiently systematic and rigorous study. Although many researchers are aware that these sorts of anomalies can occur, the prevailing view appears to have been that they are sufficiently deviant from existing models for them to be ignored. Of course, no anomaly is too trivial to be attended to. One also wonders the extent to which our intuitions about the prevalence of such phenomena are biased by the inconsistency that we manage to detect?

2.3 Properties associated with symbolic implementations of language processing

Symbolic views of language treat lexical items as constituents which can combine to form propositions that, in a discourse, bear logical relations to one another. Syntactically and semantically well-formed items (or atoms) can form sentences which are interpretable by our language processing facility as descriptions of the world. Such atomic models probably date back (at latest) to the Cartesian revolution in sixteenth century philosophy, and under-pin Doctor Johnston's pioneering dictionary compilation in the seventeenth century. In the late nineteenth century, there emerged great philosophical interest in developing logical formalisms for natural language (Frege, 1892), and this gained increasing impetus in this century with the advent of Turing-based computational devices. Indeed, such work was influential in shaping the form of that adventure. Present day linguists and semanticists continue to develop formalisms which attempt to capture the richness of meaning manifest in natural language (e.g. Montague, 1970). The past twenty years has seen an increasing intellectual investment in making such formalisms computationally tractable for Von Neuman machines (the modern digital successors to the Turing machine). Such activity constitutes the domain of computational linguistics. As we shall observe, the whole formalist drive has played an influential role in cognitive psychology, as it now does in the multi-disciplinary cognitive science.

The basic framework of symbolic processing is well presented by Fodor and Pylyshyn (1988). Under their view, semantic content is assigned to expressions which have a combinatorial syntax and semantics. These expressions are describable as propositions and constitute the basis for a 'language of thought' (Fodor, 1975). Because of their atomic form, 'structure-sensitive' operations can be applied to them. These are rules which operate with respect to the form of an expression and without recourse to the content of the atoms in question. This framework has consequences for the sort of cohesion that is established from a linguistic input, and also for the type of processing machinery that effects such establishment. Of course, these are not unrelated issues.

In this section, the main interest is concerned with how such a framework can account for the incompleteness phenomena that have been introduced. Prior to this assessment, more detail is presented concerning the cohesion and machinery-type issues. Working examples of natural language parsing systems will illustrate some of the themes that are introduced.

a: Atomic cohesion-establishment

How does an atomic processor try to establish cohesion? The principal strategy in a computational linguistic system is to design a program which specifies rules for combining linguistic entities together. This combination has both a syntactic and semantic aspect. The job of the program (or parser) is to try and prove that a given input-string (sentence) is satisfied by the procedures specified in the program, and that it therefore constitutes an acceptable and meaningful expression of the language. In practice, the control of most parsers lies with the syntactic component. Clauses that satisfy specified syntactic configurations are then subject to the selection restrictions of semantic procedures.

Such systems assume compositionality and combinability constraints. Compositionality can be described as follows. Words have definable meanings that when bound together in acceptable formulations result in definable sentence representations. The semantic contribution of any particular word is the same across contexts, and any variation in a word's impact is a reflection of the process of combination with other words. This compositional view will be assessed more closely later. Assuming compositionality licenses the application of combinability restrictions. As mentioned, the semantic component of a parser effects selection restrictions on permissable combinations of linguistic objects. These restrictions work by checking the semantic relations between entities on the basis of features attributed to them. This attribution is made in a pre-specification of semantic features for all lexical items interpretable by the parser.

So the basic format for a program is this. A 'dictionary' component has a list of lexical entries which are defined with respect to the syntactic category they occupy (noun, verb, adjective), together with the stipulated semantic features for each item. Grammatical rules are encoded in order to build syntactic trees from the constituents attributed to the lexical items. A semantic feature-matching process refines the capability of the parser to prevent the acceptance of input strings which are meaningless, contradictory or anomalous. An example of a parsing system which works under these principles is Bobrow and Webber's (1980) PSI-KLONE parser.

(8) David ran the half-marathon.

PSI-KLONE's syntactic processor could propose to the semantic component that the noun 'David', in subject position in sentence (8), could be bound to the verb 'ran' as part of its schematic requirements. The semantic processor tests for the presence of appropriate feature-relations between these entities. The pre-defined verb-schema for 'ran' has slots which have to be occupied by appropriate fillers. In this case, the agent slot of the schema will take a pre-specified 'animate' feature. The slot can be satisfied by the lexical entry 'David', also pre-defined with an 'animate' semantic feature. We can see that this is a computational extension of the verb-schema idea presented in section 1.4(c). The cohesion for the string is established through checking that relations between lexical entries are properly matched. Importantly, these relations are never partially computed or based upon minimal evidence. Cohesive structures result from the unbending application of procedures which test for specified conditions. Procedures which fail to establish an appropriate mapping simply return a message to that effect to the syntactic component, and it then attempts a different parse on the clause in question. Items either fulfil the specified requirements or do not. There is no room for the development of weak relationships between semantically related items. The resulting coherent structures are established as a set of logically consistent bindings.

In practice there are a number of important design issues for parsing systems. Some of these, such as programming language implemetation issues, are not directly relevant to this discussion. Some of those that are relevant are introduced in the following sub-sections.

b: Syntactic and semantic processing interaction

Because natural language parsing systems have separate syntactic and semantic components, procedures determining the control relations of their interaction have to be specified. We have already observed that control tends to lie with the syntactic component which sends either partial or complete syntactic configurations for assessment by the semantic component. The main issue, then, is determining under what conditions the semantic component is brought into play. If semantic evaluation of a noun-phrase, for example, occurs only when that structure has been syntactically closed by the parser, then such a strategy is unrealistic. Since phrase-structure rules for noun-phrase constitution permit arbitrarily long input strings, semantically unacceptable clauses may only be discovered after time-consuming parses have been completed. This form of reference evaluation, which is utilised in Winograd's (1972) SHRDLU program, seems to be workable in well-defined domains, but is in fundamental opposition to desirable semantic evaluation strategies. Optimal processing procedures should accumulate semantic information as a sentence is read form left to right, since by this method unacceptable strings are ruled out at an early stage. Such semantic interpretation is termed incremental.

There have been a variety of strategies developed in order to maximise the incremental potential of parsers. Mellish (1981, 1982, 1983, 1985) utilises a principle of Incremental Description Refinement (IDR). Under this system, semantic operations are invoked when a major syntactic constituent is established. This, however, does not directly correspond to a left-to-right sentence scan. For example, verb-based information is only utilised when all the potential slot fillers for the verb have been established. Hence, subject position noun-phrases can only be mapped into an adjacent verb structure when subsequent fillers for the verb have been found later in the string. Processing the arguments for constituent structures prior to the structure itself means that the semantic interpretation of such a constituent is not immediately related to the existing sentence representation. As Haddock (1989) writes,

'...while the progressive steps of the semantic interpreter constitute incremental description refinement, these steps are not in a one-to-one correspondence with the steps of the processor which reads the input string from left to right.'

Another incremental strategy has been to develop grammatical formulations which produce left branching phrase-structure trees (Altmann and Steedman, 1988). Unlike traditional grammars for English, these so-called 'combinatory' grammars form new syntactic constituents as each word is read, thereby allowing a closer correspondence between the time-course of syntactic and semantic evaluation.

The issue of the interaction between syntactic and semantic processing is important for the present discussion. The apparent impact of contextually-driven expectations which obtain in the interpretation of sentences (1) to (5), suggests that the time-course of involvement of different aspects of background knowledge is a crucial factor in understanding the sort of cohesion that is satisfied for such sentences.

We have seen that symbolic parsing systems instantiate feature-matching procedures and bind variables to predicates. These methods do not sound comparable to the psychological process models introduced in Chapter 1. In fact, the underlying rationale is similar. The activity that supports the establishment of atomic relations in a parser can be properly viewed as inferential activity, since any structure built out of basic item-strings is an elaboration on the the given input. In the psychological review, support for inferential activity was found to derive from the input of relevant background knowledge. Similarly, a parser's general knowledge resides in the case specifications for its verbs and in the semantic features for lexical items. Establishing semantic relations between items is analogous to the semantic relations the human processor licenses. This is driven by general knowledge about plausible relationships in the world. Of course, opinions differ on whether parsing systems exhibit genuine comprehension, or merely simulate such activity. Under either view, operations that can be proceduralised in symbolic computation provide a valuable source of information about the tractability of relational activity in language processing.

For example, differences in parser's performance capabilities can be compared with the form of procedures that they embody. A relevant instance is that systems vary in the extent to which they can access various levels of contextual information. The PSI-KLONE system does not have access to knowledge about discourse-specific relations such as a record of entities that have been previously referred to. Other systems, like Winograd's (1972) SHRDLU, take into account such local context issues when resolving references. Psychological evidence (e.g. Crain and Steedman, 1985) suggests that under conditions of ambiguity, the human parser tends to make interpretations consistent with factors at the discourse level, even if, other things being equal, text-independent knowledge would promote stronger expectations for a different resolution.

Perhaps the implemetation issue of most relevance to the present discussion concerns the way in which the meaning associated with lexical items is envisaged and defined. This in turn effects the type of contribution a word can make to an interpretation of any given message.

c: Word meaning and compositionality

The fundamental assumption behind atomic approaches to word meaning is that words map onto stable underlying objects. Further, these objects can be defined and described. Different approaches have been taken in attempting to make such descriptions. Carnap (1956) attempted a formalism using 'meaning postulates' which define words with respect to other implied lexical items combined by logical operators (e.g. if x is a 'girl' then x is 'female'). Other approaches have constituted word meaning in terms of semantic components which are again combined by logical constants, but are in themselves theoretical elements rather than implied lexical entries (Katz and Fodor, 1963; Weinreich, 1966). Such has been the rationale behind the lexical specifications in the parsing systems that have been mentioned.

Importantly, computational linguistic programs set definitions for lexical entries in terms of <u>discrete</u> lists of features or components (e.g. Pereira and Shieber, 1987). Feature-matched structures are complete only in so far as a specified set of deductions can be made from them. This is what is meant by having a logically consistent set of bindings. Notice that it is through lexical pre-specification that symbolic accounts can generate valid inferences for combinations of those items. This is an established tradition in logical formalisms of language.

One wonders whether the converse view will become more favoured. Could it be that capturing inference patterns is the key to defining parameters of word meaning? For example, Moxey and Sanford (1987), and Moxey, Sanford and Barton

(1990) have discovered richer (and unexpected) models of meaning for non-logical quantifiers as a result of investigating inferential patterns for these words. Existing formal definitions have been found to under-represent the full meaning of the items. Their findings do not appear to be compatible with discrete-definition approaches to word meaning, and an example will be described in the following sub-section. Of course, such early endeavour does not sink the atomic view. It merely indicates a successful change of emphasis in one domain that may prove to be of value in others.

Unquestionably, the dictionary approach to word meaning captures some of the important relations signalled by lexical entries. There is the danger, however, that our introspections about word meanings may only afford us the more stable aspects of a word's potential. Our own language about meaning is very revealing. The underlying metaphors for meaning (Lakoff and Johnson, 1980) are ones of containment and composition. We speak of 'the meaning of X containing Y' ('recepticle' metaphor) and 'the meaning of X consisting of or being connected to Y' ('Meccano' metaphor). An outcome of this way of thinking about meaning may be that we pre-empt the nature of the composition and containment signalled by a lexical entry. The success of discontinuous feature-composition (e.g. 'girl' clearly implies 'female') may disguise the need for more complex formulations of meaning (e.g. continuous relations between objects). However neat, our atomic lists may be idealised and oversimplified.

These points may be best illustrated with an example taken from Fodor and Pylyshyn (1988). Under their compositional formulation, a word like 'good' makes a constant contribution to everything that it modifies. A 'good man' answers to the relevant interest in men, a 'good game' answers to the relevant interest in games, etc. Although there is a definite consistency in the form of modification that occurs (and it would be a strange language where this was not the case), composite expressions seem to have a meaning which access a particular nuance of 'goodness' appropriate for that combination. The difficulty is that this implies an infinite set of semantic features for lexical items like 'good'. If natural language is as expressive as is imagined to be the case, then there is no finite set of things which 'good' can modify. Rather, as new concepts emerge in the language, they can combine with 'good' to create a new meaning, and a different shade of 'goodness' results.

A non-compositional view would assert that 'good man' and 'good game' refer to different things and simply have different meanings. The subtlety of meaning for combinations of words results because of the richness of the background knowledge that is brought to bear in constructing a combination. So perhaps lexical items are better thought of as pointers to knowledge, the content of which is under constant revision and open to original combination with other knowledge. Under this view, the underlying objects for lexical items have continuous and unstable properties. At first glance, such continuity seems more compatible with the notion of partial inferential activity.

An anecdotal illustration may help the exposition of the points being made.

(9) Imagine a piano made out of cheese.

In (9), the compound expression 'a piano made out of cheese' should bring to mind a possible model for such a thing, despite it being a remote possibility that the reader has ever considered this before. It seems unlikely that the particular set of semantic features for getting the subtle 'piano' nuance of 'cheese' have at last been given the oppurtunity to combine with the pre-specified 'cheesy' nuance of meaning for 'piano'. The rich model which we can develop from the description must proceed through the combinative potential of the relevant lexical items, but its richness need not reside <u>in</u> their specification. Rather, the combined items can make reference to background knowledge existing in a form that permits the creation of the resultant subtle representation. One could argue that symbolic parsers don't operate at such a level of creative thought, and this is almost certainly true. But this only serves to illustrate the possible limitations of defining lexical items in relationship with stable underlying objects subject to compositional processing. It risks underestimating the immense contribution of rich general knowledge in the translation of description into thought.

Of course, various questions arise. If words are not just defined in a discrete way with respect to salient features, and their full meaning-potential only emerges when they serve as pointers to variety in background knowledge, then how do they operate effectively as such pointers and what range of combinatory impact can they manifest with other lexical items?

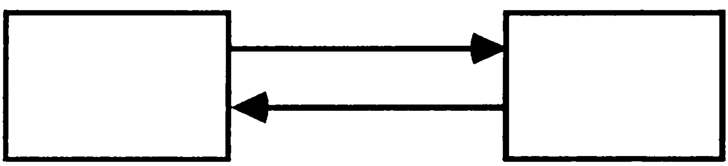
d: Semantic specification under a 'Language of Thought'

The appearance of lexical entries as the front end of the language/thought system can give the impression that language provides the entire constitution for thought. Indeed the extreme position on this is taken in Fodor's (1975) 'Language of Thought'. Here an identity relation is posited between an expression of the language and the corresponding thought: they are just viewed as being the same stuff. This position forces a specific type of semantic theory for the language. As Bierwisch (1970) writes,

'...a semantic theory must....show how the structure of the meaning of words and the syntactic relations interact in order to constitute the interpretation of sentences ...(and)... indicate how these interpretations are related to the things spoken about.'

If language and thought are the same thing, then these indications demand a specification of a one-to-one correspondence between the language and the world. The idea is illustrated in Figure 2(a).

FIGURE 2(a): Semantic model for a 'language of thought'.



Language = Thought **World**

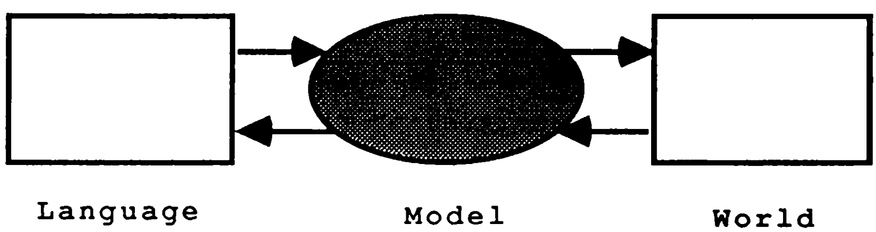
Under this model, the job of the semanticist is to define the conditions that have to hold in the world for descriptions of the language to be true. But the assumption is implicit in the model that everything the language does in 'constituting an interpretation' can be fully captured with sole reference to configurations of world objects.

A different view posits that mental representations or models (c.f. Johnson-Laird, 1983) are constructed out of the language and these then stand for states of affairs in the world. Here, the mind or brain acts as an interface between the language and the world (Garrod and Sanford, 1988). This model is presented in Figure 2(b).

The working advantage for this model is that a semantic theory specified with respect to properties of mental representations can make reference to such psychological notions as perspective, expectation and focus. In the previously mentioned work of Sanford, Moxey and Barton

(1990), the semantics for non-logical quantifying expressions (e.g. 'few', 'a few', 'occasionally') have been better captured with reference to sets of entities in the world mediated by a notion of focus.

FIGURE 2(b): Mind as interface between language and world.



For example, although the items 'few' and 'a few' are found to denote similarly small proportions (which could theoretically be captured under a 'language of thought', description-to-world semantics), 'few' (unlike 'a few') is found to put focus on the set of entities which have not been predicated by the quantifier. Hence, the linguistic description 'Few MP's were at the meeting' controls the focus of one's thought onto the MP's that were not there! Part of the meaning of this word is captured in a semantic framework which can posit perspective characteristics on sets of individuals in the world. Specification of the configuration of world objects is not enough. Similarly, Moxey and Sanford (1991) have shown how the interpretation of some quantifiers is determined with reference to an interaction between both the writer's and the reader's prior expectations about the events in question. Under this view, then, the job of understanding how language works takes on a radically different appearance. It becomes a question of how messages are designed by communicators in order to move around the thoughts of readers and listeners. It is these thoughts that represent the world that is being described.

The crucial factor in this set of assumptions is that **a mind-interfacing semantic theory does not have to assume a direct correspondence between a description of the language and a state of affairs in the world.** Because of this, the mathematical framework used to represent semantic relations can take on continuous, algebraic properties (as would be the case to capture the set-focusing properties of the non-logical quantifiers). To date, the whole formalist endeavour has assumed that language-to-world mappings can be adequately formalised in a logical system that employs discrete mathematics, but such recent work is questioning the validity of this assumption. Again, at first glance, continuous semantic specifications seem more compatible with the notion of partial inferential activity.

One can speculate on the extent to which the success of discrete mathematical formulations for other languages (e.g. computer languages, artificial languages) has resulted in the importation of that paradigm for natural language formalisms. This is particularly crucial when one considers that such languages are essentially human-made (or at least human-discovered), and retain capacity for description in domains far more easily definable than that of natural language. There is a need for increased awareness in cognitive science that such, and not the converse, has been the true order of things.

In light of the observations made in these first four sub-sections (2.3 [a] to [d]), this section is completed with a preliminary assessment of incompleteness in cohesion-establishment under the atomic processing approach.

e: Incompleteness in cohesion-establishment under atomic processing

We firstly have to ask the question in what sense do the phenomena presented in (1) to (5) exhibit incompleteness in semantic processing? Initially, two senses of incompleteness come to mind. The first concerns the component devices that make cohesive mappings. The second concerns other devices presumed to make checks on the resulting representations. For present purposes we shall assume that these devices are one and the same, but this relationship will be scrutinised more fully later.

The sorts of mappings that we have been discussing could be deemed to be incomplete on the basis of the evidence that is required for their establishment. That is, a partial satisfaction of a specification could be taken as sufficient for a mapping to be made. Hence, in (5) the activity or evidence resulting from the lexical impact of the word 'survivors' seems to satisfy the component processes enough for this item to be mapped onto the preferred interpretation.

(5) When an aircraft crashes where should the survivors be buried?

A similar process must be embodied in (2), where 'Moses' has a sufficient overlap with 'Noah' to satsify the mapping. Indeed, Erickson and Mattson (1981) speculated that the greater the number of features the correct filler and its 'surrogate' have in common, the more likely a mis-mapping will pass un-detected.

(2) How many animals of each kind did Moses take on the ark?

In both these examples, then, **not all of the semantic**

information made available by lexical items seems to be attended to by the processor. Shallow processing seems to occur. This class of example raises difficulties for atomic processors on two counts. The first is that a discrete feature-definition approach to such lexical items should necessarily make available information from core meaning (like 'alive' in 'survivors'). This availability leads to the second difficulty. No existing atomic processor satisfies mapping specifications on the basis of partial evidence (or partial inferential activity), and so such availability would necessarily result in the detection and rejection of the item as a slot-filler. The non-compositional framework which we have alluded to would not demand lexical specification on such a discontinuous basis. Neither would such a view demand that Erickson and Mattson's hypothesis be couched in terms of discrete features, but rather in configurations of continuous activity. Further, a processor that made sub-atomic rather than atomic mappings could accept partial evidence for slot specifications. Despite dealing with cohesion for semantic information, the controlling parameters of such sub-atomic devices could also be open to influence from more global, contextual factors.

This notion is best illustrated with reference to sentence (1).

(1) This book fills a much needed gap.

A symbolic processor that firstly finds potential syntactic configurations and sends them for semantic evaluation would not have any reason to propose an adjectival modifier for the noun-phrase 'This book'. Further, the relational properties of the qualification 'much needed' and the noun 'gap' are semantically compatible, and so such a combination should be successfully established. In this instance, it appears that the low pragmatic plausibility of the asserted qualification ('a much needed gap') seems to be over-ridden by the high plausibility of this qualification transferred to the other noun-phrase ('a much needed book'). In effect, the search for pragmatic satisfaction seems to be strong enough to over-ride the syntactic and semantic constraints. (Notice how the pattern in sentence (5) also straddles this semantic/pragmatic boundary. 'Burying survivors' is not semantically anomalous but it is pragmatically implausible and also of questionable moral status. Evaluation of the relationship between these levels will recur throughout this thesis.)

Existing atomic processors exhibit no properties of the kind described above. In no case can the impact of a semantic or pragmatic hypothesis be sufficiently strong for the syntactic rules to be over-turned. Nor is there any sense in which existing atomic processors appear to display strong bias towards finding pragmatically acceptable messages. Indeed, it is even difficult to characterise what atomic processors may be manifesting as expectations. If proposals from a syntactic component to a semantic component are to be thought of in this fashion, then this sort of expectation exerts no biasing effect on the form of resolution that is established for the sentence. In the human processor, **expectation at the message-level seems to manifest considerable impact on the cohesion that obtains**. This takes an important form in the other two examples, (3) and (4).

(3) Can a man marry his widow's sister?

(4) How much soil is there in a hole that is two metres deep, six metres long and three metres wide?

We have already noted that one reason why these tricks go un-noticed is that one part of the message takes the attention of the reader away from the the part displaying the anomalous relation, and even biases the cohesion process towards a mis-mapping. This discovery strongly suggests that devices responsible for establishing cohesion are effected by the satisfaction of constraints at <u>levels removed from</u>

relation-checking processing. Hence, the mention of 'sister' in sentence (3) has the confirming effect in establishing the significance of the message by mapping it onto a kinship/marital problem-type. The establishment of this message-level significance appears to relax the constraints on the devices checking for appropriate relations at the lexical level. So the valid deduction that the man must be dead if he has a widow can pass undetected, and the inconsistency is not noticed. A similar rationale can be presented for sentence (4). Attention focused on the particular dimension-characteristics of the hole, seems to distract the processor from recognising (or evaluating?) the presupposition that an empty hole can contain things

Importantly, atomic cohesion establishment exhibits no properties of focus-sensitivity. That is, there is no set of mechanisms that focus attention onto the significance of one part of a message over any other part. This may be due to the fact that particular parts of messages are not identified as significant with respect to thematic or problem-type content. If there is no thematic development from mapping to a background model, then such identification is not possible. On this count also, atomic processing fails to display the sort of machinery that the psychological phenomena seem to demand.

Perhaps the crucial property of a processor is the form of interaction of the various types of knowledge being brought to bear on an interpretation (syntactic, semantic, pragmatic etc.). However much effort is expended on making semantic interpretation as incremental as possible, the phenomena presented in (1) to (5) strongly suggest that the impact of syntactic, semantic and contextual constraints is highly interactive. It is only under such an assumption that the potential force of pragmatic constraints over syntactic ones (for example) can make any sense. Even in symbolic systems where local-contextual information is brought to bear on interpretation (e.g. those systems which keep a record of past reference resolutions, like Winograd's SHRDLU), this information doesn't come into play other than to disambiguate rival syntactic and semantic hypotheses, and it never distorts the reality of the form of the input.

What has to be kept in mind throughout this thesis is the author's view that these results are of importance because they display activity of comparable form to the activity that occurs in forming cohesive representations from 'normal' texts. The incompleteness is manifest from the system in error, but the contention is that such processing patterns will also obtain in standard comprehension. Some speculation on the relationship between communicative impact and completeness of processing will be undertaken later in the chapter.

In view of the areas in which classical symbolic approaches appear unable to capture the full range of human comprehension activity, the following section reviews the properties of recent sub-symbolic connectionist architectures to investigate whether such forms of computation can account for incompleteness phenomena in a more satisfactory way.

2.4 Properties associated with sub-symbolic connectionist models of language processing

a: A brief overview

Connectionist architectures are highly parallel computational devices. They are made up of a number of processing units or nodes which are interconnected. Each node has an activation level which has to be reached for that node to be 'switched on', and weights obtain on the connections between adjacent nodes. So when a set of input nodes is activated (with an input signal) the level of activation for those nodes is related to other nodes via the strength of the weights between them. Groups of these nodes (called 'networks') can learn to make mappings between configurations of input signals and expected outputs. Networks are trained by being given pairings of input/output patterns. Under such conditions an adjustment of the connection strengths between nodes occurs, this being driven by an algorithm correcting for any error occurring in the input-to-output mapping. It is the settling of these parameters that constitutes learning. After training, networks are often able to reliably make input/output mappings of the same type to the original training materials.

Networks can differ in many respects. The number of nodes can vary, as can the number of so-called 'hidden layers' between the input and output levels. Such variation effects the form of learning that can occur. Importantly, the training process can result in generalised learning capacity. That is, networks can create sensible and ordered output configurations from novel input patterns. This, however, is not acheived through the application of structure-sensitive rules. Rather, constraints become manifest in the configurations of activity over sets of processing units, and novel inputs realise outputs on the basis of a constraint satisfaction process.

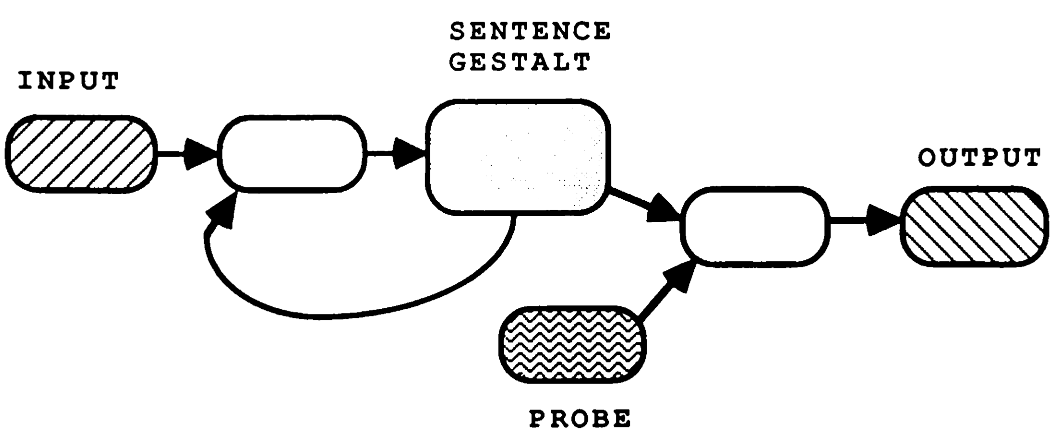
Research on connectionist architectures has been prevalent in the 1980's, and applications have been extremely varied. The present discussion does not attempt to illustrate this variety. Rather, one of the most recent language applications will be introduced, and relevant properties of this model will dominate the discussion. It will be made clear where properties of this model are particular to it only, or are typical of connectionist architectures in general.

b: The 'Sentence Gestalt' model (St.John and McClelland, 1990)

This particular model conforms to a general schematic structure for connectionist sentence comprehension mechanisms. This structure is presented in McClelland, St.John and Taraban (1989), and an adapted version is given in Figure 2(c). In this diagram the boxes represent pools of processing units, and the arrows represent connections between these pools. The 'input' pool represents the impact of successive lexical items, and the 'sentence gestalt' is the evolving sentence representation. The unmarked pool between these sets is a hidden layer of units which allows patterns of activity in the input to have a combined effect on the output to the sentence gestalt. The probe units can be used to 'question' the activity in the gestalt after any cycle. The probe acts like a window on the gestalt's current 'thoughts'. The effect of the probe is mediated through another layer of hidden units to an output layer.

Under this format there are two types of 'graded' constraints which have an effect on the form of the sentence gestalt at any time. These are the constraints imposed by the input, and those imposed by the existing gestalt. They are graded in the sense that they can exert weak or strong influence on the resulting pattern of activity.

FIGURE 2(c): Schematic structure for a connectionist sentence comprehension mechanism. (Adapted from McClelland, St.John and Taraban, 1989)



The stronger a constraint, the more difficult it is to over-ride its impact. In this model, each successive input-word serves to update the form of the representation. As McClelland, St.John and Taraban write, this representation, and the lexical input, can be

'taken to be a pattern of activation over a set of processing units.'

This particular 'Sentence Gestalt' model maps sentence descriptions onto event formulations. That is, the input nodes take a sentence description, and the corresponding gestalt denotes a situated event. This event is represented as a set of role-slots which are taken by fillers. The event formulations are very simple. There are eight possible roles in this model, some of which may be irrelevant or indeterminate for any particular description. The roles are agent, action, object, recipient, location, manner, instrument and accompanist. <u>The sort of cohesion that is established for a description is the allocation of its parts to roles in the event</u>. It should be immediately stated that these mappings are made to a very limited domain, and this observation conditions some of the conclusions about the model's breadth of applicability.

Similarly, the input formulation for sentences is simplified. Descriptions are not presented word by word as appearance in an English sentence. Rather, a sequence of constituents is presented where each consists of a unitary noun (or verb, or adverb) combined with a preposition (this could also be the auxiliary 'was'). Examples of the constituent format, together with the corresponding English structures are given in (10), (11) and (12). The correct event formulations with their role-filler pairings are also given.

(10) "BUSDRIVER" "HIT BALL" "WITH BAT"
(10') The bus driver hit the ball with the bat.
(10") (agent: bus driver)

(action: hit)
(object: the ball)
(instrument: the bat)

(11) "TEACHER" "WAS KISSED" "BY BOY"
(11') The teacher was kissed by the boy.
(11") (agent: boy)
(action: kiss)
(patient: teacher)

(12) "ADULT" "ATE SOMETHING"
(12') The adult ate something.
(12") (agent: 'adult')
[indeterminate individual, open to inference.]
(action: ate)
(object: 'something')
[indeterminate object, open to inference.]

The training for this network consists of matching input constituents with appropriate event descriptions in the gestalt. In this system there is a small vocabulary of 58 words, and each of these corresponds directly to an input unit. The set of possible role fillers is more restricted, since some of the lexical items are function words, and some can also stand for two different underlying concepts (e.g. 'bat' in both the 'animal' and 'racket' sense). Notice also that certain items are superordinate to others (e.g. 'adult' for 'school teacher' and 'bus driver'). No superordinate concept can act as a role-filler, so on the basis of context the network makes an inferential commitment to the most plausible individual under conditions of uncertainty. After any constituent input, the probe units can reveal the gestalt's assessment of the likely content of relevant fillers. For example, the 'agent' role could be probed, and the corresponding output would indicate the probability of each of the network's concepts taking this role. Again, this probing can be done at any point in the presentation of the sentence, so an index is available of the continuous commitment of the gestalt to item-to-filler possibilities.

For example, after the first constituent presentation in sentence (12), the 'adult' item may activate the concepts for 'bus driver' and 'school teacher' (the only two adults in this model) to some degree, without committing either to a fully-blown inferential status. Successive constituents are likely to constrain the representation further. There is a large amount of indeterminacy after the first constituent, but the impact of successive constituents tightens the constraints on the activation patterns of the gestalt. Under a similar rationale, the role specifications left empty by the initial description-set (e.g. an instrument role in (12)) can manifest potential candidates on the basis of the constraint satisfaction of the system. For this sentence, the nodes representing 'knife' and 'spoon' may be activated as possible instruments for the verb.

The difficulty for the network is to correctly find the appropriate role fillers from the surface structure of the initial input. Because of the structural properties of English, this can be a complicated business. For example, both (13) and (14) have the same surface form, but in (14) 'the schoolboy' clearly has to take an accompanist slot, rather than the instrument slot naturally taken by 'the knife' in (13).

(13) The bus-driver ate the steak with the knife.

(14) The bus-driver ate the steak with the schoolboy.

So semantic as well as syntactic constraints have to be established in the settling of the networks activation levels. This is comparable to the semantic selection restrictions imposed by atomic parsers. However, in the connectionist framework, the decision making is not done via absolute specification testing, but rather manifests out of the simultaneous constraint satisfaction process.

After a very long training schedule, this particular network, when tested on a number of randomly generated sentences, was able to make appropriate role-filler pairings at 99% reliability. Importantly, there is no guarantee that the system will necessarily make correct mappings, and critics of this approach could seize on this result as a weakness of the model. However, it is this susceptibility to error which is of particular relevance to the present discussion.

(15) The adult ate the steak with daintiness.

In the interpretation of sentence (15), the 'ADULT' constituent activates both the 'male' and 'female' concepts about equally, and similarly with the 'school-teacher' and 'bus driver' concepts. The 'ATE' and 'STEAK' constituents serve to increase the activity on the 'bus driver' (who is associated with eating steak). The most interesting result occurs in response to the final constituent 'WITH DAINTINESS'. Because daintiness is only associated with the school-teacher in this model, the effect of this input is to increase the activity on those concepts associated with her. In fact, one of these constraints (that she eats soup and not steak) is sufficiently strong for the activity on the 'soup' concept to outweigh the activity on the 'steak' concept which has been explicitly stated! Notice also that although sentence (15) is not anomalous or ambiguous to our minds, to the 'Sentence Gestalt' model it constitutes a set of inputs which signal conflicting information about the world. Bear in mind that it was in terms similar to these that a notion of coherence was formulated in section 1.3. It was suggested that coherence was established by interpreting

'text-based references as descriptions of <u>the same states of affairs in the world</u>.'

In the present case, the model's world is sufficiently small for sentence (15) to create ambiguity. Notice two things. First, the network naturally attempts to establish coherence in the sense postulated in section 1.3. Second, the constraint satisfaction supports an inappropriate mapping because of the conflicting relations signalled by the input-set. The system seems to manifest an in-built trust that the inputs will be coherent, and error can result when they are not.

The importance of this general finding is in indicating that this sort of connectionist system can license mis-mappings from lexical inputs to background slots. This seems to be influenced both by the impact of contextual constraints on interpretation, and also by conflicting signals at input. The lexical item in question seems unable to impose strong enough constraints in order to over-ride those imposed from elsewhere. This, of course, is the sort of occurrence that we have been speculating for the phenomena of sentences (1) to (5) ((2) & (5) in particular).

The constraint satisfaction processing displays properties that appear compatible with the idea of partial inferential activity. There is a continuous shift in the status of particular concepts as possible inferences. Lexical items do not require semantic pre-specification, and nor is an item's impact restricted to discrete feature-combination with other items. Non-linnearities in the activity of the hidden layers of nodes permit a continuous range of impact for one item in combination with others. Compositionality assumptions are not pre-requisite. Although atomic parsers give the appearance of increased capacity to interpret complex English word strings, our investigation has revealed that the form of cohesion satisfied for such strings could be severely limited with respect to the richness of thought manifest in natural language usage. There is no direct evidence that connectionist systems embody such complexity of representation either, but the properties manifest in simple systems like 'Sentence Gestalt' point towards the relaxation of atomic constraints which all but preclude such complexity. In the connectionist model, the interactions evident between the sources of constraint satisfaction both from the lexical input and background constraints are wholly consistent with the developing notion of partial inferential activity.

The 'Sentence Gestalt' model is interesting because of these properties, but in no way can it be viewed as a satisfactory framework for understanding how sentences are interpreted. We have already noted that the domain in question is very limited. This is true both for the vocabulary size and the conceptual range. Further, it is difficult to imagine that our knowledge about situations can be satisfactorily captured by simple role-filler pairings. The input format is also artificial with respect to normal sentence construction, something that symbolic parsers cope with much better. Of course, connectionist research is in its infancy, and it will be of immense interest to see whether more complex and realistic interpretive devices can be developed. Irrespective of this outcome, the class of computational devices of which this connectionist system is an example exhibit general properties of multiple constraint satisfaction. It is this factor which underlies the processing characteristics that are attractive to the developing view of inference. Before making preliminary conclusions about the properties of cohesion-establishing mechanisms, the following sub-section summarises important contrasts between symbolic and sub-symbolic perspectives on processing machinery.

c: Contrasts between symbolic and sub-symbolic representation

There has been considerable debate concerning the status of symbolic and sub-symbolic activity. One camp views the sub-symbolic level as implemetations of symbolic activity (e.g. Fodor and Pylyshyn, 1988), while the other sees symbolic form as an idealised approximation to the activity of sub-symbolic architectures (e.g. Rumelhart, Smolensky, McClelland and Hinton, 1987). The phenomena presented in this chapter would tend to support the latter view. One advantage of sub-symbolic implementations (as in a connectionist network) is that computational activity becomes tailored to the demands made of the system. There is no prescription of the form that a representation has to take, and consequently the prior assumptions made about a representation are minimised. Under a symbolic model, assumptions have to be made in order to preserve the formal status of the representation. Hence, maintaining such pre-conditions as combinatorial syntax and semantics, atomic processing, lexical pre-definition and compositionality, results in the development of models that seem too restrictive for the range of phenomena that are found to obtain in human comprehension.

The most relevant application of the symbolic framework in psycholinguistics has been in proposition-based theories of text comprehension, briefly mentioned in the previous chapter (e.g. Kintsch and van Dijk, 1978). Such theories view text representations as concatenations of propositions structured in memory almost as they appear in the text. We have already noted that models like this, which use the proposition as the primitive of representation, have difficulty in accounting for embellishment activity in the way it has been presented in this thesis. Similarly in relation to cohesion, it seems that any theory which recovers the basic propositional content of a text prior to assessing it in relation to background knowledge (i.e. prior to interpretation), has very little to say about the sort of phenomena that have been introduced. A processor that exhaustively checked the relations of items constituting propositions would, under the same rationale as a selection restriction of a parser, detect the anomalous or ambiguous nature of the phenomena in question. This argument will be expanded in later chapters.

It is not just the cohesion-related phenomena that foreground the limitations of atomic processing. Sanford and McGinley (1990) have reported on phenomena that they term 'extrasymbolic', which seem to implicate the incorporation of component activity beyond the confines of symbolic and sub-symbolic form. Many of their subjects report that vignette (16) sounds strange in some way.

(16) The car came around the bend. It almost hit a car.

For some reason, it sounds like the car in the second sentence is co-referential with the car in the first sentence. This is despite the fact that such an interpretation is effectively ruled out by all of the available syntactic, semantic and pragmatic cues. Despite this, the processor still seems to entertain the possibility of co-referentiality. In this case, an autonomous matching seems to occur for items with the same sense but different reference (Frege, 1892), an effect that would not be predicted, for example, under a formal symbolic system. The authors have called the phenomenon the 'sounds-like' effect. Interestingly, the 'sounds-like' effect can be attenuated by the incorporation of additional material in the second sentence of the vignette. This is true even for embellishments that modify the situation in general, and not the second car in particular. An example is given in (17).

(17) The car came around the bend. It almost hit a car despite some skilful and defensive driving.

What this result suggests is that stylistic variability in textual construction has a considerable effect on the sort of cohesion that the processor establishes for any particular message. Factors such as the balance of material over a sentence, or variability of word length within a message seem to effect the ease with which the processor can operate. Sanford and McGinley make an important comment.

'The (stylistic) effect does not veto a correct interpretation. Subjects all know what the intended message was. But no good communicator would ever use such sentences,

and it is not clear why on the basis of existing models.'

It is basically good science to develop theories which are as parsimonious as possible, and account for all the relevant data. The above effects, together with the cohesion-related phenomena of (1) to (5), confront existing symbolic (and sub-symbolic) frameworks with a new sort of data that calls for a richer model of text comprehension. If assumptions are being made about the formal properties of natural language, and resulting psychological models are unable to cope with the full breadth of interpretation characteristics, then there exist good grounds on which to question the philosophy of such an approach.

2.5 Further conclusions and research directions

We have observed, then, that rule-based processors establishing cohesion at the atomic level have difficulty in accounting for the incompleteness phenomena which have been presented. The multiple constraint satisfaction processing evident in sub-symbolic connectionist systems display properties that appear to be more consistent with the type of cohesion machinery envisaged for the human processor. Such connectionist modelling is very much in its infancy, and only time will tell whether or not models can be developed beyond the demonstration level. We have alluded briefly to some speculation about possible characteristics of the component processes of cohesion. The following sections look more closely at these, and in light of the findings aim to present a viable research direction.

a: Cohesion-establishing mechanisms

One observation concerning cohesion-establishing mechanisms was that they may be of two types. The first type is a class of device which licenses mapping activity. These are assumed to test specifications for slot-fillers, mediate expectation activity, authorise anaphoric reference resolutions, etc. The second type is a class of device which checks that established representations are maintaining a satisfactory level of cohesion. They ensure that only the things that ought be fitting together do so, and that they do so in a satisfactory way. Importantly, because the interpretative process manifests information and representations at a variety of levels, so cohesion-establishment can be checked at a number of levels. More on this shortly. Whether or not these functions are carried out by the same set of devices is not crucial at this stage of inquiry. The important point is that, at a functional level, there appear to be two sorts of cohesion-related activity. We will look at these in turn, and question in what sense either could be said to embody complete or incomplete sets of processes.

{1} Mapping devices

We have seen that the mapping devices used in symbolic parsing systems establish cohesion at the atomic level. The type of tests employed for such cohesion preclude the occurrence of the incompleteness phenomena. The partial inferential activity evident in these examples suggest the need for devices which process information and **license** relations below the atomic level. With 'sub-atomic' cohesion, the sort of information required to fill a slot, for example, could be a partial match to a role-specification. Licensing a mapping of one object into another could be done on the basis of knowing part of what the first object is (or part of what it has caused to happen to other things). We can see how the same rationale can be extended to functions like anaphoric resolution, where an identity relation has to be established between different discourse-objects. Cohesion at the sub-atomic level could **license** mappings on the basis of a partial match for co-reference, revealing an underlying continuous process. Such an account seems plausible for the effects found by Ehrlich and Loridant (1990). The theme, then, is that these devices may work by testing for satisfactory levels of 'goodness-of-fit'.

Continuous evidence accumulation has a well respected tradition in cognitive psychology. Study of relevant devices has included neural counters and evidence accumulators which become active or send signals on the basis of receiving a certain type (and/or amount) of information (e.g. Murray, 1970, Sanford, 1972). An example is in relation to models of simple reaction time, where devices detect signals on the basis of receiving sufficient information to pass some specified criterion. The signal detection process becomes a statistical one, since criterion levels can be manipulated in various ways, and the likelihood of detecting a signal is not a simple function of the rate of incoming information. Such criterion-passing and evidence accumulation processes have thus far only been employed in models of signal detection processes. We can see how the same rationale could be applied to discourse processing, say for accumulating semantic information in testing a slot specification. Notably, such continuous processing is a dominant feature of connectionist architectures. These provide a significant advance on simple evidence accumulators because of the large degree of interaction that can occur between processors in a network. The St.John and McClelland model has illustrated the attractive properties of models which exhibit such levels of interaction.

Under this general processing formulation, inconsistency could arise should disconfirming information arrive subsequent to the licensing of a mapping. Crucially, **incompleteness would be inherent in the process irrespective of the nature of the information, but inconsistency would manifest only in disconfirming cases**. The inconsistency would reside in the relation between the interpretation made from the message and its true content. Of course, it would seem likely that a processor would try to minimise the probability of inconsistency occuring. The difficulty in acheiving such a balance is two-fold.

First, like any cognitive process, it is subordinate to the goals of the cognitive agent. Increasing depth of processing for mapping tests might guarantee tighter cohesion, but perhaps at the expense of other important outcomes. One example might be time-related factors. There is some survival value in making fast interpretations of incoming messages. Tight cohesion may decrease the likelihood of getting the wrong message, but only at the risk of getting the right one when it is too late! Indeed, linguistic interpretation, in comparison to other cognitive processes, is very fast.

The second problem is that it is difficult to formulate what it would be to have a fully complete processing in this domain. **There is nothing inherent in the items of a text which guarantees the level of completeness required to establish their cohesion.** Sanford and McGinley (1990) make this point:

'...a complete and coherent set of relations is just that which is required to enable the making of a specified set of conclusions by inference: unless we know what conclusions we are to make, we cannot specify what constitutes completeness and coherence. Yet with many texts, it is not clear what conclusions we are supposed to make.'

If lexical items have unstable underlying properties, and can combine to form new creative representations, then what limit is there on the number of relations that are potentially relevant to checking a proposed mapping? In (18), does the reader have to access <u>every</u> possible relevant relation between Noah and the ark before it can be confident that this is what the message about?

(18) How many animals of each kind did Noah take on the ark?

If this were the case, surely the resultant over-processing would be undesirable and time-consuming. The same effect could be acheived with shallower processing, providing that some assumptions were made, perhaps concerning the communicative form of the message. These assumptions need not be explicit or even available to the system. They may just fold out as consequences of adopted processing patterns. Of course, we have seen that when assumptions are not held up in case of fact, infelicitously constructed messages can make the component processes look error-prone and stupid. This does not mean that people are bad at processing language. It means that the conditions under which processing occurs succesfully seem to be tailored to specific classes of communicative contract (i.e. co-operative ones: c.f. Grice, 1975),

The main issue is this. **Under what conditions do certain relations become subject to careful checks but others do not?** Where will the system invest in processing depth, and where will it not? This is the same as asking what inferential fields are brought into focus during interpretation, and we can see now that the issues raised in section 1.5 concerning thematic focus and inference are as relevant to cohesion-based activity as they are to embellishment. In summary, what is it about the way a text is constructed that controls attention onto some things but away from others, and how do such constructions relate to cohesion-inferences?

{2} Cohesion checkers

One might speculate that some device alerts the processor when an attempted mapping (or established mapping) is inappropriately made. Such a process could operate at a variety of levels, for instance, monitoring more global aspects of cohesion than have been considered in the above sub-section. Perhaps establishing a satisfactory overall level of coherent mappings could be sufficient for these devices to be happy that everything is in order. There are quite a few issues raised by this possibility.

First, what are the things that have to be checked in order to assess global cohesion? We have already discussed slot mappings and anaphoric resolutions. But there are a number of less obvious candidates, such as checking the truth value of presuppositions behind a question. A good example of this was presented in sentence (4).

(4) How much soil is there in a hole that is two metres deep, six metres long and three metres wide?

This presupposes that there can be such a thing as a hole which contains soil but is at the same time defined as being empty of soil. This is obviously a contradiction. Is it the case that attention can be drawn away from checking this presupposition, perhaps if successful cohesion is established for other parts of the question? For instance, the salience of the depth, length and width parameters may effect a tight mapping into an arithmetic-type problem structure. It is almost as though the processor is happy that it has found the point of the message, and conditions for cohesion elswhere are relaxed. Such patterns of commitment to topic-establishment have all the appearance of multiple constraint satisfaction. A build-up in successful mappings into a background model may be the way in which the topic or theme of the message is established. It may become harder for other items to impose strong enough constraints to overide existing commitments to the 'aboutness' of the text. Of course, when staring at a blank page, the theme or topic for the following discourse is up for grabs. The reader-to-be could be thinking about anything. Schema-based theories can be guilty of giving the impression that the background model is evoked as a natural consequence of there being a text at all. In less extended texts without titles, such as sentence (18), commitments to interpretation must proceed with far greater degrees of freedom. It is not even clear that the most contextually rich information will be signalled by items early in the sentence.

These points open up more general issues in discourse processing. For example, what is the time-course of interpretative activity for different lexical items in a text? Could it be that temporary cohesion is established on one part of a message until other items manage to signal a theme or topic? Perhaps semantic information becomes available incrementally but is not necessarily used to its full potential as it becomes available?

Related to this question, considerable research has been carried out investigating the time-course and properties of lexical access. This concerns the processes that make available the meaning of a word when it is recognised. One issue, then, concerns if, and when, contextual information comes into play in lexical access. Because a word can have more than one meaning, disambiguation may be necessary in some cases. The relationship between context and ambiguity resolution provides a test case for contextual involvement.

There are two basic positions on this issue, both of which have received experimental support. One position holds that all possible meanings of a word are activated to the same degree, and are then assessed for plausibility in view of contextual information. The other position views context more active in enhancing the most plausible meaning as activation occurs. These views result in different predictions about the state of the comprehension system. For example, an early suppression of an innappropriate meaning should reduce the effect the relevant concept can have on the representation and subsequent processing. To test for these effects, priming techniques have been used, such as lexical decision and word naming, for target items related to different meanings of words. An interactive activation approach would predict that less plausible meanings would be suppressed by context early in processing and be less able to prime other concepts. Some researchers have found evidence in support of this view (Kintsch and Mross, 1985), while others have found small, but unreliable, supporting trends in their data (Seidenberg, Tanenhaus, Leiman and Bienkowski, 1982; Blutner and Sommer, 1988). A recent meta-analysis of eighteen such priming experiments (St.John, 1991) has had increased statistical power in testing such variance. This analysis supports the interactive view for early contextual influence in lexical access.

The present thesis is more concerned with cohesion processes than lexical access. However, this result supports the general view that the availability of semantic information cued by lexical items is not a 'module' independent of the global context. **Context can interact with processing at the lexical semantics level**. Some of the experiments to be described in this thesis will investigate interactions at this level for cohesion inferencing.

Throughout this thesis there has been considerable reference to the interpretive role of 'expectations' set up via background information, and also to the enigmatic concept of 'relevance'. It is possible that both are important in the global assessment of cohesion. We have already noted that 'expectation' can mean more than one thing. It can be about how things fit together in general (i.e. accommodation). It can also be about how specific things have fitted together within a particular time-span. Under either sense, there could be interpretation-independent devices whose job it is to ensure that incoming references are compatible with the currently evoked background model. For example, perhaps part of the reason behind mis-interpretation of sentence (5) is that the item 'survivors' fits well into the background scenario for an aircrash, <u>irrespective of its role in that model under the present message</u>.

(5) When an aircrash crashes where should the survivors be buried?

(Notice, unlike (4), that a <u>semantic</u> check on the status of the presupposition that 'survivors can be buried' would detect nothing amiss.)

Whether or not a build-up in interpretative constraints can effect expectations particular to conceptual ranges, or even lexical items, is not clear. Nor is it clear what sort of interactions would take place between such levels of top-down expectation and incoming signals. This issue will be addressed more closely in the following chapters.

The 'relevance' issue seems more general and thematic, perhaps even subsuming much to do with expectation-based activity. How Grice's (1975) maxim to 'be relevant' relates to processing activity is unclear. Some attempts have been made, however, to examine this issue. Sperber and Wilson (1986) have developed an account of relevance in terms of 'contextual effects' and 'processing effort'. Their idea is that relevance is a function of maximal contextual effect (i.e. effects on the background model) and minimal processing effort. The account does not specify whether such a configuration guarantees relevance, or simply increases the likelihood of acheiving it. Perhaps, even, these effects are just outcomes of processing relevant batches of material.

Aside from this uncertainty, there is clearly some mileage in their general view. **Relevance must be a property of the relation between an existing representation and incoming information**. It seems pre-requisite that relevant information will have an effect on what has come before. We would tend to think of <u>less</u> relevant information as having more tenuous links with the current theme, and presumably, therefore, having less natural effect on it. Sperber and Wilson are not explicit in what they mean by 'processing effort'. Bananas are yellow. However, the previous sentence almost certainly demanded increased attention and effort for (hopefully unsuccessful) integration into the current theme. Clearly, a message that does not have an obvious effect on what has come before is going to demand extra interpretive resources, and in this sense low processing effort may be a consequence of reading a well-written text. For information that is not easily integrated, deeper processing may be required for interpretation. The relevance issue is something that will be returned to in later chapters. For the time-being it is sufficient to note that 'relevance relations' may provide a way of thinking about inferential control.

By way of more general comment, there may be stable differences in the grading of constraints obtaining on different aspects of cohesion. For example, it may be a strong constraint that successive inputs maintain a degree of relevance, but constraints obtaining on the completeness of anaphoric mappings may be weaker. This suggests a system where satisfaction obtained from 'important' cohesion-relations could exempt the less important relations from stringent examination. Of course, the global assessment may be far more passive than is suggested by this metaphor. Overall cohesion may even be the decision of a single device accumulating information from other sources. One of the attractive properties of this multiple-constraint view is that there is room for the graded constraints to vary as function of a whole range of things. These could be cues from the text, or derivations from the currently evoked background model. It could even be that different reading strategies or reading speeds will effect the strength of the cohesion parameters. Is there a different kind of cohesion to be established from a legal document than from a billboard? Similarly, individual differences may occur in form of reading style.

b: Focus-based inferential control

We have seen the broad issues raised by possible characteristics of sub-atomic mapping and global evaluation of cohesion. The same underlying psychological model holds for both of these issues. Writers select particular lexical items and expressions in order to effect some control over what the reader thinks about as a consequence. Part of this thought is constituted by information that goes beyond what is explicitly introduced in the text. Such inferential activity is seen to be continuous and partial, brought to bear on particular areas or fields of attention as a result of the things the writer is focusing on. This model holds for both embellishment and cohesion types of inferential activity.

In order to develop a better understanding of cohesion-based inferential control, the relationship between textual construction, focus, and cohesion activity has to be investigated. We can summarise basic principles that are open to investigation by posing as-yet unanswered questions.

1: 'Sub-atomic' mapping:

How prevalent is mapping activity that establishes cohesion at the sub-atomic level, and to what extent can this be revealed by the incompleteness phenomena that have been introduced? Despite recent research (e.g. Erickson and Mattson, 1981), such phenomena are not well understood and need to be pushed beyond the anecdotal. Also, what form does incompleteness take: what sort of 'partial' is partial inferential activity?

2: Global cohesion:

Assuming evidence for incomplete cohesion, what role might the overall status of message-cohesion have on the depth of processing for particular relations? Is it possible that processing is less complete on relations deemed subordinate to other aspects of a message (e.g. parts that signal a problem-definition)? Despite shallow processing on some relations, might the processor be satisfied with a 'good' global levels of fit?

3: Text construction and focus:

What is it about the way texts are constructed that controls the focus of the processor's attention onto some things and away from others? Crucially, how does this interact with depth of processing? For example, might cohesion vary as a function of global text-form (e.g. narrative, expository, exclamation, instruction, question etc.)? Also, assuming that manipulating emphasis involves controlling inferences, is such manipulation dependent on a sufficient depth of processing during interpretation?

4: Scenario and relevance:

If depth of processing is not constant across relation-types, then what is the source of inferential control? What role might the background scenario play in providing expectations and shaping the relations that are looked at in-depth? Could information that is found to be relevant to the established theme be given a different inferential status to information that is less relevant?

It is these issues that are under investigation in the case-study to be presented over the following chapters.

Chapter Three

A Brief Introduction to the Present Case-Study Investigating Detection Rates for Various Manipulations of a Text-Based Anomaly

3.1 Summary

The present chapter explains the methodology used to investigate the principles outlined in Chapter 2. The rate of detection for a text-based anomaly is taken to represent the depth of inferential processing for that part of the text. The items used in the case-study are introduced, and arguments are presented in favour of their selection. Some details are presented outlining the context in which the investigation took place, together with notes on the procedures adopted for assessing the detection-status of any particular reader.

3.2 The rationale behind the case-study

The empirical component of this thesis reports a case-study on the detection rates for a particular text-based anomaly. Various manipulations of this anomaly have probed the language/focus/inference relationship for cohesion inferencing. The rate of detection for a text-based anomaly is a potentially rich source of information. Although the sentences (1) to (5) have helped illustrate the sort of phenomena under discussion, there has been no indication of how likely it is that readers will notice the inconsistencies in the sentences. On the basis of general experience, the argument has simply been that there are some, and perhaps many, readers who fail to do so.

(1) This book fills a much needed gap.

(2) How many animals of each kind did Moses take on the ark?

(3) Can a man marry his widow's sister?

(4) How much soil is there in a hole that is two metres deep, six metres long and three metres wide?

(5) When an aircraft crashes where should the survivors be buried?

The rationale behind obtaining data for such rates of detection is as follows. First, the detection rate over a group of readers is taken to be a measure of the degree to which a particular inconsistency is likely to become apparent to the processing system. In effect, the detection rate acts as an evaluation of the depth of processing at the relevant point(s) of inconsistency for that text. Changing the construction of the text, and possibly thereby the salience of the inconsistency, allows comparative examination of the depth of processing for various text-forms. Sentence (6) presents an interesting example of this.

(6) Can a dead man marry his widow's sister?

Although the author has no data on the detection rate for this version, it seems likely, in comparison with (3), that the explicit mention of the man's loss of life will alert the processor to the consequent restriction on his marital potential. Of course, this change in likely detectibility must come as a result of different inferential patterns being cued by the message. Explicit mention of his death must place more focus on this condition: it may even stake a claim to being the theme of the message. When mention of marriage occurs (which presupposes being alive) the processor may be alerted to the inconsistency in the question. Perhaps a special interpretation will be made in order to discard the presupposition. For example, the man and the sister-in-law may marry in a future life, or whatever. However, this is not central to the present argument.

The important point is that it is the language that is used that brings attention onto some things rather than others, and controls the focus of the inferential activity of the processor. For both sentences (3) and (6) the same inferences are licensed, presumably with the same levels of plausibility, but the respective question-constructions must put focus on different inferential fields. Of course, there is no apparent limit to the possible alterations in question construction. Asking question (7) would surely guarantee detection.

(7) Can a man who has been dead for two years marry his widow's sister?

Even in this version, one is conscious of the effect of the 'widow's sister' expression in distracting attention from the anomaly. So, clearly, construction changes may tap into global cohesion and constraint satisfaction as well as sematic-based emphasis. The latter have everything to do with sub-atomic mapping, since in (3), (6) or (7) a non-detection would suggest an interpretation where an individual (man) was bound to a background role (groom) despite having a property inconsistent with selection restrictions on potential role-fillers (dead). In this sense, (3) is analogous to both (2) and (5) where Moses gets bound inappropriately to the role of ark-supervisor, and the alive survivors are bound to their coffins.

a: Items for the case-study

The choice of item selected for the present studies derives from sentence (5). There are two reasons for this. One limitation in the research of Erickson and Mattson (1981), who used the 'Moses' example, was that the inconsistencies were anchored to a particular narrative (or historical) event. It is likely that readers would have various amounts of knowledge about the event, and this could have a swamping effect on the detection rate comparisons over conditions. Further, a proper name is fairly limited in the ways in which it can be manipulated in a text. It can be introduced in different parts of the question, or be replaced by other names, but aside from forcing quasi-prosodic alteration (for example by underlining) there is a limitation on the interesting variations that can be tried out, such as forming more complex noun-phrases.

In contrast, the advantage of investigating inconsistencies with the 'survivor' item is that its meaning is anchored in general semantic knowledge rather than knowledge of specific events. Hence, the possibility of variable knowledge of the significance of the item is virtually ruled out. Readers either know that a survivor is an individual who is alive subsequent to a life-threatening event, or they do not. Because of the simplicity of the word, and the level of educational background of the subjects used, it was thought safe to assume the latter possibility non-applicable. Further, compared to a name, a simple noun allows more interesting manipulations such as adjectival qualification (e.g. survivors: shocked survivors), or replacement with semantically-related terms (e.g. survivors: wounded). Such manipulations will be our set of tools for investigating the language/focus/inference relationship.

It was decided also to investigate a wide variety of manipulations within one anomaly, rather than a limited range within a few anomalies, because the present research is not really about text-based anomalies. It is about inference, and text-based anomalies are relevant to inference for the reasons outlined in Chapter 2. Hence, although looking at more than one anomaly would stretch our understanding of the properties of such phenomena, the author felt that there would be more to be learned about language and inference-patterns by maximising the number of manipulations within a single domain. It is hoped that the following chapters detailing those manipulations will bear this point out. There is also a strong practical consideration in that no less than 36 conditions are to be reported. Had any given reader been exposed to more than one anomalous text, detection of one anomaly might have resulted in special attention being paid to other items. Hence, testing batches of anomalous items would not have been an informative measure of text-driven inference.

The problem this strategy raises, as with any in-depth case-study, is in generalising the findings to other domains. In this respect, arguments will be made in support of the present study in the final chapter of the thesis. Such arguments can only be satisfactorily made subsequent to presentation of the work.

b: The narrative text format

The initial experiments were carried out using an expanded version of (5), taking the form of a narrative discourse, (8).

(8) There was a tourist flight travelling from Vienna to Barcelona. On the last leg of its journey, it developed engine trouble. Over the Pyrenees, the pilot started to lose control. The plane eventually crashed right on the border. The wreckage was equally strewn in France and Spain. The authorities were deciding where to bury the survivors. They couldn't make up their minds.

Question: What would your solution to the problem be?

Notice that this format is amenable to the sorts of manipulation that have been discussed. The length of the discourse would suggest that for coherent interpretation to occur, some thematic commitment would need to be established prior to encountering the anomaly in the sixth sentence. So issues arise concerning the impact later items may need to produce in order to effect a change of direction on the thematic flow. Similarly, under a constraint satisfaction model of interpretation, processing configurations may obtain that display top-down expectation. Could the mere mention of burial within an aircrash context constrain expectations of likely objects towards dead persons? What effect would this have on detectibility for items that deviate from such an expectation, and also for the sort of cohesion satisfied for the text? The domain of uncertainty is broader still since these issues must interact with the type of manipulations that have been mentioned.

Also signalled by this text is the mapping to a unique kind of problem structure: the occurrence of wreckage equally strewn on the border of two countries. This fifth sentence is presumably a main candidate for the 'point' of the story, and provides the major source of information for establishing the problem that the authorities face: where to site the burials? Of course, this supposed mapping to a problem-structure is the same kind of thing that has been speculated in the 'distracting' effect of the rich information signalled by the relevant items in (3) and (4).

(3) Can a man marry his <u>widow's sister</u>?

(4) How much soil is there in a hole that is <u>two metres deep, six metres long and three metres wide</u>?

Could these mapping-to-problem relations result in tight bindings that boost global cohesion, but as a consequence slacken cohesion-requirements elsewhere?

c: The experimental context

It was Erickson and Mattson's (1981) finding that warning subjects of possible inconsistencies failed to increase rate of detection. It is the author's view that issuing such warnings in the present research is not a central manipulation of interest, and is even potentially confounding. Subjects may have different views on why they think they are reading the text: is it to detect anomalies or for some other purpose? Clearly, a subject reading a text in order to check for inconsistency is carrying out an abnormal task if we are interested in normal reading. Such skills as these are interesting in themselves, but the drive of this research is not on the characteristics of people's detection strategies. Rather, it is on the comprehension activity that is implicated by non-detection of such inconsistencies. Patterns of ease of detectibility may help to fill out the picture on the properties of 'normal' comprehension. As such, it was intended that subjects would read the texts in a 'normal' way, that is, with no instructions that may force the adoption of special processing strategies.

The task selected was therefore one of reading to solve a problem. The subject's task was to provide the solution they thought most appropriate in the situation. No indication what-so-ever was given that there would be any errors in the text, and subjects took part on the basis of expecting a 'problem-solving psychology experiment'. More details of the experimental method will be presented in the relevant section for Experiment 1.

d: Boundaries of detection

The main dependent variable for this case-study will be rate of detection, and it is appropriate at this point to introduce some procedures used to determine what does and what does not constitute detecting an anomaly. This is in fact a more rigourous process than may be supposed, for two main reasons.

First, the content of a subject's solution is not an unambiguous source of information on which to base a detection decision. Sentence (9) is an example of a possible solution.

(9) The dead should be buried in their home countries.

The mention of burying the dead would suggest that the subject had failed to notice the true description ('survivors') and had based their solution on an erroneous interpretation. However, the possibility exists that detection of the anomaly had been made, but that it was followed by a pragmatic evaluation to the effect that a solution should be offered 'as though it had said dead'. This pragmatic inference may even be a function of the type of label in question. For instance, the subject may think that what is being meant is that the survivors go on to die, although the present text would be a strange way of communicating that message! Or subjects may feel under communicative/experimental constraints to ignore the apparent inconsistency and attribute it to a typing error or some other factor at that level.

The other problem, not unrelated to the first, is that a reader may detect the true description but not evaluate it to be anomalous. The nature of the present 'anomaly' is basically a pragmatic one. There is nothing semantically anomalous about burying survivors: the expression denotes a comprehensible and consistent state of affairs. The anomaly arises because that state of affairs does not fit into generally held views on what ought to happen in the world. Indeed, such a proposition can (thankfully) arouse strong emotions of disapproval. At the experimental level, however, it is not sufficient to assume that solutions such as (10), for example, have arisen from non-detecting interpretations.

(10) They should be buried in the nearest possible site.

The consequence of these observations is that experimenter detection decisions have to be made on the basis of a de-brief interview subsequent to the subject providing a solution. In the first instance, this simply involves asking the subject if they had noticed anything strange or anomalous about the text. If this fails to produce a reaction, then a transcript of the text is presented, and they <u>then</u> re-read the narrative in knowledge that it may have some inconsistency. At this point, the experimenter points out the content of the sixth sentence, and in the vast majority of cases there is a clear-cut reaction to the true content. It transpires that subjects either express surprise (even dismay) at having missed the anomaly, or they assert that they had noticed the description and had made an interpretation and solution accordingly. The nature of the text and the anomaly makes it very difficult for a non-detector to give the appearance of a detector in the de-brief phase. It is always on the basis of this de-brief interview that the evaluation of detection is made.

The empirical work is presented in the following six chapters.

Part Two

Chapter Four

Preliminary Narrative-Based Manipulations: Experiment 1

4.1 Summary

The experiment described in this chapter investigates the prevalence of 'sub-atomic' mapping in the aircrash narrative. The anomalous 'survivors' item is contrasted with manipulations incorporating other crash-related individuals. 'Sub-atomic' mapping is found to be prevalent. The detection rate comparisons across conditions suggest a cohesion process that is incomplete by testing for generalised classes of information. It is argued that the findings rule out models of incomplete processing based on simple overlap of semantic features.

4.2 Experiment 1: Basic manipulations investigating sub-atomic mapping

a: Introduction

The manipulations tested in this experiment investigated properties of sub-atomic inferential activity. Non-detection of the basic '...where to bury the survivors' expression would suggest the following underlying process. Under whatever conditions of expectation, the semantic specification for 'survivors' must be treated as a filler for the object slot of the 'bury' verb-schema. The item effectively <u>stands for</u> 'dead persons' in the resulting model, having been authorised as an item supporting that interpretation. The resultant inconsistency between the model and the text shows that the processor has failed to incorporate (or evaluate) the full meaning of 'survivor'. We have speculated that such omissions are the outcome of cohesion-establishing procedures that are necessarily incomplete. Where such inconsistency arises, the mis-mapping will be described as a process of **semantic surrogation,** since the text-based description is an artificial implant into the text-model. Presumably, incompleteness is just as likely to obtain in the interpretation of non-anomalous descriptions (e.g. '...where to bury the dead'). In such cases, however, no inconsistent outcome would result. There is no reason known to the author to suggest that processing is less complete just on those occasions when inconsistency arises.

This particular framework raises two related issues. First, how prevalent is surrogation of this sort? Are non-detections simply the consequence of casual engagement with the text, or is surrogation an outcome of genuine comprehension activity? Second, what relationships between the text and the background model are most susceptible to surrogation? Presumably, surrogation is dependent on a close relation between the meaning of the 'expected' description and the actual one. A main issue is how to characterise 'closeness' in terms of semantic processing.

In the present case, 'dead' and 'survivor' both denote accident-related individuals which are predicted by an 'aircrash' scenario. Could it be that an insensitivity to the negation, or otherwise, of an 'alive' semantic feature effects a close enough match for acceptance into the schema? Perhaps the processing configuration accepts an accident-related individual who has 'something to do with living or dying' ('still alive', or 'was alive'), and this level of examination is deemed sufficient to **license** the mapping. The chances of an inconsistency arising at such depths of processing might be minimal under general communicative assumptions. Why process information in more depth than seems necessary to obtain an appropriate interpretation? **Under this view, cohesion processes access the full definitional content of the 'filler' term, but license relations on the basis of partial tests on the status of its semantic components.**

Of course, this speculation presupposes that surrogation has any prevalence in interpretation of this sort, and this was tested in the current experiment. The relationship between expectations and would-be surrogates was probed further. For the present materials, the above view suggests that mapping tests are carried out as follows: information wanted, dead or alive, providing the description is of an accident-related individual. So the obvious manipulation in this accident-related 'space' was to test detection rates for descriptions of other predicted individuals. In the present case, three other separate items were tested: **injured, maimed** and **wounded**.

Injuries and wounds can vary in their severity, though perhaps, *a priori*, an injury is likely to be more serious than a wound (which could be a cut or a bruise). The damage to a maimed person suggests the loss of a particular physical function. Compared to a survivor, the life-status of these individuals is less clear-cut. They all suggest some kind of damage as a result of an accident (in this context), but whether the individual survives or dies as a consequence is open to inference. It could be that in not blocking an inference for subsequent death, **wounded** (for example) is interpreted as meaning 'wounded and died later'. The corresponding inference for survivors, 'survived and died later', is anomalous (but see discussion section for a fuller analysis). Hence, for the **wounded** set, readers may show no detection of an anomaly, but no acceptance of the idea that there had been an anomaly either.

Arguments can also be made in the opposite direction. **Injured, maimed** and **wounded** are all terms which are highly consistent with a generalised notion of the destruction of life which follows most aircrashes. On these grounds it could be argued that these terms will be more prone to surrogation than the **survivors** item which is destruction-related but implies continued life as an outcome. **Under this view, the fit required depends more on a generalised notion of damage and destruction**. The issue is more complicated still, because these entailments might not become fully available to a processor operating under shallow semantic processing!

Any differences in the detection rates for these items might help to clarify the type of incompleteness that is suspected to occur. If definitional information is what is most likely to be processed, then one might expect the highest detection rate from **survivors**, and a lower rate (but not based on anomaly) from the other terms. If the fit of the expression to some generalised notion of destruction is what is taking place, then the other pattern would be predicted.

b: Method

{1} Materials and design

The materials used were three texts each of seven sentences in length. Each of these texts described a problem that an individual or group of individuals was experiencing in a particular situation. Two of the texts were used to familiarise the subjects with the experimental procedure, and the other was the experimental aircrash narrative as presented in vignette (8) of Chapter 3, and re-presented below. Neither of the other texts contained anomalies.

There was a tourist flight travelling from Vienna to Barcelona. On the last leg of its journey, it developed engine trouble. Over the Pyrenees, the pilot started to lose control. The plane eventually crashed right on the border. The wreckage was equally strewn in France and Spain. The authorities were deciding where to bury the survivors. They couldn't make up their minds.

Question: What would your solution to the problem be?

There were four independent conditions in this experiment, each condition reflecting a text-based manipulation. The first condition was as printed above, and the other three conditions were identical except for the substitution of **injured, maimed** or **wounded** in place of **survivors**. Four independent groups of subjects saw only one of these versions. All subjects saw the same first two texts, which are given in Appendix A.

{2} Subjects

These were mostly undergraduates at the University of Glasgow. A total of 90 was tested. After reading, some subjects claimed to have familiarity with the aircrash 'trick', these subjects being removed from the initial analysis. Because of this factor, attempts were made to balance the subjects without previous knowledge across conditions.

{3} Apparatus

The materials were presented on an Apple MacIntosh computer screen (MacPlus version) calibrated for self-paced reading with millisecond accuracy. By pressing the space-bar subjects could pace the rate at which successive sentences of the text were presented to them, and the computer measured the dwell-times on each sentence. A pen and paper were provided for the subjects to write down their solution after reading the problem.

{4} Procedure

Each subject was informed that the experiment involved answering some problems set on a computer screen. The experimenter displayed the use of the computer equipment by using the space-bar press to move through a test trial (using the sentences of the first text [see Appendix A]). A space-bar press moved the screen-display from "READY" into the first sentence of the text. Subsequent presses moved the reader from sentence to sentence. The current sentence was removed from the screen as the next appeared.

Subjects were asked to read at their normal pace. They were encouraged to imagine that they were reading a newspaper report. They were told that each of the texts would be seven sentences in length, and that following the seventh sentence the computer would set them a question about the text. In each case the computer asked that the subject write down their own solution to the problem that has been described. After writing down a solution, the subject made a space-bar press which moved the program from the "QUESTION" display to the "READY" display for the next trial. Each subject was asked to inform the experimenter when the computer displayed "FINISH" after the last trial. Each subject provided a solution to two problems, one on a pre-test filler problem, and the other on the aircrash narrative.

When each subject had finished the experiment, they were de-briefed as described in the final section of Chapter 3. A record was kept of each subject's detection status, and whether or not they had previous knowledge of the problem.

{5} Treatment of results

There are four areas of measurement for this experiment. First, there are the detection rates which are expressed in percentage form (i.e. the percentage of subjects who managed to detect the true label in sentence six, <u>irrespective of whether or not they evaluated it to be anomalous</u>). Second, there is the content of each subject's solution, which has been analysed with respect to both the type of solution provided, and also the references used for the passengers in question. Third, there are the dwell-times for the sixth sentence of interest, and last, a reading speed measure that has been computed as the sum of the dwell times over the four sentences prior to the anomalous one.

c: Results

{1} Detection rates

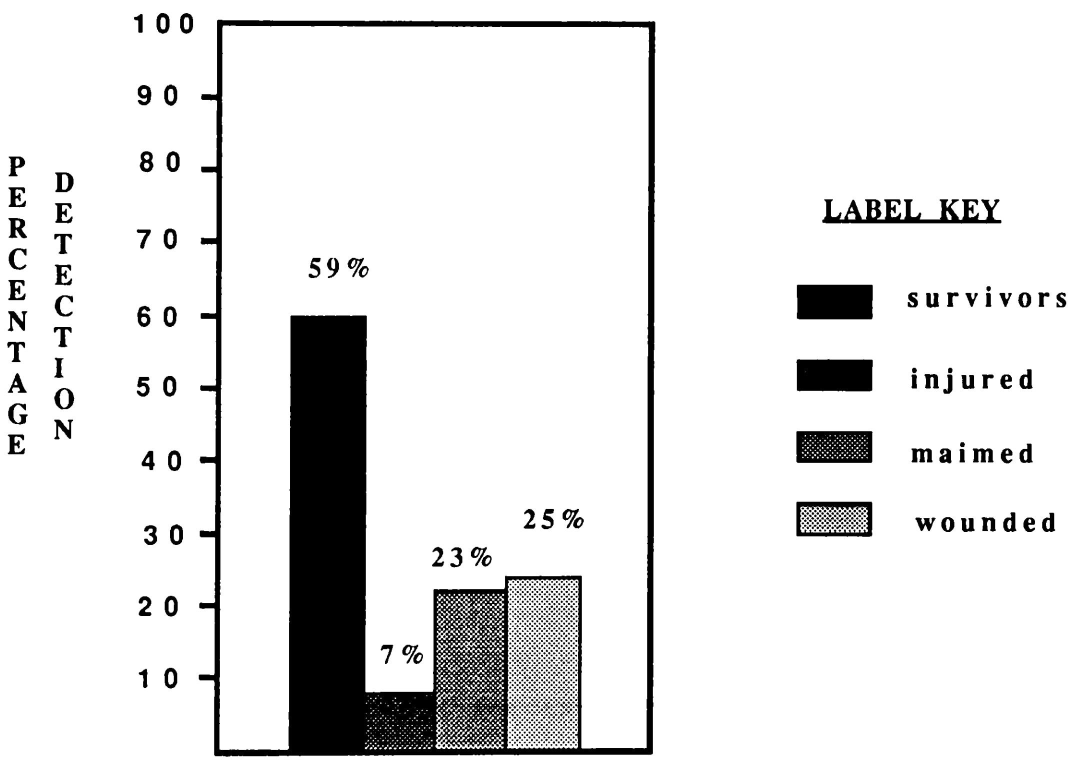
Of the 90 subjects tested, 33 had previous knowledge of the problem and they were excluded from the main results. In fact, despite having encountered the problem before, 3 of this total failed to detect the anomaly. This result will be evaluated later. The number of subjects detecting and non-detecting in each condition is presented, with the percentage detection rate, in Table 4.1 below. The percentage detection rates are also presented in diagrammatic form in Figure 4(a).

TABLE 4.1: Detection rates for Experiment 1.

Condition	**Detections**	**Non-detections**	**Percentage detection**
survivors	10	7	(59%)
injured	1	14	(7%)
maimed	3	10	(23%)
wounded	3	9	(25%)
All conditions	17	40	(30%)
Subjects with previous knowledge			
survivors	12	1	
injured	4	1	
maimed	7	0	
wounded	7	1	
All conditions	30	3	

By any standards, the overall detection rate of 30% is astonishingly low. More than two in every five subjects <u>failing</u> to detect the anomaly seems surprising enough in the **survivors** condition, but this is tame in comparison with the **injured, maimed** and **wounded** conditions. An overall Chi-square analysis on detection values for the four conditions yields a significant effect (Chi-square=13.03, d.f.=3, $p<0.01$). From inspection of the rates across conditions, **injured**, **maimed** and **wounded** appear to act as <u>more</u> effective surrogates. In these conditions, less than one in five subjects detect the anomaly, and this is as low as one in fifteen for the **injured** label.

FIGURE 4(a): Detection rates for Experiment 1.



These non-detections should not be confused with pragmatic-type inferences to the effect that **injured**, for example, must mean 'injured and died later'. A non-detection literally is a non-detection. All of the forty subjects attributed non-detection status (70% of readers!) were <u>not conscious of reading about anything other than dead passengers</u>, and typically reacted with surprise or even dis-belief on being shown a transcript of the text.

{2} Anomalous sentence dwell-times and reading speeds

The mean dwell-times for the anomalous sixth sentence are presented in Table 4.2, and the mean reading speeds (sum of the dwell-times on sentences two, three, four and five) are presented in Table 4.3.

TABLE 4.2: Anomalous sentence dwell-times for Experiment 1.

Dwell-Times (milliseconds)			
Condition	**Means for Detectors**	**Means for Non-detect**	**Total Means**
survivors	5595	3125	**4360**
injured	9727	3729	**6728**
maimed	8766	3886	**6326**
wounded	3003	3125	**3064**
Total	5940	3527	**4247**

TABLE 4.3: Reading Speeds for Experiment 1.

Reading Speeds (Seconds)			
Condition	**Means for Detectors**	**Means for Non-detect**	**Total Means**
survivors	15.3	13.1	**14.2**
injured	16.7	11.8	**14.3**
maimed	17.5	11.1	**14.3**
wounded	9.7	12.5	**11.1**
Total	14.8	12.0	**12.8**

The magnitude of the reading speed and dwell-time means suggests that the low detection rates are not a simple artifact of superficial engagement with the text. A mean reading speed of over twelve seconds for the four sentences prior to the anomalous one indicates that subjects were not simply fast-scanning the materials, but were actively engaged in the task. A mean dwell-time of over four seconds on the sixth sentence suggests this also.

From inspection of the means, dwell-times look to be longer for subjects who detect the anomaly, the exception being in the **wounded** condition. The number of subjects detecting is so low in the **injured, maimed** and **wounded** conditions that it is difficult to infer too much from these apparent differences. Parametric statistical analysis has been left until further manipulations have been reported. A general analysis for all relevant conditions will take place in Chapter 6. A similar analysis will take place for the reading speeds which, again, look to be longer for detectors (with the exception of the **wounded** condition). Discussion of these possible effects will be left until the general analysis.

{3} Content analysis for the written solutions.

There were two aspects of the written solution that seemed important. The first concerned the type of solution provided (e.g. that no problem existed at all, or that the bodies should be flown home, etc.), and the second concerned the expressions used to refer to the passengers. Of course, we might expect that only those subjects detecting the true label would use it to refer to the passengers, or that non-detectors would predominantly provide solutions that involved burying dead people.

To test such ideas, two independent judges rated each of the written solutions on pre-set scales for both solution-type and reference-expression. Only solutions for subjects contributing to the detection rate were assessed. The judges were unaware whether or not a particular solution had arisen from a detecting or non-detecting subject. They were simply informed that the solutions had come from a problem concerning where people should be buried after an aircrash.

In the description of the problem, rather than asking about burying the 'dead', other expressions had been used. For whatever reason, some readers had failed to detect this, and had thought that they had read about dead people. The pre-set categories for solution-type are given below in Table 4.4, and those for reference-expression in Table 4.5.

TABLE 4.4: Pre-set categories for content analysis of written solutions: solution-type.

Five Solution-types	
1: Family	Family, relatives or friends to decide on the burial place.
2: Home	Burial to take place in the passenger's home, home country, country of origin, place of birth etc.
3: Crash site	Burial to take place at the site of the crash.
4: No problem	The solution is that no problem exists since there is no need to bury people who are still alive (e.g. 'You don't bury survivors' or 'Is this a trick question?' etc.)
5: Other	Any other solution offered.

It was made clear to the judges that in some instances a particular solution may appear to fit more than one category. In such cases, judges were asked to fill in <u>all</u> categories that appeared applicable for that solution. Also, there was no demand that an evaluation for each type be made for each solution. Some solutions may not make explicit reference to the passengers, for example.

The measure taken for overall evaluation is the area of agreement between the two judges at the level of the solution. That is, where both judges make the same evaluation for a particular solution (on either the solution-type or reference-expression scales), that evaluation is credited towards a general concordance measure. Evaluations that are not agreed upon are ignored and left out of the analysis. Agreement, of course, refers to an observation of the judges ratings when made independently, not to any sort of negotiation between the judges.

TABLE 4.5: Pre-set categories for content analysis of written solutions: reference-expressions.

Five Types of Reference-Expression	
1: Dead	(Also **deceased, victims**)
2: Bodies	(Also **remains**)
3: General	(i.e. **people/passengers/tourists**)
4: Anomalous	(i.e. **survivors, injured, maimed,** etc)
5: Other	(e.g. pronominal references)

The content analysis for the solution-types is presented, for each condition, in Table 4.6, and the corresponding analysis for reference-expressions in Table 4.7. A transcript of the solutions is presented in Appendix B.

Note: Since a single solution can result in more than one attribution to the pre-set categories, for some conditions the number of concordant evaluations may exceed the number of solutions. In other cases, because of the agreement constraint for concordance evaluation, the solutions may be greater in number.

For the 40 non-detecting subjects, none was found to write that no problem existed: detection of the anomaly would, as expected, seem to be a pre-requisite for forming such an answer. Thirty-five of the 45 concordant evaluations for non-detectors were in the 'family' and 'home' categories.

These are clearly the dominant solution-types for subjects who are, presumably, deciding where to bury dead passengers. Few subjects think that burial should take place at the 'crash site', and only a small number of subjects presented other solutions outwith these three categories.

TABLE 4.6: Content analysis for solution-types in Experiment 1.

	Solution-type				
Condition	**Family**	**Home**	**Crash**	**No prob**	**Other**
survivors					
10 detectors	-	2	-	8	-
7 non-detect	5	2	1	-	-
injured					
1 detector	-	-	-	-	-
14 non-detect	8	6	2	-	1
maimed					
3 detectors	-	1	-	1	1
10 non-detect	5	4	1	-	2
wounded					
3 detectors	-	-	-	3	-
9 non-detect	2	3	-	-	3
Total					
(17 detectors)	(-)	(3)	(-)	(12)	(1)
(40 non-detect)	(20)	(15)	(4)	(-)	(6)

For the 17 detecting subjects, 15 concordant evaluations were made and 12 of these were that no problem existed. Clearly, the majority of detectors think that the description is anomalous: they do not tend to answer as though it had said 'dead', or happily infer that what is meant is that the individuals go on and die as a consequence of their present state. The small number of detectors makes comparison across conditions difficult.

Consistent with the above findings, detectors used the anomalous label to refer to the passengers on 12 out of 17 evaluations. Two references were made to the 'dead' and one to 'bodies', these presumably in the solutions of the subjects making pragmatic inferences (of whatever type). The non-detector's reference patterns are especially interesting. Of the 39 concordant evaluations, 22 referred to items in the 'dead' or 'bodies' category, and 8 used a general reference expression (like 'passengers'). Clearly, most solutions for non-detectors involve the passenger's homes and families, and these passengers are believed to be dead (since this is the dominant form of reference used).

TABLE 4.7: Content analysis for reference-expressions in Experiment 1.

	Reference-expression				
Condition	**Dead**	**Bodies**	**Gener-al**	**Anoma-lous**	**Other**
survivors					
10 detectors	1	-	-	7	2
7 non-detect	3	-	1	3	-
injured					
1 detector	-	-	-	1	-
14 non-detect	4	5	3	-	2
maimed					
3 detectors	1	1	-	1	-
10 non-detect	2	3	1	-	3
wounded					
3 detectors	-	-	-	3	-
9 non-detect	2	3	3	-	1
Total					
(17 detectors)	(2)	(1)	(-)	(12)	(2)
(40 non-detect)	(11)	(11)	(8)	(3)	(6)

There were, however, 3 concordant evaluations where a non-detector referred to the anomalous label, and these all occured in the **survivors** condition. How can it be that a reader unaware of the true description in sentence six used that expression in their solution? In fact, this is exactly what happened. The three solutions in question, <u>all for subjects who had failed to notice the anomaly</u>, are given below in (1), (2) and (3).

(1) The survivors relatives should be contacted and their remains flown home for burial.

(2) Survivors should be flown back and buried in their home countries.

(3) The survivors should be buried wherever their relatives wanted them buried.

On first reading, these solutions look as though they are written from a rather uncaring perspective: the writer seems happy to bury people who are still alive. In fact, not only were these subjects unaware that they had read about where 'to bury the <u>survivors</u>', they were also unaware that they had written about survivors in their solution! The whole reading and problem-solving operation had occurred under an interpretation of burial for <u>dead persons</u>, and the first time each subject was aware of having read, and written, 'survivors' was when the experimenter pointed out both facts to them. This is an astonishing result. Although it is only found for three subjects, it is obvious even from this level of effect that semantic surrogation can occur at a depth far removed from the reader's or writer's awareness. There can clearly be a considerable separation between the surface structure and the interpretation of lexical items, this distance allowing inconsistency to arise between the two, even in cases of language production. This finding further discredits proposition-based views of text comprehension and interpretation. Text-representations clearly do not necessarily bear the direct correspondance with their input that such views suggest (e.g. Kintsch and van Dijk, 1978).

Further, the content of the solutions for both of the answered problem-texts (trial and experimental) suggest that all subjects were engaged attentively in the task. There can be no doubt, then, that semantic surrogation is an outcome of genuine comprehension activity and, from a scientific perspective, can no longer be attributed mere anecdotal status.

4.3 Discussion

a: Issues arising at the level of the present manipulations

The results have confirmed that surrogation under these conditions is prevalent, supporting the view that incompleteness of semantic processing is inherent to text comprehension. The pattern of detection rates can be used to investigate the type of incompleteness obtaining on comprehension. The results are consistent with the view that a fit is being attempted to some generalised notion of destruction cued by the context. Hence, the terms most consistent with damage-related outcomes (**injured, maimed** and **wounded**) are mapped more consistently into the background model. As a consequence of the effectiveness of this mapping, the full implication of these terms does not seem to develop. Consequently, non-detecting subjects are not aware of having read about injured, maimed or wounded (or survivors for that matter), but only about dead people. Clearly, what does <u>not</u> tend to happen is that readers become conscious of processing the relevant term and then of assessing how consistent it is with a death-outcome. This pattern emerges <u>only for those readers who manage to detect the real label</u> (30% of subjects), and we have seen from the content of their solutions that they are not well disposed towards a pragmatic inference for subsequent death. (Note: However, detectors are predominantly from the **survivors** condition). Had processing involved close inspection of the definitional components of the terms, the opposite pattern of detections could have resulted. **Survivors** is more clearly defined for entailments of life-status, and searches for well-defined semantic features at this level would have made surrogation more likely for this term.

In support of this argument, there are a number of discourse expressions which illustrate differences in the death-enabling inferential patterns of the items used. Clearly, everyone goes onto die, but in this context some constraints must come into play to narrow the time-span for effects localised to the s cenario in question. For example, it is unlikely to be the authorities problem where to bury the wounded should they die forty years later.

Sentences (4), (5) and (6) illustrate the general view that subsequent death is licensed and plausible from the **injured**, **maimed** and **wounded** items. In contrast, sentence (7), positing the same relation to subsequent death for the **survivors** label, sounds very strange and inappropriate.

(4) The men were injured and later died in hospital.

(5) The men were maimed and later died in hospital.

(6) The men were wounded and later died in hospital.

(7) The men survived and later died in hospital.

Survival conjoined with later death causes an inconsistency. Employing a contrastive conjuction seems to improve this relation, but the other terms fail to generate a sufficient contrast with subsequent death to warrant use of this conjunct. Sentences (8), (9) and (10), at best, fail to provide more information than (4), (5) and (6), and, at worst, sound strained. However, (11) is far more acceptable than (7).

(8) The men were injured but later died in hospital.

(9) The men were maimed but later died in hospital.

(10) The men were wounded but later died in hospital.

(11) The men survived but later died in hospital.

In summary, all of the terms presuppose enabling conditions for subsequent death, but they differ in the complexity of the conditions required for that enablement. Hence, since **survivors** is, in this respect, a stage removed from the destruction associated with death, it is detected more frequently than the other labels. The fit is less good under tests to some generalised specification of destruction.

b: General issues and new manipulations

The emerging patterns of surrogation suggest that cohesion-based activity cannot be fully described with sole reference to semantic features and the degree of overlap between 'expected' features and input information. A processing model of this type was proposed by Erickson and Mattson (1981). Rather, **the present results suggest that the important relations for cohesion rely more on properties of the inferential fields signalled by terms**. These inferential fields defy discrete feature-definition. For example, what sort of 'feature' would describe the likelihood that an injured person would die as the result of their injuries? Such forms of description take us back down the road of infinite feature lists, since every possible inferential relation would need to be specified at the lexical semantics level. Under present conditions, cohesion-establishment is satisfied in the relation between two sets of inferential fields: those owned by the background model, and those brought into focus by the text. This relationship can result in inconsistency between the words used to express the model, and the actual words present in the text. Such a relation is not necessarily inconsistent. For example, the gist of a text may be communicated using different words but still be consistent with its original. In the present case, the interpretation is not consistent with the message.

The present results also have implications for general models of sentence processing. That **wounded** can be interpreted in the way suggested goes against any claim that cohesion for noun-phrases, for example, has to be satisfied prior to interpretation. At least, such processing configurations do not appear necassary for global cohesion to be established. Also, theories promoting the view that propositions are 'extracted' from texts prior to interpretation (e.g. Kintsch, 1974) are further discredited by these findings. The present evidence suggests that inferential activity can be far more localised to sub-propositional levels. Were such local relations necessarily subject to assessment within the larger representation at the propositional-level, detection of anomalies of the present type would be guaranteed. That so many subjects fail not just to detect an anomaly, but also to become aware of the meaning of the words they are reading, may suggest some kind of top-down influence. At what level such expectations may be operating is not clear.

Importantly, incompleteness in the cohesion process supports the view that interpretation is not necessarily carried out in a compositional manner. Expectations derivable from background scenarios are able to exert constraints on interpretation which bias against the lexical items of the text 'combining' to a compositional specification.

One property of a build-up in thematic and interpretative constraints may be that successive items require increasing focusing power to shift the attention of the processor onto new fields of information. Of course, this model could obtain at a variety of levels within the system, from shifting global topic, to alerting special attention for a particular lexical item or phrase. For the timebeing we can think of this in terms of the depth or shallowness with which parts of a text are attended to. In the present case, there are a couple of reasons why the inclusion of the anomaly in its position in the sixth sentence may be optimal for its non-detection. There is the confirming effect of mapping to a problem-structure signalled by the content of sentence five, which may effectively relax subsequent cohesion demands. Since the text comes in a narrative format, the 'ending' in interpretation terms is likely to co-occur with the 'end' of the text (in serial presentation terms). Hence, it becomes likely that top-down expectations for particular inferential fields will be maximal at this point, as will the impact required of relevant items to disconfirm such expectations.

Hence, new issues emerge for a deeper understanding of the language/focus/inference relationship. For example, what is it about the item **injured** (for example) that prevents it from putting focus onto non-dead but injured people? It is presumably for this purpose that such a word exists in the first place. We have already discussed the overall interpretative conditions that may have contributed to non-detection, but what about conditions at the level of textual construction: it is, after all, the text that is driving the processor? Presumably, for things about injured people to come into focus, the item signalling those things would require to generate more impact than the basic **injured** term seems able to do. One possible manipulation would be to place theme-independent emphasis on the item, say by using upper-case print. Increased attention may result as a consequence of text presentation. More interestingly from the present perspective, is the possibility of emphasis manipulation which would increase the overall surface structure of the signalling item(s). *A priori*, the more work the processor has to do on a particular part of a message, the more likely that things signalled at that part of the text will effect the thematic flow of interpretation. Hence, one possible manipulation would be to replace **injured** with the relative clause **people who were injured** which **licenses** the same inferential fields but may have sufficient surface-structure complexity to put focus on the (non-dead) injured rather than the dead. Manipulations of this type will be incorporated in the next two experiments. As well as this form of emphasis manipulation, there are forms of emphasis which are more semantic-based.

If we return to sentences (4) to (11), it became apparent that because **injured, maimed** and **wounded** were indeterminate with respect to subsequent death, their inferential fields were more compatible with notions of general destruction. Qualification of the **injured, maimed** and **wounded** with an adjective that biased against death as an outcome should **license** against semantic surrogation to the same degree. Hence, if the **surviving injured** were described, one might expect that this manipulation of emphasis would be sufficient to set up a contrast with the non-surviving injured (i.e. dead). As a consequence, the inconsistency should be more apparent to the processor. Such semantic-based forms of emphasis are potentially informative about the relationship between the surface structure of a text, the resultant focusing impact of the terms, and the inferential activity of the processor as an outcome. It was under this rationale that manipulations of this type were also incorporated within Experiments 2 and 3. Both of these experiments will be described in the following chapter.

Chapter Five

Emphasis-Based Manipulations: Experiments 2 and 3

5.1 Summary

Both surface-complex and semantic-based emphasis manipulations were tested in Experiment 2. The experimental procedures were identical to those of Experiment 1. An increase of anomaly detection was observed in only some of the conditions, suggesting that expected focusing impacts had been attenuated. Stronger manipulations in Experiment 3 again failed to increase detection. The detection patterns suggest that interpretation may be supported by the inferential fields owned by part of a composite expression. However surprising, mappings may be made on the basis of evidence from these confirming 'fields', independent of the significance of the combined description (which may not be fully computed). Such outcomes may be conditional on top-down testing.

5.2 Experiment 2: Emphasis-based manipulations investigating focus and sub-atomic mapping

a: Introduction

Chapter 4 has opened up investigation of the relationship between textual construction, focus and inferential activity. Two related types of text-based manipulation emerged: increased surface-structure complexity, and semantic-based emphasis. The first involves description-types that **refer to the same set of individuals but do so through a more complex linguistic form**. The example given in Chapter 4 was a relative clause construction for the **people who were injured**. More basic than this manipulation is the incorporation of a noun with an added qualifier. For example, the **surviving passengers** as an expanded version of the **survivors**. There are prior grounds giving rise to interest in this type of manipulation. From continuation-based studies text) it is known that nounphrases of this type are more frequently referred to, and that they are also more likely, in comparison with unqualified nouns, to be included in memory protocols for texts (c.f. Sanford and Garrod, 1981).

These results suggest that the surface complexity of a description can have a controlling effect on the type of representation arising from a text. Importantly, these contrasts, such as **survivors** and **surviving passengers**, license the same inferential fields, but may effect different levels of impact on a representation. Hence, across whatever sets of processing conditions, the latter description is more likely to put focus on the individuals in question. Under this hypothesis, it would be expected that the incorporation of **surviving passengers** into the present format would increase the rate of detection for the anomaly. This manipulation was tested in Experiment 2.

The semantic-based emphasis described in Chapter 4 was applied to the **injured**, **maimed** and **wounded** conditions resulting in **surviving injured**, **surviving maimed** and **surviving wounded** descriptions. To re-cap briefly, although these items are more surface-complex expressions, the principal manipulation is that **the meaning of the composite expression licenses more restrictive inferential fields than the unqualified version**. A **surviving injured** person lives despite injury, and contrasts with people who die because of injury. The qualification should bias against the inference of subsequent death as a (local) outcome.

Hence, the four conditions tested in Experiment 2 were direct, emphasis-based manipulations of the conditions employed in Experiment 1 (as below).

Expt. One		Expt. Two
survivors	:	surviving passengers
injured	:	surviving injured
maimed	:	surviving maimed

wounded : surviving wounded

The expectation in each case was that the rate of detection would show a highly significant increase on the corresponding condition from Experiment 1. No significant amount of surrogation would be expected with these emphasis manipulations.

b: Method

The materials, design, subjects, apparatus and procedure were identical to those for Experiment 1 (clearly, a different group of subjects was used). The only difference in design was that the terms **surviving passengers, surviving injured, surviving maimed** and **surviving wounded** were substituted as items in the **survivors** position of the aircrash narrative. These constituted the four conditions of the experiment, tested on independent groups of subjects. Again, a total of 90 subjects was tested.

c: Results

(1) Detection rates

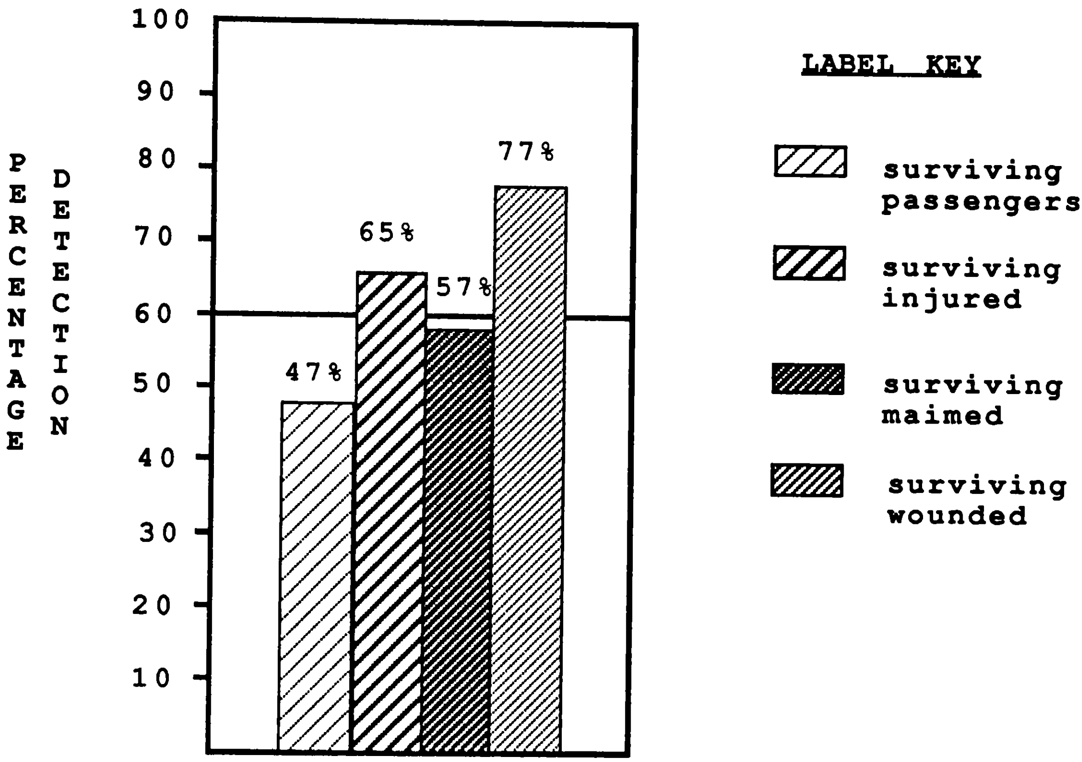
Of the 90 subjects tested, 27 had previous knowledge of the problem and their detection status was excluded from the results. Despite having encountered the problem before, 6 of this total failed to detect the anomaly. The number of subjects detecting and non-detecting in each condition is presented, with the percentage detection rate, in Table 5.1 below. The percentage detection rates are also presented in diagrammatic form in Figure 5(a), superimposed over a line representing the baseline detection rate for the **survivors** condition.

TABLE 5.1: Detection rates for Experiment 2.

Condition	**Detections**	**Non-detections**	**Percentage detection**
surv.pass	9	10	(47%)
surv.inj	11	6	(65%)
surv.maim	8	6	(57%)
surv.wound	10	3	(77%)
All conditions	38	25	(60%)
Subjects with previous knowledge			
surv.pass	7	4	
surv.inj	2	1	
surv.maim	6	0	
surv.wound	6	1	
All conditions	21	6	

Although the overall detection rate of 60% is double that of Experiment 1, it is lower than was expected. There are no reliable differences in rate of detection across conditions (Chi-square=3.026, d.f.=3, NS). Notably, the **surviving passengers** condition failed to show <u>any</u> increase on the corresponding rate for **survivors**. In fact, there is a numerical decrease. The other manipulations all show an increase on the unqualified items. These contrasts will be evaluated in the discussion. As before, non-detections were literally non-detections. These subjects were not aware of reading about anything other than dead people.

FIGURE 5(a): Detection rates for Experiment 2.



{2} Anomalous sentence dwell-times and reading speeds

The mean dwell-times for the anomalous sixth sentence are presented in Table 5.2, and the mean reading speeds (sum of the dwell-times on sentences two, three, four and five) are presented in Table 5.3. As in Experiment 1, both the reading speeds and dwell-times look to be longer for detectors, but statistical analysis and discussion of these possible effects will be left to a general analysis in the following chapter.

TABLE 5.2: Anomalous sentence dwell-times for Experiment 2.

Dwell-Times (milliseconds)			
Condition	**Means for Detectors**	**Means for Non-detect**	**Total Means**
surv.pass	9645	4440	7043
surv.inj	14256	3257	8756
surv.maim	8476	3260	5868
surv.wound	9765	6837	8301
Total	10765	4161	8144

TABLE 5.3: Reading speeds for Experiment 2.

Reading Speeds (seconds)			
Condition	**Means for Detectors**	**Means for Non-detect**	**Total Means**
surv.pass	15.3	14.0	14.6
surv.inj	20.0	9.4	16.2
surv.maim	15.8	10.5	13.5
surv.wound	15.9	13.4	15.3
Total	16.9	12.0	15.0

{3} Content analysis for the written solutions

The same analysis procedure was carried out as before using the same independent judges. The concordant evaluations for solution-type are presented in Table 5.4, and the corresponding analysis for reference-expression in Table 5.5. A transcript of the solutions for those subjects contributing to the detection rate is presented in Appendix C.

TABLE 5.4: Content analysis for solution-types in Experiment 2.

	Solution-type				
Condition	**Family**	**Home**	**Crash**	**No prob**	**Other**
surv.pass					
9 detectors	2	-	1	5	1
10 non-detect	3	3	-	-	4
surv.inj					
11 detector	-	-	-	10	1
6 non-detect	2	3	-	-	1
surv.maim					
8 detectors	1	2	-	5	1
6 non-detect	2	3	1	-	3
surv.wound					
10 detectors	1	-	-	5	1
3 non-detect	2	1	-	-	-
Totals					
(38 detectors)	(4)	(2)	(1)	(25)	(4)
(25 non-detect)	(9)	(10)	(1)	(-)	(8)

A similar pattern of solution-types emerged as in Experiment 1. Detectors mainly assert that no problem exists, and although there were 11 concordant evalutions for pragmatic inferences for subsequent death (7 more than in Experiment 1), this came from a total of 36 evalutions (in comparison with 16 in Experiment 1). Hence, 70% of concordant evaluations still found that no problem existed, in comparison with 75% in Experiment 1. The predominant interpretation of detectors is that the description is anomalous. Non-detectors, though fewer in number to Experiment 1, display a similar pattern of solutions. As before, most concentrated on involving the (dead) passenger's homes and families. As expected, no non-detector asserted that no problem existed.

TABLE 5.5: Content analysis for reference-expressions in Experiment 2.

Condition	Dead	Bodies	General	Anomalous	Other
			Reference-expression		
surv.pass					
9 detectors	-	-	-	4	1
12 non-detect	3	1	1	-	-
surv.inj					
11 detectors	-	-	-	9	1
6 non-detect	4	1	1	-	-
surv.maim					
8 detectors	-	1	1	5	1
6 non-detect	-	4	1	1	1
surv.wound					
10 detectors	1	-	1	5	2
3 non-detect	1	2	-	-	-
Totals					
(17 detectors)	(1)	(1)	(2)	(23)	(5)
(40 non-detect)	(8)	(8)	(3)	(1)	(1)

The pattern for reference-expressions is also similar to that of Experiment 1. Twenty-three of the 32 concordant evaluations for detectors used the anomalous label, the others presumably making pragmatic inferences for subsequent death after detection. Most non-detectors referred to the 'dead' and the 'bodies', with one instance of a non-detector carrying the anomalous label to their solution (again without being aware of doing so). This (rather amusing) solution is given in sentence (1).

(1) Bury the survivors in both territories depending on where each landed - to be fair.

d: Discussion

Separate analyses were carried out on the detection rate contrasts between Experiments 1 and 2. The surface-complex manipulation of **surviving passengers** from **survivors** showed no reliable difference (Chi-square=0.474, d.f.=1, NS). Two of the semantic-based manipulations showed highly significant increases on the original manipulations. These were **surviving injured** (Chi-square=11.543, d.f.=1, p<0.01) and **surviving wounded** (Chi-square=6.74, d.f.=1, p<0.01). **Surviving maimed** showed no reliable detection increase over **maimed** (Chi-square=3.238, d.f.=1, NS), and was not above the basic detection rate for **survivors**.

These results are difficult to interpret. Clearly, the surface-complex manipulation has had no effect, and the semantic-based manipulations show large numerical increases that are highly reliable for two of the conditions. Restricting the inferential fields for the basic noun by semantic-emphasis <u>can</u> reduce semantic surrogation substantially, but this result does not hold statistically for all manipulations. Even detection rates of 65% and 77% are somewhat smaller than expected. Why should the detection rates have come out lower than expected, particularily in the **surviving passengers** condition?

For some reason the focusing power of the expressions seems to be attenuated. Perhaps even these manipulations have insufficient power, under present conditions, to bring 'aliveness' into focus and result in detection. One hypothesis relies upon there being some kind of top-down testing present. Assuming that the inferential fields for <u>both</u> the noun **injured** and the qualifier **surviving** are available to the processor, mapping might take place **on the basis of evidence from the injured component only**. If the significance of the composite expression is not fully computed prior to interpretation, then surrogation might proceed **irrespective of the relationship between the two components**. This would be particularily plausible if the processor was <u>actively searching</u> for information consistent with global destruction. The argument is not that the qualifier is not read. Rather, attention is not paid to it in processing, and the full meaning of the composite expression may not always be given the oppurtunity to make an impact. This clearly does not happen in all instances, but it may account for some of the non-detections.

A similar argument can be presented for the failure of the more surface-complex **surviving passengers** to increase focus on alive people. If the significance of the whole adjectival phrase is not necessarily computed prior to interpretation, then the level of surrogation effected by the **survivors** inferential fields may be maintained by those fields signalled by the **surviving** qualification. If reference to passengers is expected anyway, then the noun **passengers** may only serve as confirmation, and the qualifier may have an independent effect.

In summary, if increased focusing power depends on a description's surface complexity, and this is conditional on its full manifestation as a meaningful item, then that effect may be attenuated where such composition is not completely established. Another way of thinking about this hypothesis is that if shallow processing is occurring at a particular point in a text, then descriptions at that point are predisposed to have a weaker impact than would otherwise be the case. In the present case, the speculation is that shallow processing may be a consequence of a build-up in interpretative constraints. This top-down processing may instigate active searches for general classes of information. It was with this hypothesis in mind that Experiment 3 was undertaken.

5.3 Experiment 3: Further emphasis-based manipulations

a: Introduction

Experiment 3 was undertaken to test the boundary conditions of detection under emphasis manipulation. Only two conditions were tested in this experiment. The first attempted to further increase the surface structure description on the set of survivors by using the relative clause form **passengers who survived**. The interest was whether this phrase would have more focusing power than the adjectival phrase **surviving passengers** which failed to increase detection rate in Experiment 2. Clearly, it would be expected that a description with this level of surface structure would have sufficient impact on the representation to increase the detection rate, even though there is reason to suspect that the present narrative conditions may be optimal for non-detection.

The second condition in Experiment 3 tested the notion that in processing a composite phrase, the processor can access the inferential fields of separate parts of the description. The test for this was to include a phrase that is anomalous in itself, but has a component which is highly consistent with notions of global destruction. This item was **surviving dead**. Hence, if the significance of composite phrases were necessarily worked out prior to interpretation, readers would detect this description and, presumably, have difficulty interpreting it. They may attempt a contrast with 'non-surviving' dead: perhaps the **surviving dead** are the dead whose bodies do not become mutilated, or something of that sort? But clearly we would expect detectors to be confronted with a considerable challenge in searching for a viable interpretation for the description. Any significant degree of non-detection under this manipulation would suggest that the noun **dead** (perhaps independent of its qualifier) was having a direct confirming effect on the 'preferred' interpretation.

b: Method

The materials, design, subjects, apparatus and procedure were identical to those for Experiments 1 and 2. The only difference in design was that the terms **passengers who survived** and **surviving dead** were substituted as items in the **survivors** position of the standard aircrash narrative. These constituted the two conditions of the experiment, tested on independent groups of subjects. A total of 51 subjects was tested.

c: Results

{1} Detection rates

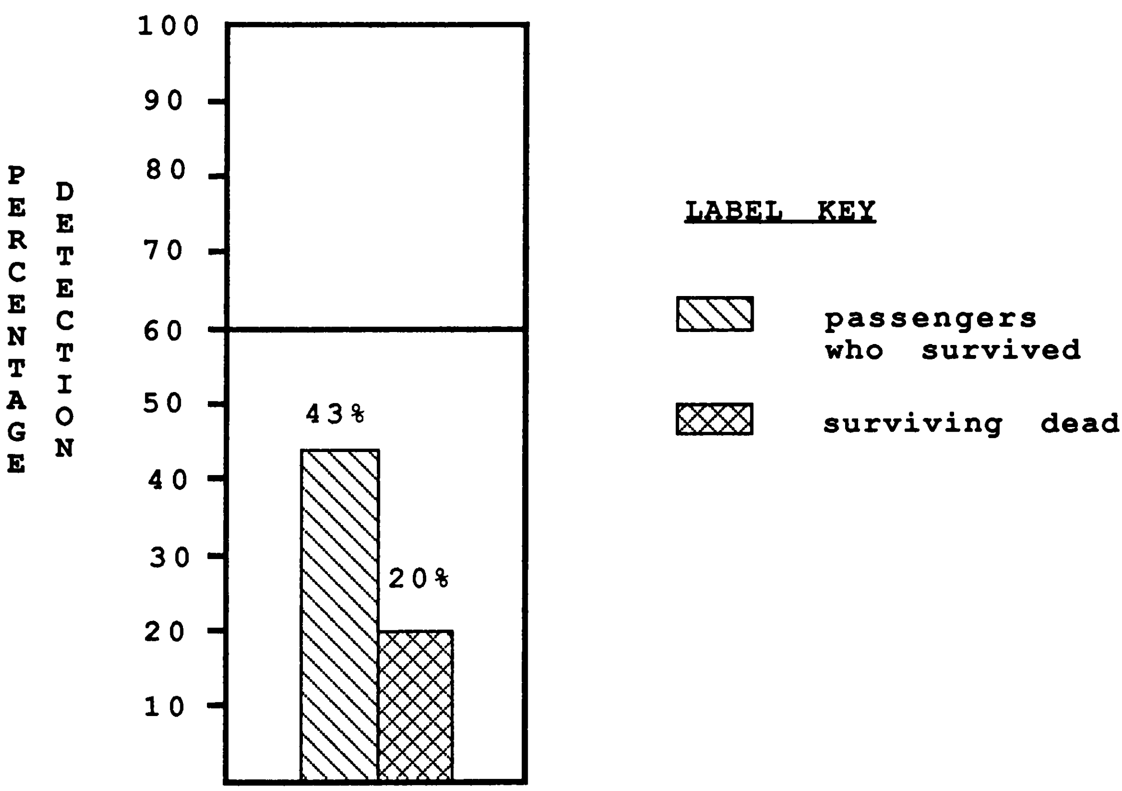
Of the 51 subjects tested, 10 had previous knowledge of the problem and their detection status was excluded from the results. Despite having encountered the problem before, half of this total failed to detect the anomaly. The number of subjects detecting and non-detecting in each condition is presented, with the percentage detection rate, in Table 5.6. The percentage detection rates are also presented in diagrammatic form in Figure 5(b), superimposed over a line representing the baseline detection rate for the **survivors** condition.

A combined detection rate of only 32% approaches the original low level of 30% in Experiment 1. Far from increasing detection, these manipulations have shown numerical decreases on the corresponding rates from Experiment 2. A detection rate of only 20% for the **surviving dead** manipulation is absolutely astonishing. Again, non-detecters were not aware of having read about anything other than dead people. These contrasts will be investigated further in the discussion. There was no reliable difference in detection rate between conditions (Chi-square=2.469, d.f.=1, NS).

TABLE 5.6: Detection rates for Experiment 3.

Condition	**Detections**	**Non-detections**	**Percentage detection**
pass.who.surv	9	12	(43%)
surv.dead	4	16	(20%)
Both conditions	13	28	(32%)
Subjects with previous knowledge			
pass.who.surv	4	5	
surv.dead	1	0	
Both conditions	5	5	

FIGURE 5(b): Detection rates for Experiment 3.



{2} Anomalous sentence dwell-times and reading speeds

The mean dwell-times for the anomalous sixth sentence are presented in Table 5.7, and the mean reading speeds (sum of the dwell-times on sentences two, three, four and five) are presented in Table 5.8.

TABLE 5.7: Anomalous sentence dwell-times for Experiment 3.

Dwell-Times (milliseconds)			
Condition	**Means for Detectors**	**Means for Non-detect**	**Total Means**
pass.who.surv	7780	4659	6219
surv.dead	13097	4764	8930
Total	9416	4719	7542

As before, both the reading speeds and dwell-times appear to be longer for detecting subjects. Statistical analysis and evaluation of these possible effects will be included in the following chapter.

TABLE 5.8: Reading speeds for Experiment 3.

Reading Speeds (seconds)			
Condition	**Means for Detectors**	**Means for Non-detect**	**Total Means**
pass.who.surv	13.7	12.8	13.2
surv.dead	14.6	12.6	13.0
Total	14.0	12.7	13.1

{3} Content analysis for the written solutions

The same analysis procedure was carried out as before using the same independent judges. The concordant evaluations for solution-type are presented in Table 5.8, and the corresponding analysis for reference-expression in Table 5.9. A transcript of the solutions for those subjects contributing to the detection rate in both conditions is included in Appendix D.

TABLE 5.9: Content analysis for solution-types in Experiment 3.

	Solution-type				
Condition	**Family**	**Home**	**Crash**	**No prob**	**Other**
pass.who.surv					
9 detectors	1	1	-	7	-
12 non-detect	2	8	-	-	1
surv.dead					
4 detectors	1	-	-	-	-
16 non-detect	6	4	-	-	2
Totals					
(13 detectors)	(2)	(1)	(-)	(7)	(-)
(28 non-detect)	(8)	(12)	(-)	(-)	(3)

The pattern of solution content is also as before. Seventy percent of concordant evaluations for detectors asserted that no problem existed, and 20 out of 23 concordant evaluations for non-detectors concentrated on involving the passenger's families or homes.

TABLE 5.10: Content analysis for reference-expressions in Experiment 3.

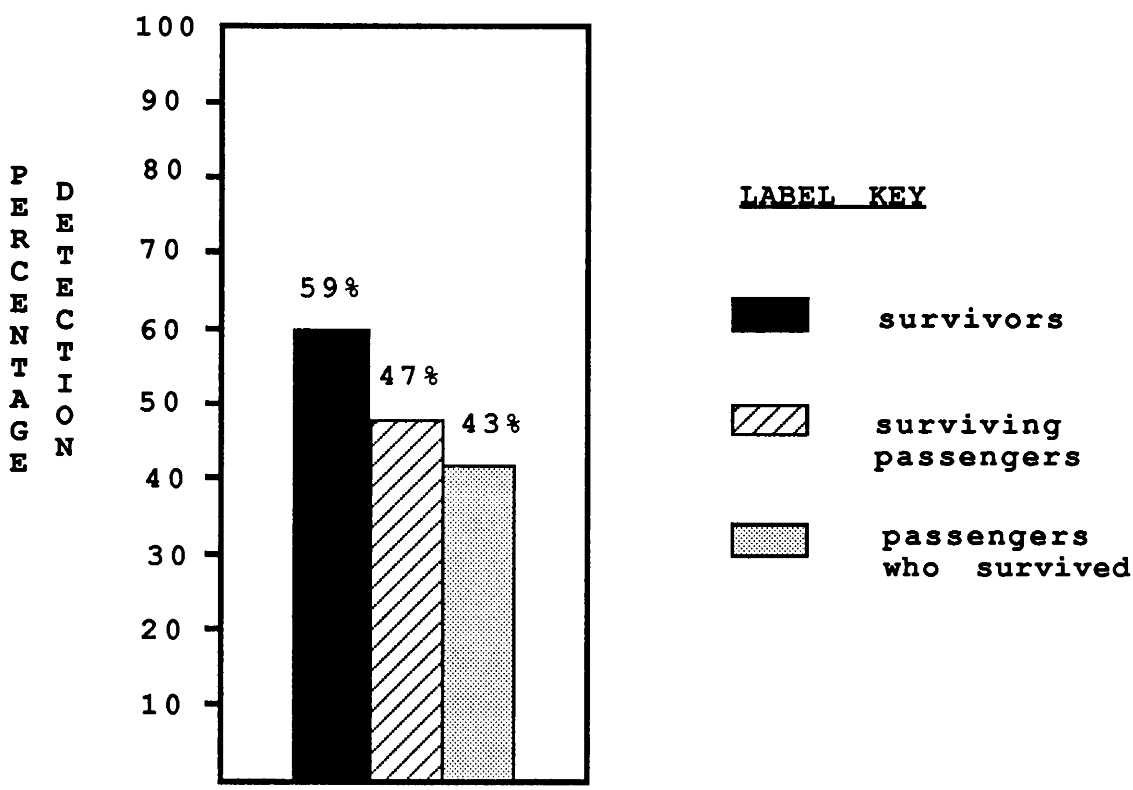
	Reference-expression				
Condition	**Dead**	**Bodies**	**Gener-al**	**Anoma-lous**	**Other**
pass.who.surv					
9 detectors	1	-	-	3	2
12 non-detect	-	2	6	-	2
surv.dead					
4 detectors	1	-	-	1	-
16 non-detect	4	1	2	-	5
Totals					
(13 detectors)	(2)	(-)	(-)	(4)	(2)
(28 non-detect)	(4)	(3)	(8)	(-)	(7)

There were only 8 reference-expression evaluations for detectors and half of these used the anomalous label. For the non-detectors, no subject carried the anomaly to their solution, and of the 22 concordant evaluations, 7 referred to dead/bodies, and 8 used a general description. This general form of description seemed particular ly evident in the **passengers who survived** condition, suggesting that the general nature of the text-based description had effected the form of description chosen for the solution. This may have occured despite suspicion that the full meaning of composite phrases is not necessarily computed prior to interpretation.

d: Discussion

The detection rate for **passengers who survived** shows a numerical decrease on the **survivors** and **surviving passengers** manipulations. These rates are illustrated in Figure 5(c). Neither of the surface complex manipulations have had any greater focusing impact on the representation. Separate analysis of these three conditions shows no reliable difference in detection rate (Chi-square=0.996, d.f.=2, NS). For whatever reason, the cohesion-establishing mechanisms seem to go unaffected by these local focus manipulations.

FIGURE 5(c): Comparison of detection rates for the surface structure manipulations.



The most surprising result is for the **surviving dead** manipulation. In this condition, fully 80% of the subjects tested were only aware of reading about dead people, and the local anomaly, **surviving dead**, passed un-detected. It was not the case that these readers could not work out what the qualification meant, or that they somehow found a contrast with 'non-surviving dead' individuals. Non-detectors were simply not aware of their having been a qualification for the **dead** at all. The only explanation the author can find for this is that the **dead** noun is has had a direct effect on interpretation **independent of its combined significance**

with the qualifier. The magnitude of this effect is so great that one suspects some type of local top-down processing initiating an active search for 'preferred' or expected information. This is consistent with the view that the expected information contains some general notion of destruction. It is possible, then, that the low detection rates for all the semantic-based manipulations reflect patterns of cohesion under conditions of maximal prediction, rather than any inherent incapacity on the part of the items to manipulate focus.

The argument in this thesis is certainly not that reader's expectations about what is coming next are so high that they do not read the words in the text, although it is always possible that this could happen. Some early pilot work on a small number of subjects tested the basic view that readers could be so certain about burying 'the dead' that the relevant nounphrase was not even read. The manipulation tested was where the authorities should bury the **hatchet**. All subjects noticed the label and attempted interpretation on the basis of the metaphorical meaning, mainly concerning possible grudges with the involved parties.

All these results suggest subtle patterns of top-down and bottom-up interactions. The expectations are not so high that any old lexical item will satisfy the test, but rather the build-up in interpretative commitment effects some kind of bias towards preferred classes of information. Of course, the ramifications of these views are not restricted to non-detection of anomaly. Rather, undetected anomalies act as signals for processing configurations that are as basic to successful interpretation. So if one were to read the narrative text concerning **passengers who died**, there may be no sign of incomplete processing, but the components licensing the relevant mappings may be doing so on the basis of some general, shallow type of test. Cohesion for the phrase <u>as a unit</u> may not even be attempted.

A general overview of the first three experiments will form the content of the following chapter.

Chapter Six

A General Analysis of Experiments 1, 2 and 3

6.1 Summary

The first part of this chapter summarises the results for the ten narrative-based conditions tested in Experiments 1, 2 and 3, and carries out statistical analysis on the reading speed and dwell-time results. There is discussion both of these results and other patterns emerging from the detection rates and solution content evaluations. The second part of the chapter assesses the emerging characteristics of cohesion-establishment in view of these findings.

6.2 A general analysis of the narrative-based manipulations

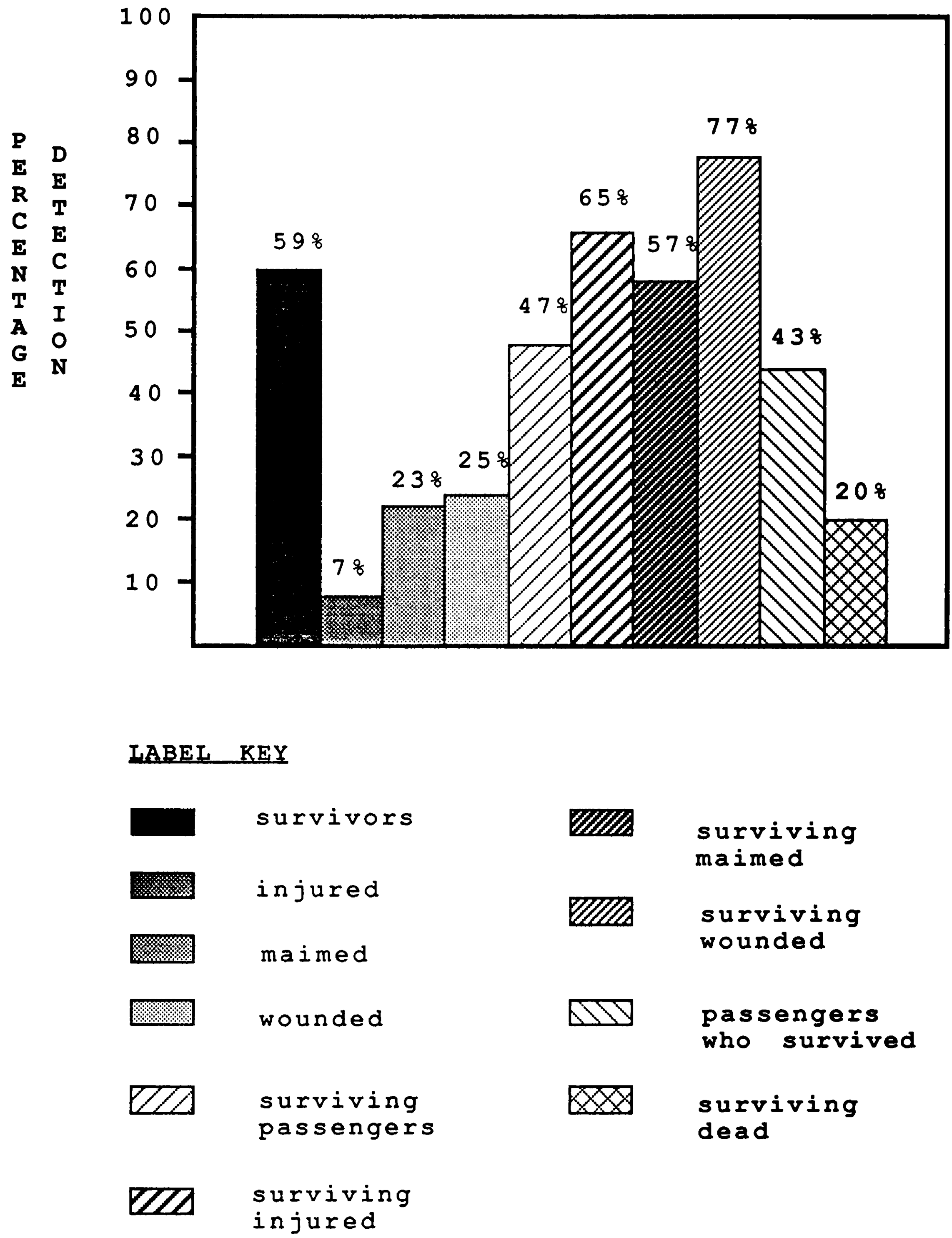
a: Detection rates

Table 6.1 details the detection rates for all ten narrative-based conditions tested in Experiments 1, 2 and 3. Figure 6(a) illustrates these detection rates in percentage format.

TABLE 6.1: Detection rates for all ten narrative-based conditions.

Condition	Detections	Non-detections	**Percentage detection**
survivors	10	7	(59%)
injured	1	14	(7%)
maimed	3	10	(23%)
wounded	3	9	(25%)
surv.pass	9	10	(47%)
surv.inj	11	6	(65%)
surv.maim	8	6	(57%)
surv.wound	10	3	(77%)
pass.who.surv	9	12	(43%)
surv.dead	4	16	(20%)
All conditions	68	93	(42%)

FIGURE 6(a): Percentage detection rates for all ten narrative-based conditions.



Not surprisingly, there is a highly reliable influence of condition on detection rate (Chi-square=29.07, d.f.=9, $p<0.01$)

b: Analysis of dwell-times and reading speeds

Table 6.2 lists the dwell-times for all ten narrative-based conditions, and Figure 6(b) illustrates these in diagrammatic form.

TABLE 6.2: Anomalous sentence dwell-times for all ten narrative-based conditions.

Dwell-Times (milliseconds)

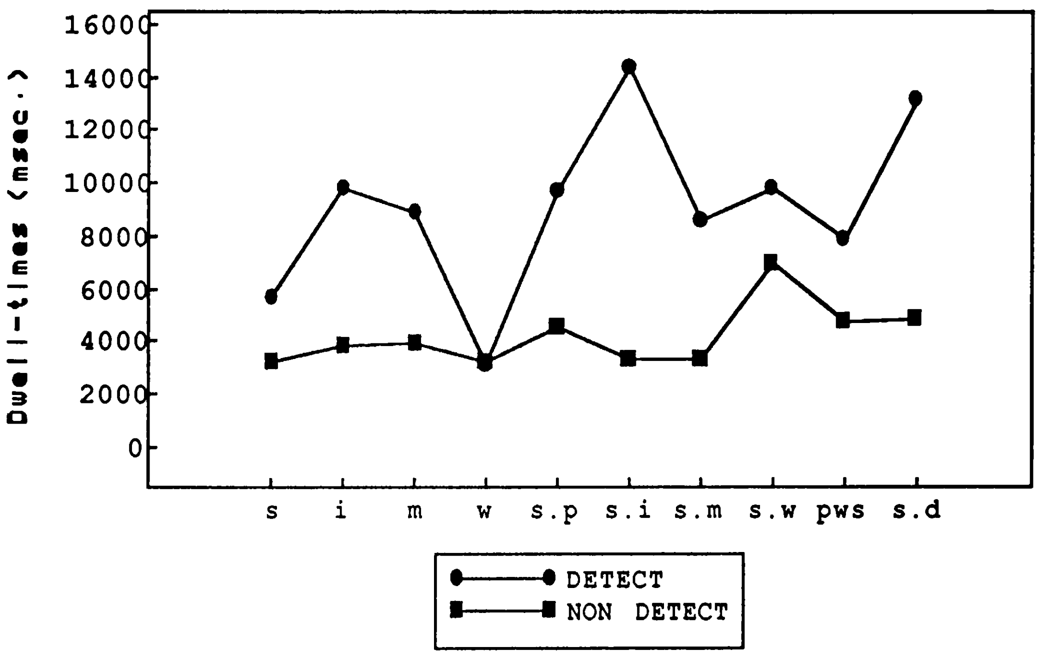
Condition	**Means for Detectors**	**Means for Non-detect**	**Total Means**
survivors	5595	3125	**4360**
injured	9727	3729	**6728**
maimed	8766	3886	**6326**
wounded	3003	3125	**3064**
surv.pass	9645	4440	**7043**
surv.inj	14256	3257	**8756**
surv.maim	8476	3260	**5868**
surv.wound	9765	6837	**8301**
pass.who.surv	7780	4659	**6219**
surv.dead	13097	4764	**8930**
Total	9011	4108	**6179**

In nine of the ten conditions detectors have longer latencies suggesting a possible effect of detection on dwell-time. A two-by-ten analysis of variance was carried out on the dwell-time means. The results of this analysis are shown in Table 6.3.

TABLE 6.3: Analysis of variance for dwell-times in all ten narrative-based conditions.

Source of Variation	**Degrees of Freedom**	**F ratios**	**Probability values**
Detection	1	20.191	0.0000
Conditions	9	1.174	0.3162
Interaction	9	0.827	0.5928

FIGURE 6(b): Diagram of means for dwell-times in all ten narrative-based conditions.



The corresponding means for reading speeds are presented in Table 6.4, and illustrated in Figure 6(c). Again, nine out of ten conditions show a larger latency for detectors. A two-by-ten analysis of variance was carried out on these means and the results of this are shown in Table 6.5.

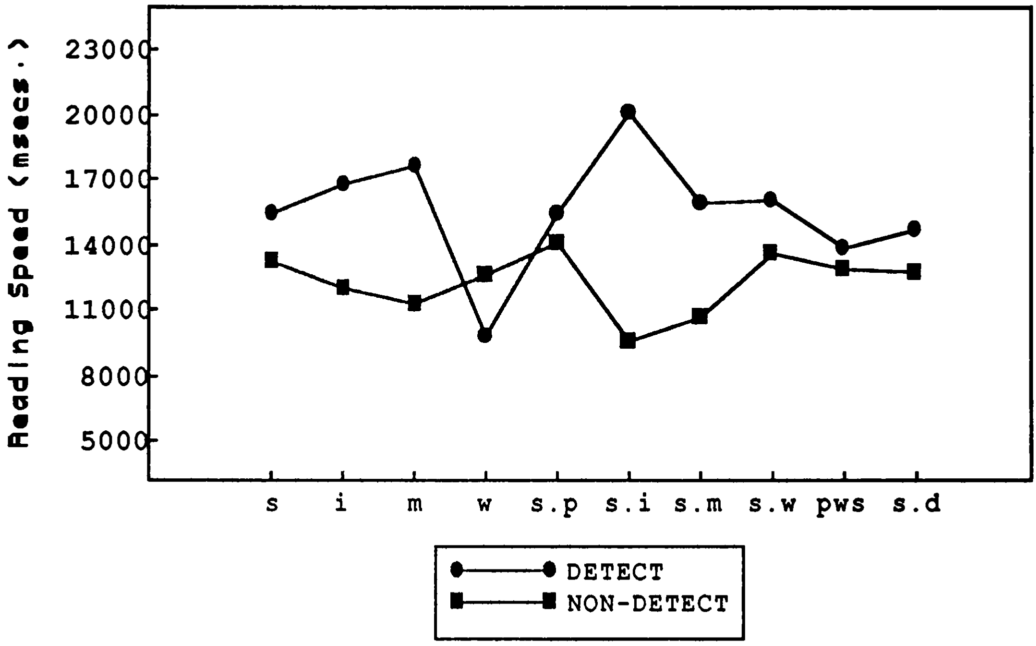
TABLE 6.4: Reading speeds for all ten narrative-based conditions.

Reading Speeds (seconds)			
Condition	**Means for Detectors**	**Means for Non-detect**	**Total Means**
survivors	15.3	13.1	**14.2**
injured	16.7	11.8	**14.3**
maimed	17.5	11.1	**14.3**
wounded	9.7	12.5	**11.1**
surv.pass	15.3	14.0	**14.6**
surv.inj	20.0	9.4	**16.2**
surv.maim	15.8	10.5	**13.5**
surv.wound	15.9	13.4	**15.3**
pass.who.surv	13.7	12.8	**13.3**
surv.dead	14.6	12.6	**13.0**
Total	15.4	12.1	**13.5**

TABLE 6.5: Analysis of variance for reading speeds in all ten narrative-based conditions.

Source of Variation	Degrees of Freedom	**F ratios**	**Probability values**
Detection	1	8.905	0.0034
Conditions	9	0.398	0.9344
Interaction	9	1.079	0.3817

FIGURE 6(c): Diagram of means for reading speeds in all ten narrative-based conditions.



Both sets of results show highly significant effects for the detection source of variation. The probability value is especially low for the dwell-time analysis. There are no effects for conditions, and no interaction effects in either analysis. From inspection of the means it is clear that slower readers over the four sentences prior to the anomalous one are more likely detect the anomaly. Also, detectors dwell significantly longer on the anomalous sentence.

These are both interesting results. The effect for dwell-times could be reflecting part of the reading speed outcome that, for whatever reason, slower readers on the sixth sentence are more likely to detect the description. The difference in means for dwell-times across detection-type is so great, however, that it seems likely that there is also considerable effect due to re-reading the sentence after detection of the anomaly. Indeed, instances occurred when detectors dwelled on the sixth sentence for over 20 seconds, and in one case a dwell-time of 58 seconds was recorded. Some detectors report re-reading the sentence after detection having failed to establish an interpretation, often unable to 'believe their eyes' at what they had read!

The reading speed effect presents a range of interesting issues. First, it is possib le that reading speed predicts the likelihood of detection across description-types. However, we have already observed that there is no overall interaction between detection and conditions. Correlations computed for the relationship between reading speeds and detection rate in each condition were as low as r=0.17 (for reading speed <u>means</u>) and r=0.08 (for reading speed <u>medians</u>). This suggests that reading speed only accounts for an extremely small percentage of the detection variance (at best fractionally less than 3%). Hence, although speed of reading has an overall effect on detection rate, it does not predict the rate for particular conditions. Clearly then, its effect does not operate directly at the level of the manipulations which have been incorporated within these experiments.

Given this, what sort of effect is reading speed having? It is known that faster readers make fewer fixations per line and tend to pay less attention to shorter words in a text (Rayner, 1978), but it is not clear in this case how such differences may be interacting with interpretative activity. For instance, could it be that tighter cohesion varies monotonically as a function of fixation duration? At one level it seems trivially true that information which is closely attended to will be processed in a more thorough way. But the possibility exists that subjects detect, not necessarily because they are reading the text more slowly (and as a consequence in greater depth), but because their reading speed pattern <u>correlates</u> with comprehension machinery that effects cohesion. Doing something fast means that a large number of operations are carried out over unit-time. The nature of these operations may or may not change as an outcome of pace variation. Perhaps, then, slower readers forced to read the narrative at a faster pace than their usual may still be more likely to detect the true description. By the same rationale, faster readers may not be helped by having the text presented to them at a rate slower than their normal. Although this was a testable hypothesis, effort was concentrated on expanding the conditions under which detection rates were tested.

The fact that reading speed has an overall effect, but does not predict detectibility of particular items, might suggest that there is at least some separation between the fixation duration factors and depth of processing at the lexical level. Of course, fixation duration itself may not be the most important consequence of fast reading. At one extreme of the reading speed 'spectrum', there are speed-reading techniques which encourage fewer fixations with the aim of decreasing global reading time. Although proponents of such techniques claim that speed can increase without effecting comprehension levels, the measurement of such levels is often done in very general terms, not accounting for the possibility of massive elaborative inferential activity in filling the gaps in information lost as a result of minimising fixations. Speed of reading may be so fast that comprehesion activity is compromised. Conversely, the pace may be so slow that it resembles exhaustive analysis rather than reading. The latencies in the present experiements were clearly within these boundaries.

c: Solution content

The total content evaluations for solutions in all conditions are presented in Table 6.6 for solution type, and Table 6.7 for reference expression. The solution-type patterns in Experiments 1, 2 and 3 are confirmed by the general pattern observed after summing over all conditions. Just over 70% of concordant evaluations for detectors find that no problem exists. For the non-detectors, no subject makes this assertion, and over 75% involve the (dead) passenger's homes and families.

TABLE 6.6: Content analysis for solution-types in all ten narrative-based conditions.

	Solution-type				
Condition	**Family**	**Home**	**Crash**	**No prob**	**Other**
All ten narrative-based conditions					
68 detectors	6	6	1	44	5
93 non-detect	37	37	5	-	17

Detectors predominantly use the anomalous label when they refer to the passengers (accounting for 68% of concordant evaluations). The non-detectors references are more evenly spread across the types, with a small number carrying the anomalous label to their solution unawares (accounting for about 5% of concordant evaluations). References to the dead, deceased, bodies, remains, etc. constitute 67% of evaluations.

TABLE 6.7: Content analysis for reference-expressions in all ten narrative-based conditions.

	Reference-expression				
Condition	**Dead**	**Bodies**	**Gener-al**	**Anoma-lous**	**Other**
All ten narrative-based conditions					
68 detectors	5	2	2	39	9
93 non-detect	23	22	19	4	14

6.3 Characteristics of cohesion-establishment

a: 'Sub-atomic' mapping

(1) Inferential fields and patterns of surrogation

One main finding from all the results is that **the surrogation phenomenon cannot be simply explained with sole reference to amounts of over-lap in semantic features**. The claim is that surrogation patterns are reflected in the detection rates, and that the mechanics of surrogation reflect underlying cohesion-based processes. The patterns of surrogation, particularily arising from Experiment 1, support the view that the mapping from semantic input to background model operates on the basis of fields of relations signalled by the incoming items. Although these are very difficult to characterise in a meaningful way, they clearly cannot reside in any pre-defined lexical definition of a word, since this implies infinite listing of possible inferential relations for that word.

Surrogation can occur because the mechanisms operating on a proposed relation cannot carry out exhaustive checks on it. This is because it is not clear what finite set of things would have to be checked. In this sense, **cohesion-based activity cannot be entirely complete, but only more or less so depending on the depth of processing on any particular relation**. The main issues are what text-based configurations control the focus of the processor onto some relations but away from others, and on what basis the processor displays satisfaction with a set of bindings as an interpretation of a message.

Two points of evidence suggest that surrogation occurs partly because interpretative activity can operate autonomously from other processes or memory traces. First, a small number of non-detectors carry the anomalous label to their written solution still unaware of its inconsistency with their interpretation. Clearly, if these subjects are representative of all readers, the <u>words</u> can be retained in some kind of memory trace, but this must be separate from the trace for the working interpretation. The former can obviously be accessed when the interpretative representation is put to use (in this case for solving a problem) and the anomalous label can be used in production still un-detected. On one occasion, a subject with previous knowledge of the problem failed to detect the anomaly, but in his solution noticed that he had written 'wounded' rather than 'dead' and could not work out why he had done so! This relation between the trace for the words and the interpretation could, of course, be thought of as modularity within the same representation.

Second, as mentioned above, subjects who have come across the exact same problem before, and who often have vivid memories of their past encounters with it, can still fall into the 'surrogation trap'. Although the vast majority with previous knowledge spot the anomaly during the experiment, prior knowledge does not guarantee future detection. This also implies some degree of detachment from other representations or memories, this pattern reminicent of Erickson and Mattson's finding that even warning subjects about anomalies failed to increase detection rate. In some of the present author's early pilot work, more than one of the experimental texts contained an anomaly, and detection of the first did not guarantee that subjects would detect the second. This was despite the fact that subjects often reported looking for anomalies in the following texts once they had realised that this may be the point of the experiment.

<u>{2} Local inferential relations</u>

Sub-atomic mapping activity can bind relations that are localised within an interpretation. **This finding suggests that the checks on a particular relation are not necessarily evaluated at levels in the broader interpretation where overall consistency has to be established.** This counts as evidence for a processor that <u>could</u> operate within compositional constraints, but need not necessarily do so. Maintaining compositionality would demand a deeper form of processing than has been observed in these studies. Such depths may risk over-processing. There may be processing configurations, reflecting underlying assumptions about the structure of a message, which can deliver interpretations consistent with the text without aspiring to compositionality. These may result in outcomes which are of value to the cognitive agent (e.g. time saved, resources allocated elsewhere etc.).

{3} Global cohesion

The above comments imply some management of the various local relations some of which may be computed independently. The results from the first three experiments are consistent with the view that some mappings into a background model may be central to establishing the point of a message, or be important in evoking appropriate contextual information. A global cohesion mechanism that operated under multiple constraint satisfaction may be boosted by such mappings which could effectively slacken cohesion requirements elsewhere. Establishing significant text-to-model mappings may set up thematic constraints which then determine what information is important: that is, the things that will be brought into focus. Expectations, in the weaker sense, are formed at the level of the evoked scenario. So maintaining relevance in successive inputs is achieved providing descriptions are 'expected' or at least compatible with that model. Hence, the author's speculation is that burying the **hatchet** is reliably detected because attention is drawn to it after it fails to map easily into the background 'aircrash' scenario. The terms used in all other conditions presumably map into the model 'as expected'.

Build-up in thematic contraints may extend to the point of expectation in the strong sense, where it becomes harder than normal for successive items to alter the focus of the processor's attention. Hence, the focusing impact of incoming expressions may be attenuated under such conditions. The results presented suggest that expectations for particular inferential fields may be supported through access to the fields signalled by an item that is part of a composite expression. Although the exact conditions of interpretation are not known, **this suggests that the processor can carry out interpretative activity on information yet to be qualified as directed by the text**. This casts doubt on any model of processing that presupposes the computation of phrasal or propositional components prior to interpretation. Also, the focusing impact of a lexical item or phrase would seem to depend crucially on it being able to manifest its full inferential fields in interpretation. The present evidence suggests that this may depend on the context in which interpretation is being carried out, and the following section develops some of these themes.

b: Interpretative context

{1} Narrative structure and problem definition

Importantly, the present experiments have been carried out using a narrative text. It is known that some characters in narrative texts are assigned to subsidiary roles that render them less significant in the overall model of events. Anderson, Garrod and Sanford (1983) showed that it is harder to retrieve information about subsidiary characters after a change of episode has occurred in a narrative. Status seems to be given to main characters in narratives, often reflected by the use of proper names. Characters in supporting roles seem to be less in focus, or are in focus only in relation to a main character. Hence, in the narrative format used in the present experiments, burying the 'dead' may determine those individual's status as supporting characters in relation to the (agent) authorities. The main point is that there could be a relationship between the expected 'supporting role' of a character-description and the sort of interpretation carried out on its text-based description. *A priori*, the objects of burial may be low in status, and be less likely to be brought into focus, and less likely to exert enough focusing power to alter any expectations the processor may have about them.

This factor is compounded by the use of the narrative as a problem-description the reader is expected to solve. Hence, the prior information in sentence five that 'the wreckage was equally strewn in France and Spain' may bias the processor towards 'knowing' what the problem is going to be: feelings of predictibility about burial sites may be high. An effective enhancement in mapping the text to a problem-structure, may make it difficult for succesive items to alter that structure. These contributary effects may be part of the build-up in constraints that makes it difficult for the final nounphrase in sentence six to exert the focusing impact due to it. The possibility that this level of expectation filters down to the on-line mapping cannot be ruled out, particularily as we have already seen contextual effects operating at this level in lexical access processes (St.John, 1991).

{2} Time-course of interpretation and thematic commitment

Although it is known that, allowing for backward saccades, reading occurs in a generally left-to-right fashion, it does not follow that interpretation follows in the same manner. We have already noted that separate memory traces may exist for the explicit text content and the interpreted model, detail in the former suspected to fade after comprehension (c.f. Sachs, 1967). Clearly, there is no demand that the interpretation of an input occurs immediately as the relevant word or phrase is recognised. Perhaps in resolving a sentence some phrases are put on hold and then interpreted in light of other mappings. This would be consistent with the view that some parts of a text evoke rich background structure providing the basis for the interpretation of other items (e.g. verbs signalling underlying schemata, contextually rich items evoking background scenarios). Time-course of commitment to interpretation could operate at a variety of levels. The same rationale applies in the resolution of a narrative as to the resolution of a sentence. The usual narrative structure of beginnings, middles and endings with associated sub-events has presumably evolved in order to make interpretation and commitment to the theme easy to establish (Rumelhart, 1975; Johnson-Laird, 1983). Readers know what to expect. Narratives that do not conform to this pattern are more difficult to understand.

Certain types of interpretation, such as pronoun resolution, seem to occur quite rapidly (Sanford and Garrod, 1989), but the time-course of interpretation for other items may not be so fast. This may be particularily true for some phrases that take roles within verb structures, where the signalling verb has yet to appear in the sentence. How this relates to cohesion is not clear. It is possible that preliminary levels of cohesion have to be established on successive inputs in a more-or-less incremental fashion, but that final cohesion for the sentence may only be resolved when more information has arrived. The issue of partial cohesion is raised again.

A limitation in the first three experiments has been the suspicion that interpretative commitment may be so great that active searches for compatible information come into play. **From these results it is not decidable whether semantic surrogation and its related effects are conditional on top-down processes**. Further, the range of manipulations varying time-course factors is limited within the narrative domain. Because time-course may be a function of focus, this forms an important domain for the present study. As such, a different paradigm was developed in an attempt to carry forward the issues raised by the narrative-based experiments. This paradigm returned to the original question-construction for the anomaly described in Chapters 2 and 3. The original example is presented in sentence (1)

(1) When an aircraft crashes where should the survivors be buried?

Should surrogation occur under these conditions, then it would clearly not be a simple function of top-down searches resulting from thematic commitment. A single sentence message does not provide enough content for that level of commitment to develop. Further, manipulations in question-construction, as well as variation in passenger description, are easy to achieve and test. Research completed using this format constitutes most of the remainder of the case-study. This work will be presented in Chapters 7, 8 and 9.

Part Three

Chapter Seven

The Questionnaire Format: Experiments 4 and 5

7.1 Summary

A questionnaire-based format for the anomaly is introduced. Using this format, Experiment 4 incorporates the basic manipulations from Experiment 1 and surrogation is still found to be prevalent. Non-detection occurs despite the absence of thematic build-up in the new test materials. Clearly, shallow processing is not conditional on local top-down searches for preferred information. Experiment 5 re-tests some of the emphasis-based manipulations, and reliable detection increases are observed on the corresponding narrative ratings. This result supports the view that attenuated focusing impact has occurred for these items under the narrative format. Discussion is made of possible factors controlling focus and depth of processing under question-construction.

7.2 Introduction to the questionnaire format

The case-study was developed further by testing detection rates in various question-construction formats of the anomaly. The standard question-construction is presented in sentence (1).

(1) When an aircraft crashes where should the survivors be buried?

In question-constructions of this type, there is reduced oppurtunity for extensive thematic and interpretative commitment to be made prior to encountering the anomaly. Indeed, in (1) the anomalous label occurs <u>before</u> the verb in the sentence. Any significant degree of surrogation under this format would suggest that processing incompleteness is not conditional on the sort of top-down expectations suspected to obtain in the narrative interpretation.

Of course, in the same way that the narrative text required the context of the problem-solving experiment, so the question-constructions demanded a context also. The context selected was a questionnaire which included questions about social and related issues. Ten questions were used in all, and the fifth of these was always the (only) anomalous one. A list of the questions is presented in Table 7.1.

TABLE 7.1: A list of the ten questions used in the questionnaire format.

1.	Should homeless people have the right to be housed in their home town?
2.	At what age should people be permitted to hold a driver's licence?
3.	Does the teacher have ultimate responsibility for a child's education?
4.	Should all citizens have the right to peaceful protest?
5.	When an aircraft crashes where should the survivors be buried?
6.	At what age should it be legal to smoke tobacco in Britain?
7.	How many MPs should there be in the British parliamentary system?
8.	After a death who officially should be informed?
9.	Should National Service be reinstated?
10.	Should students have the right to a state-funded education?

Subjects were told that they were to read and fill in a questionnaire which they were informed was part of a general survey into people's attitudes on various social issues. In the instructions they are asked to answer as best they can without seeking further information, and that it is their own opinion on the issues raised that is sought. They are also asked to work through the questions 'reasonably quickly', avoiding very complex answers. Subjects could be as brief in their answers as they wished (note-form answers were encouraged), and enough space was given below each question for an answer to be written. No indication of the true purpose of the questionnaire was given prior to the subject's participation, and subjects were left for about five-to-ten minutes to fill out their answers.

As with any reading task, the global context will have considerable effect on the sort of processing characteristics that emerge. This context was designed in order that no special attention would be drawn to the anomalous question. A general theme of social and related issues was thought to be the most appropriate. A more defined context (e.g. aviation-type issues) may not have been relevant enough to the subject's general interests, and such thematic content may have fine-tuned expectations about likely questions in an unhelpful manner for present purposes. Similarily, had the questionnaire been too general (or hypothetical) in nature, its relevance and even plausibility may have been questioned. Having completed the questionnaire, subjects were then de-briefed by the same method as the narrative-based studies, and the same records of detection-status and prior knowledge were taken.

7.3 Experiment 4: Basic manipulations

a: Introduction

The idea behind Experiment 4 was very straightforward. This was simply a test of the manipulations used in Experiment 1 (i.e. **survivors, injured, maimed, wounded**) under the conditions of the questionnaire format. The standard question-construction format was used, as in sentence (1), and the other labels were substituted for **survivors** in the appropriate conditions.

(1) When an aircraft crashes where should the survivors be buried?

If detection of the anomaly was total, or at least prevalent, then such an outcome would support the view that surrogation patterns and related effects were specific to the sort of interpretative conditions in the narrative. Any significant degree of non-detection would suggest that this was not the case. As in Experiment 1, surrogation patterns could be contrasted across conditions (and also contrasted with the patterns for the narrative).

b: Method

{1} Materials and design

The materials used were questionnaires composed of two sheets of A4 size paper, stapled together. The front sheet contained instructions as described above, and listed in Appendix J. The second sheet contained a list of ten questions all on social and related issues. This list is shown in Table 7.1. The fifth question was always the anomalous one concerning the aircrash, and no other question contained an anomaly.

The were four independent conditions in this experiment, each condition reflecting a text-based manipulation. The first condition was as printed in Table 7.1, and the other three conditions were identical except for the substitution of the labels **injured**, **maimed** or **wounded** for **survivors** in the fifth question. Four independent groups of subjects saw only one version of the questionnaire.

{2} Subjects

These were mostly undergraduates at the University of

Glasgow. Most were asked to take part while at study in one of the University's libraries. A total of 140 was tested.

{3} Procedure

Subjects were informed that the questionnaire was part of a general investigation into people's attitudes to various social issues. The experimenter went over the instructions printed on the first sheet of the questionnaire, and then left the subjects to answer the questions, informing them that he would return within five minutes or so. A pen was provided for any subject who required one.

Upon his return, the experimenter asked if the subject had noticed anything strange or out of place in the questions. If the answer was no, attention was drawn to the fifth question, and its true content was pointed out to the subject. The subject was informed of the real reasons behind the study, and a record was kept of whether detection of the true content in question five (irrespective of whether or not it was evaluated to be anomalous) had occurred. A record was also kept of subjects who claimed to have previous knowledge of the 'trick'.

{4} Treatment of results

There were two areas of measurement for this experiment. First, there were the detection rates which are expressed in the usual percentage form. Second, there was the content of each subject's answer, which has been analysed with respect to both answer-type and reference-expression by the established procedure. There were no reading speed or dwell-time data under this paradigm.

c: Results

{1} Detection rates

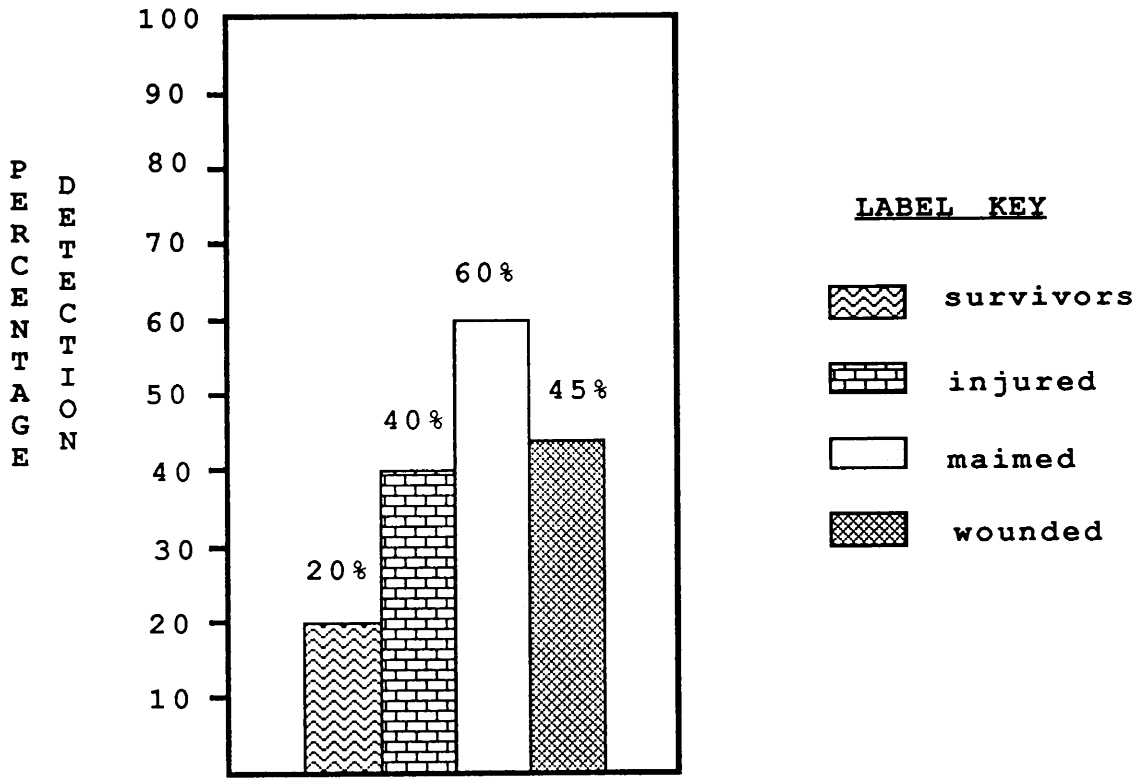
Fifty-five of the 140 subjects tested had previous knowledge of the problem and their detection status was excluded from the results. Despite having encountered the problem before, 20 of this total failed to detect the anomaly. The number of subjects detecting and non-detecting in each condition is presented, with the percentage detection rate, in Table 7.1. The percentage detection rates are also presented in diagrammatic form in Figure 7(a).

TABLE 7.2: Detection rates for Experiment 4.

Condition	**Detections**	**Non-detections**	**Percentage detection**
survivors	3	12	(20%)
injured	10	15	(40%)
maimed	15	10	(60%)
wounded	9	11	(45%)
All conditions	37	48	(44%)
Subjects with previous knowledge			
survivors	12	8	
injured	5	5	
maimed	6	4	
wounded	12	3	
All conditions	35	20	

Considering the absence of build-up in thematic constraints, an overall detection rate of 44% percent was again sursprisingly low. There can be no doubt surrounding the prevalence of surrogation under these conditions. Clearly, incompleteness in cohesion is not conditional on the sort of top-down effects thought to obtain in the narrative. There were no significant detection rate differences across conditions (Chi-square=6.275: d.f.=3, $p>0.09$).

FIGURE 7(a): Detection rates for Experiment 4.



{2} Content analysis for the written answers

The same pre-set categories of evaluation were used as before. These are shown again in Tables 7.2 and 7.3. The concordant evaluations were achieved from the ratings of the same two judges, operating under the same instructions.

The analysis for answer-types is presented, for each condition, in Table 7.4, and the corresponding analysis for reference-expressions in Table 7.5. A transcript of the solutions for those subjects contributing to the detection rate is included in Appendix E.

TABLE 7.3: Pre-set categories for content analysis of written answers: answer-type.

Five Answer-types	
1: Family	Family, relatives or friends to decide on the burial place.
2: Home	Burial to take place in the passenger's home, home country, country of origin, place of birth etc.
3: Crash site	Burial to take place at the site of the crash.
4: No problem	The answer is that no problem exists since there is no need to bury people who are still alive (e.g. 'You don't bury survivors' or 'Is this a trick question?' etc.)
5: Other	Any other answer offered.

TABLE 7.4: Pre-set categories for content analysis of written answers: reference-expressions.

Five Types of Reference-Expression	
1: Dead	(Also **deceased, victims**)
2: Bodies	(Also **remains**)
3: General	(i.e. **people/passengers/tourists**)
4: Anomalous	(i.e. **survivors**, **injured**, **maimed**, etc.)
5: Other	(e.g. pronominal references)

Note again: Since a single solution can result in more than one attribution to the pre-set categories, for some conditions the number of concordant evaluations may exceed the number of solutions. In other cases, because of the agreement constraint for concordance evaluation, the solutions may be greater in number.

TABLE 7.5: Content analysis for answer-types in Experiment 4.

	Answer-type				
Condition	**Family**	**Home**	**Crash**	**No prob**	**Other**
survivors					
3 detectors	-	-	-	3	-
12 non-detect	6	4	1	-	2
injured					
10 detector	3	1	-	4	-
15 non-detect	8	6	-	-	2
maimed					
15 detectors	4	4	1	3	-
10 non-detect	7	1	-	-	1
wounded					
9 detectors	4	-	-	3	1
11 non-detect	5	5	-	-	-
Total					
(37 detectors)	(11)	(5)	(1)	(13)	(1)
(48 non-detect)	(26)	(16)	(1)	(-)	(5)

A slightly different pattern to the narrative results emerged in the present content analysis. Although non-detector's answers displayed the same bias towards the passenger's homes and families, more detectors than before appeared to draw the pragmatic inference for subsequent death of the individuals. Only 36% of detectors asserted that no problem existed, in comparison with 75% for the corresponding labels in Experiment 1. This effect seems mainly due to the **injured, maimed** and **wounded** conditions where 18 out of 28 concordant evaluations were in categories other than the 'no problem' category. The pattern appears to be that if detection is acheived, a pragmatic inference is more likely to be made from the **injured, maimed** and **wounded** labels (since these presuppose more plausible conditions for subsequent death).

TABLE 7.6: Content analysis for reference-expressions in Experiment 4.

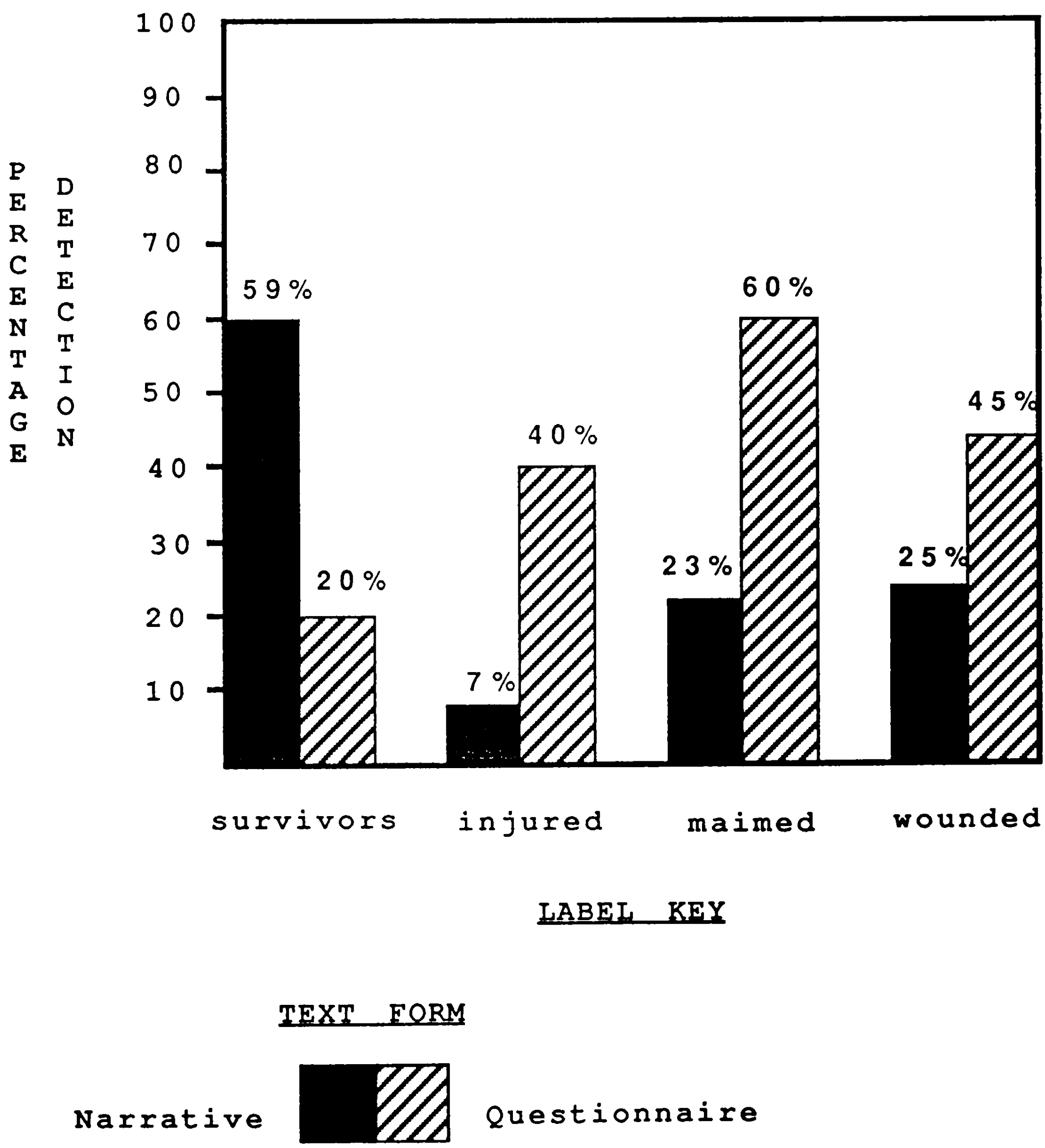
	Reference-expression				
Condition	**Dead**	**Bodies**	**General**	**Anomalous**	**Other**
survivors					
3 detectors	-	-	-	3	-
12 non-detect	-	-	1	-	3
injured					
10 detector	1	-	-	3	3
15 non-detect	1	-	-	-	1
maimed					
15 detectors	2	-	-	1	1
10 non-detect	1	-	-	-	2
wounded					
9 detectors	2	-	-	1	1
11 non-detect	1	1	-	-	3
Total					
(37 detectors)	(5)	(-)	(-)	(8)	(5)
(48 non-detect)	(3)	(1)	(1)	(-)	(9)

In comparison to the narrative results, far fewer references were made to the passengers. There were still 8 out of 18 references using the anomalous label for detectors, but for both detectors and non-detectors, the notable increase was in the number of 'other' references (44% of concordant evaluations compared with 14% in the general analysis for the narrative studies). It may be that these are mostly pronominal references, an outcome consistent with the shorter form of answer anticipated for the question responses. No non-detector carried the anomalous label to their answer.

d: Discussion

In Experiment 1, **survivors** was detected more frequently than the other labels. In this experiment, no corresponding detection bonus for this label was obtained. In fact, the <u>lowest</u> rate of 20% was received for this condition (a drop of 39% from the narrative format) and the other labels increased in rate by between 20% and 37% on their narrative rating. An illustration of these comparisons is presented in Figure 7(b).

FIGURE 7(b): Contrasts for the basic manipulations over the narrative and questionnaire formats.



A comparison of the overall detection rates across the formats (at 30% and 44% detection) showed no reliable effect (Chi-square=2.718, d.f.=1, $p>0.08$). However, separate analyses of the detection contrasts for each condition showed reliable differences for three of the labels. **Survivors** showed a relaible decrease in detection (Chi-square=4.97, d.f.=1, $p<0.05$), while both **injured** and **maimed** provided reliable increases under the question format (Chi-square=5.226, d.f.=1, $p<0.05$: Chi-square=4.68, d.f.=1, $p<0.05$). There was no reliable difference in detection rate for the **wounded** label (Chi-square=1.28, d.f.=1, NS).

A separate analysis, excluding the effect of the **survivors** label, shows a detection increase from 18% to 49% for the **injured**, **maimed** and **wounded** labels. This is highly reliable (Chi-square=10.83, d.f.=1, $p<0.001$). The increase for these labels could be due to the removal of the thematic and interpretative build-up, but why a reverse pattern should occur for the **survivors** label is not at all clear. There may be some kind of interaction between the growth of the inferential fields for particular items, and the extent to which they are effected by the presence of a top-down search.

Having observed the occurrence of considerable semantic surrogation under the question format, the natural extension to this experiment was to incorporate manipulations involving surface-structure and semantic-based emphasis. In the narrative studies, the low detection rates for these manipulations drew attention to possible top-down influences. Should the attenuated effects for these manipulations be dependent on such influences, it would be expected that detection rates would significantly increase under the present format. Experiment 5 was carried out in order to test this hypothesis.

7.4 Experiment 5: Emphasis-based manipulations

a: Introduction

This experiment tested the notion that the attenuated focusing effects observed for the emphasis-based manipulations in Experiments 2 and 3 were in part due to the presence of top-down processing conditions. A reduction in the level of build-up in interpretative commitment under the present format should give the composite expressions the oppurtunity to manifest the emphasis due to them. A selection of the previously tested items were incorporated into the present experiment. These were the **surviving passengers** which had previously failed to achieve an increase in focus over the basic **survivors** item. Second, **surviving injured** was selected as an item representative of the semantic-based manipulations of Experiment 2. Two of these had resulted in an increased rate over the unqualified expressions in Experiment 1, but with a lesser magnitude than expected. Lastly, the **surviving dead** item was tested for the apparent interpretative bias towards the meaning of the unqualified noun.

b: Method

The materials, design, subjects and procedure were the same as in Experiment 4 except that only three conditions were tested in the present experiment. Testing was carried out on a total of 59 subjects.

(1) When an aircraft crashes where should the survivors be buried?

The three conditions reflected a text-based manipulation replacing the **survivors** in the standard question-construction (1) with either **surviving passengers,**

surviving injured or **surviving dead**. Three independent groups of subjects saw only one version of the questionnaire.

c: Results

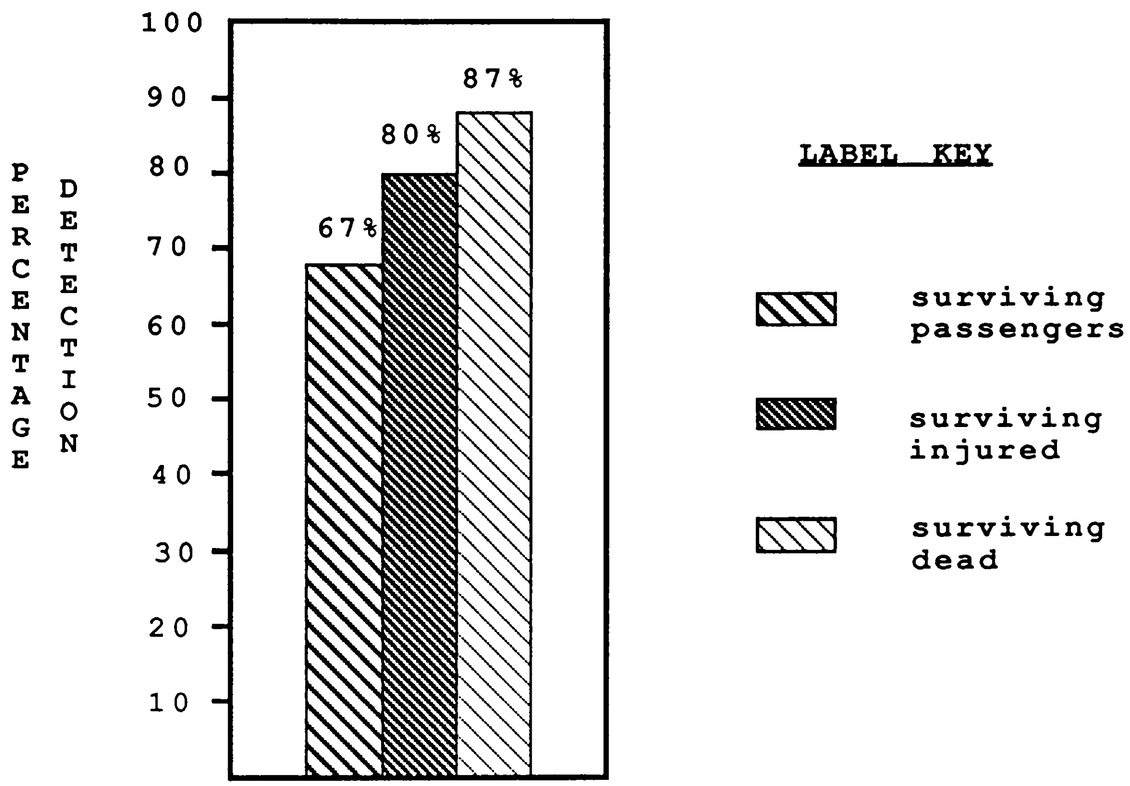
{1} Detection rates

Of the 59 subjects tested, 14 had previous knowledge of the problem and their detection status was excluded from the results. Despite having encountered the problem before, 6 of this total failed to detect the anomaly. The number of subjects detecting and non-detecting in each condition is presented, with the percentage detection rate, in Table 7.6. The percentage detection rates are also presented in diagrammatic form in Figure 7(c).

TABLE 7.7: Detection rates for Experiment 5.

Condition	**Detections**	**Non-detections**	**Percentage detection**
surv.pass	10	5	(67%)
surv.inj	12	3	(80%)
surv.dead	13	2	(87%)
All conditions	35	10	(78%)
Subjects with previous knowledge			
surv.pass	0	2	
surv.inj	4	3	
surv.dead	4	1	
All conditions	8	6	

FIGURE 7(c): Detection rates for Experiment 5.



At 78%, the detection rate is the highest obtained to this point in the case-study. There were no reliable detection rate differences across conditions (Chi-square=0.86, d.f.=2, p>0.25). The experimental hypothesis, that detection rates would significantly increase on the rates for the corresponding labels under the narrative format, was confirmed. Statistical analysis on the overall rates across text-forms showed a highly reliable effect (questionnaire=78%, narrative=43%: Chi-square=12.52, d.f=1, p<0.001). This effect will be evaluated in the discussion section.

{2} Content analysis for the written answers

The same evaluation procedure was adopted. The analysis for answer-types is presented in Table 7.7, and the corresponding analysis for reference-expressions in Table 7.8. A transcript of the solutions for those subjects contributing to the detection rate is included in Appendix F.

TABLE 7.8: Content analysis for answer-types in Experiment 5.

	Answer-type				
Condition	**Family**	**Home**	**Crash**	**No prob**	**Other**
surv.pass					
10 detectors	1	-	-	7	2
5 non-detect	4	1	-	-	1
surv.inj					
12 detector	1	-	-	8	2
3 non-detect	3	-	-	-	2
surv.dead					
13 detectors	3	-	-	5	1
2 non-detect	-	1	-	-	1
Total					
(35 detectors)	(5)	(-)	(-)	(20)	(5)
(10 non-detect)	(7)	(2)	(-)	(-)	(4)

The percentage of detectors making a pragmatic inference for subsequent death returned to approximately the same level as for the first three experiments. Sixty-seven percent of concordant evaluations found that 'no problem' existed. The small number of non-detectors makes cross-study comparison difficult. Nine of the 14 concordant evaluations involved the passenger's families and homes.

The number of references to the passengers again appears to be lower under the questionnaire format. A third of all concordant evaluations were in the 'other' category (again consistent with an increase in pronominal references for note-form answers). Detectors predominantly used the anomalous label for reference (14 out of 19 concordant evaluations). There were too few non-detector references to support any comment.

TABLE 7.9: Content analysis for reference-expressions in Experiment 5.

	Reference-expression				
Condition	**Dead**	**Bodies**	**Gener-al**	**Anoma-lous**	**Other**
surv.pass					
10 detectors	1	-	-	5	1
5 non-detect	-	-	-	-	1
surv.inj					
12 detector	-	-	-	5	3
3 non-detect	1	-	-	-	2
surv.dead					
13 detectors	-	-	-	4	-
2 non-detect	-	-	-	-	1
Total					
(35 detectors)	(1)	(-)	(-)	(14)	(4)
(10 non-detect)	(1)	(-)	(-)	(-)	(4)

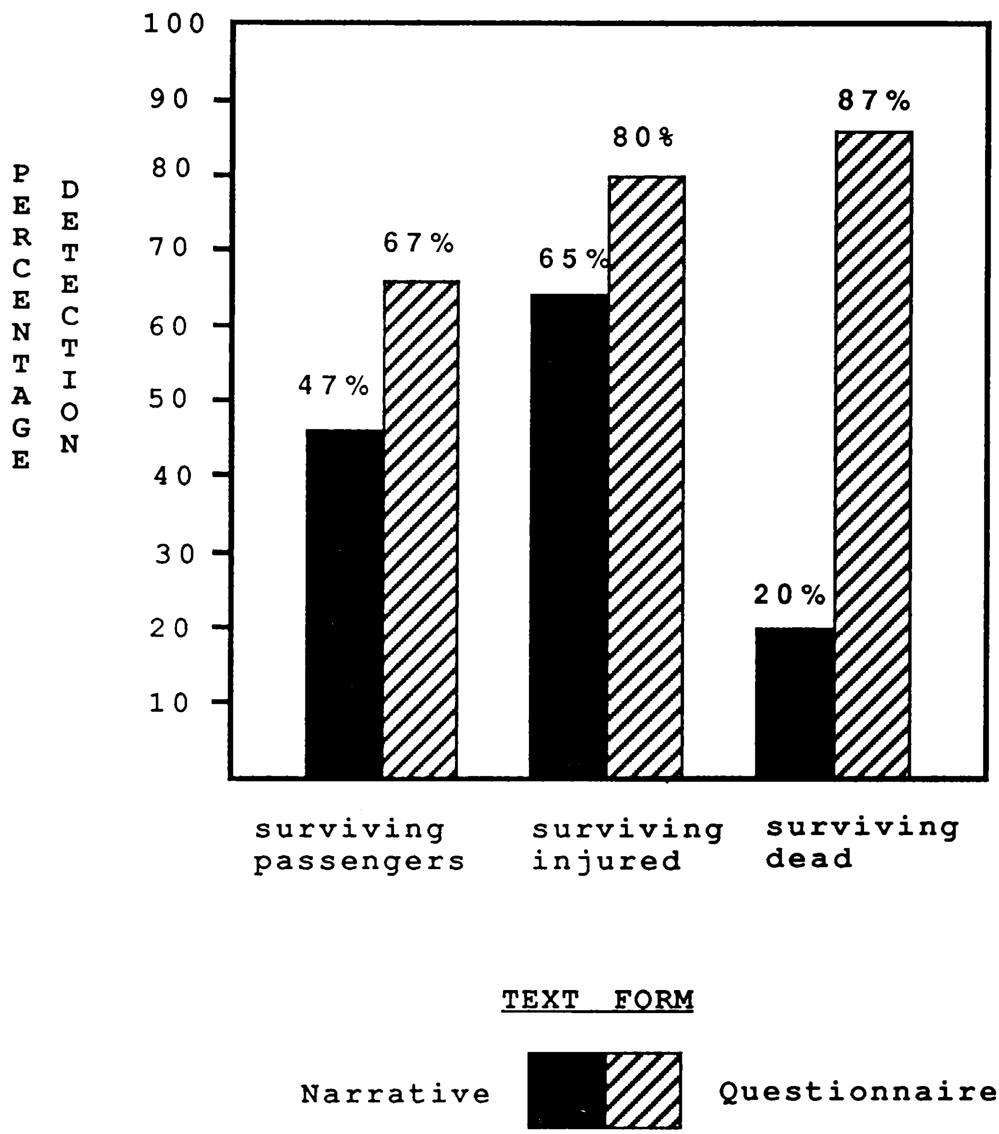
d: Discussion

The reliable overall **increase** across the text-forms was principally supported by the **surviving dead** condition. Separate analyses for each of the items were as follows:

surviving passengers; Chi-square=1.107, d.f.=1, NS:
surviving injured; Chi-square=0.924, d.f.=1, NS:
surviving dead; Chi-square=15.523, d.f.=1, p<0.001.

The global increase, and in particular the detection increase for the **surviving dead** condition, supports the hypothesis that some of the shallow processing in the narrative format had been due to top-down searches for expected inferential fields. Comparisons of the relevent detection rates are illustrated in Figure 7(d).

FIGURE 7(d): Contrasts for some emphasis-based manipulations over the narrative and questionnaire formats.



In question (2), where there is less build-up in thematic commitment effecting top-down searches, the composite expression **surviving dead** is more likely to manifest its true (and anomalous) meaning.

(2) When an aircraft crashes where should the surviving dead be buried?

Interpretation in the question format is more likely to permit growth of the relevant inferential fields for emphasis-based manipulations, and allow such manipulations to control focus in the way expected of them.

Of course, although the detection rate for the selected items has increased to 78% percent, this still represents over one in every five readers as a non-detector. For these manipulations surrogation is not as prevalent without the top-down effect, but there are clearly more passive types of expectation which license mappings on the basis of partial evidence.

The major issue of interest is in establishing the properties of question-construction which control patterns of incomplete processing. For example, could it be the case that the order of presentation for significant items in the message has a controlling effect on the processing for the anomalous item? Perhaps surrogation (and hence reduced processing completeness) depends upon an early mention of the item which evokes the background model, thus setting up a structure into which the item can be mapped? Or perhaps the time-course of item presentation is far enough removed from the time-course of interpretative commitment to make such rules-of-thumb of little value. Could detectibility (and hence depth of processing) be a function of various topicalisation factors in sentence structure, such as active and passive verb phrase constructions?

Also, we have noted that the aircrash scenario used for these materials sets up expectations for deaths and burials, therebye providing the background structure for the mappings that have been investigated. Indeed, it seems too obvious to mention that these outcomes are part of what it means to have knowledge about aircrashes. How prevalent would surrogation be were the anomaly presented in a message about a situation where death and burial was less predicted (for example, a bicycle accident)? This probes the extent to which depth of processing on a quasi-relation is effected by the plausibility of the state of affairs represented by that mapping. In interpretation, this information is signalled by the background scenario. These issues further probe the conditions under which processing completeness, and hence inference, is controlled by language. The following chapter will develop these issues further.

Chapter Eight

An Investigation into the Effects of Scenario and Question-Construction: Experiments 6 & 7

8.1 Summary

Various linguistic and contextual manipulations of the basic question-format are tested in Experiment 6. Detection rate contrasts suggest that presentation of anomalous labels prior to context-evoking items is not sufficient to increase focus on the anomaly. This result is consistent with the view that not all interpretative commitment proceeds in an incremental fashion. A powerful effect of scenario-type on detection rate is observed. The goodness-of-fit for references into a background model is found to be a function of the degree to which those items are predicted by that model. These results give further support to scenario-mapping theories of interpretation. A narrative-based scenario manipulation in Experiment 7 provides results consistent with these claims.

8.2 Experiment 6: Scenario, question-construction and focus

a: Introduction

The language/focus/inference issue has been probed in important ways over the first five experiments. The incomplete or partial nature of inferential activity has been found to vary as a function of various aspects of the language used in texts. We have seen that a developed narrative structure may effect thematic commitment in the processor to the point where it becomes harder than expected for some items to exert focus on objects that violate expectations. The interpretation of such items can occur through mapping activity that appears to operate on only part of the inferential fields owned by the items. This partial activity seems, on occasions, to display a bias towards fields of information that have been expected.

Clearly, then, depth of inferential processing on information signalled by a text is subject to control constraints. In the present paradigm, the constraints of interest concern properties of question construction **that** may increase focus on the real meaning of **survivors**, for example, by bringing that part of the message into the foreground of the processor's attention. As before, the measure of the extent of this foregrounding is taken to be the detectibility of the anomalous label in the position of interest.

Various possibilities emerge as controllers of focus in question-construction. For example, could it be the case that a late mention of a contextually-signalling item will give the inferential fields of previous items the oppurtunity to grow and establish themselves without first being mapped into a richer background model? This view assumes incremental interpretation. In all conditions thus far, the mention of the aircrash has occured <u>prior</u> to the mention of the burial objects. This is true both for the narrative and the question.

(1) When an aircraft crashes where should the survivors be buried?

It could be that the background thematic structure established prior to the occurrence of the anomaly provides the expectations for the preferred mapping. An earlier mention of the anomalous label, and a later mention of the aircrash context, may be sufficient for the full meaning of **survivors** to become established. One possible construction is given in sentence (2).

(2) Where should the survivors be buried after an aircrash?

We have already observed from the levels of semantic surrogation in Experiments 4 and 5 that it is not pre-requisite for these effects that the verb occurs before

its object in the sentence. So the time-course of interpretative activity must be such that a later mention of the verb is not, on its own, sufficient to let the full meaning of the **survivors** to make an impact. In this domain, a possible linguistic manipulation involves varying the order-presentation of the verb and its object. This can be achieved by using actice or passive verb-phrases in the question. Hence, the passive phrase 'survivors be buried' in both (1) and (2) can be 'activised' giving sentences (3) and (4).

(3) When an aircraft crashes where should you bury the survivors?

(4) Where should you bury the survivors of an aircrash?

Using these four question-constructions, the late and early mention of the scenario-evoking items occurs with both active and passive verb-phrases. The key interest is to find out whether detectibility varies as a function of these question-structure manipulations. Of course, any such variability would not just be of relevance to anomaly detection. Such parameters could reflect stable relationships between linguistic forms and consequent focusing patterns.

There is another dimension of question-structure that may be important. In all conditions thus far, the texts have been about aircrashes, and aircrashes predict death as a probable outcome. Hence, the mention of 'burial' in the context of an aircrash may be sufficient to trigger expectations about burying dead bodies: the object of the burial is highly predicted. The evoked scenario may be having a controlling effect on the depth of processing on items in the text, depending on their relation to it. A test for this factor is to change the topic of the text to one where death is a possible but not predicted outcome. In the present case, the above four question-types were also tested with the topic changed from an aircrash to a bicycle accident. The corresponding questions are presented in (5), (6), (7) and (8).

(5) When a bicycle accident occurs where should the survivors be buried?

(6) Where should the survivors be buried after a bicycle accident?

(7) When a bicycle accident occurs where should you bury the survivors?

(8) Where should you bury the survivors of a bicycle accident?

Again, both (5) and (6) have passive verb-phrases, and (7) and (8) active ones. Questions (5) and (7) introduce the topic early in the message, while (6) and (8) have a late mention. The sentences (1) to (8) represent the conditions of interest in Experiment 6. Question (1) is the standard construction that has already been tested in Experiment 4, so this experiment proceeded with seven conditions. In order to allow easy comparisions, this condition will be re-reported in the relevant results sections.

As a final note, the **survivors** label was selected for the tests because its detection level of 20% in Experiment 4 suggested that considerable scope existed for detection variation within the parameters of the present design. Variation in other labels seemed more likely to reach a ceiling level.

b: Method

The materials, design, subjects and procedure were the same as in Experiments 4 and 5. In this experiment, seven conditions were tested, composed of four sentence-construction types each of which was tested under a bicycle accident scenario-type, and (the latter) three of which were tested under an aircrash scenario-type.

Question-types A and B displayed items which introduced the scenario early in the sentence, while types C and D had late scenario-descriptions. Both A and C had passive verb-phrases, and B and D had active verb-phrases. These types are listed below, and given in full in sentences (2) to (8).

A. Early scenario-description, passive verb-phrase.
B. Early scenario-description, active verb-phrase.
C. Late scenario-description, passive verb-phrase.
D. Late scenario-description, active verb-phrase.

A. When {SCENARIO DESCRIPTION} where should the survivors be buried?

B. When {SCENARIO DESCRIPTION} where should you bury the survivors?

C. Where should the survivors be buried {SCENARIO DESCRIPTION}?

D. Where should you bury the survivors {SCENARIO DESCRIPTION}?

Seven independent groups of subjects saw only one version of the questionnaire. A total of 160 subjects was tested over the seven conditions.

c: Results

<u>{1} Detection rates</u>

Of the 160 subjects tested, 55 had previous knowledge of the problem and their detection status was excluded from the results. Despite having encountered the problem before, 15 of this total failed to detect the anomaly. The number of subjects detecting and non-detecting in each condition is presented, with the percentage detection rate, in Table 8.1 (Table 8.2 contains the results for the subjects with previous knowledge). The percentage detection rates are also presented in diagrammatic form in Figure 8(a). For comparative purposes, the results for the **survivors** condition of Experiment 4 are included in these summaries (represented as question-type A / aircrash scenario). The analysis of the detection rates will include this condition, as will all summary information from this point.

Across the 8 combined conditions of interest, then, 195 subjects were tested, 75 of these having previous knowledge of the problem. Despite prior knowledge, 22 of this total failed to detect the anomaly.

Because the number of subjects contributing to the detection rate was constant across conditions (i.e. fifteen), a chi-square factorial analysis on the detection frequencies was undertaken (c.f. Sutcliffe, 1957; Winer, 1971). This analysis tested the variance due to three factors and their interactions: scenario-type, onset of scenario-description and verb-phrase structure (the latter two factors were nested). The results of this analysis are presented in Table 8.3.

TABLE 8.1: Detection rates for Experiment 6.

Condition		Detections	Non-detections	Percentage detection
Aircrash scenario				
Question-type	A.	3	12	(20%)
"	B.	4	11	(27%)
"	C.	5	10	(33%)
"	D.	8	7	(53%)
Total		20	40	(33%)
Bicycle accident scenario				
Question-type	A.	12	3	(80%)
"	B.	11	4	(73%)
"	C.	11	4	(73%)
"	D.	14	1	(93%)
Total		48	12	(80%)
Both scenario-types		68	52	(57%)

TABLE 8.2: Detection results for subjects with previous knowledge in Experiment 6.

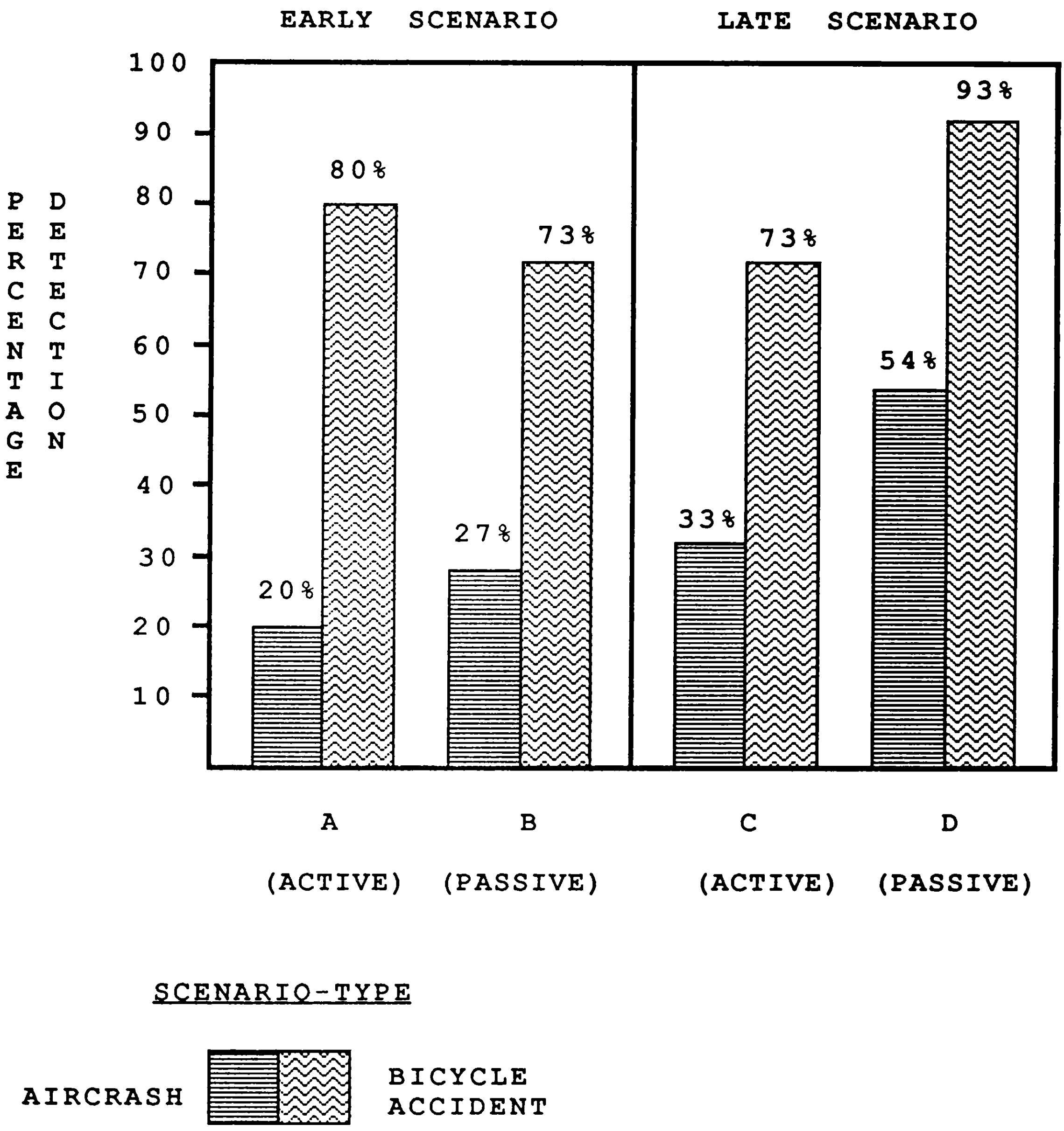
Question Type	Det	Non-det	Det	Non-det
	(Aircrash)		(Bicycle accident)	
A.	13	7	11	2
B.	7	3	6	1
C.	5	4	3	2
D.	4	2	4	1
Total	29	16	24	6

TABLE 8.3: Analysis of detection variance in Experiment 6.

Source of Variation	**Degrees of Freedom**	**Chi-square ratios**	**P values**
Total	7	14.15	$p<0.05$
Scenario-type	1	11.53	$p<0.001$
Scenario-onset	1	0.94	NS
Verb structure	1	0.53	NS
Scen-type x Onset	1	0.24	NS
Scen-type x Verb	1	0.59	NS
Onset x Verb	1	0.53	NS
Type x Onset x Verb	1	0.32	NS

The overwhelming majority of the detection variance is accounted for by the scenario-type main effect. No other main effect or interaction approached significance. The results confirm that an overall detection rate of 80% for the bicycle accident scenario is reliably larger than the 33% for the aircrash conditions. Manipulating the contextual information signalled by the question has had a powerful effect on the detectibility of the anomaly. There appears to be a weak numerical trend favouring detection in the late scenario-descriptions in the aircrash conditions but this interaction is not reliable. Similarily, a trend appears to favour detection for the active verb-phrases within these sets, but this is not reliable either. Although focus effects due to these linguistic factors cannot be ruled out, there is no evidence for them in the present data.

FIGURE 8(a): Detection rates for Experiment 6.



{2} Content analysis for the written answers

The same two judges carried out the established evaluation procedure. In the present case, references of interest could also be to cyclists as well as to passengers. The analysis for the answer-types is presented, for each condition, in Tables 8.4, 8.5 and 8.6 and the corresponding analysis for reference-expressions in Table 8.7, 8.8 and 8.9. In each case the first table provides the results for the aircrash conditions, the second the bicycle accident conditions, and the third the totals for both sets. A transcript of the solutions for those subjects contributing to the detection rate (in the seven new conditions) is included in Appendix G.

TABLE 8.4: Content analysis for answer-types in Experiment 6: aircrash scenario.

	Answer-type				
Condition	**Family**	**Home**	**Crash**	**No prob**	**Other**
AIRCRASH SCENARIO					
Type A.					
3 detectors	-	-	-	3	-
12 non-detect	6	4	1	-	2
Type B.					
4 detectors	1	-	-	3	-
11 non-detect	8	2	2	-	2
Type C.					
5 detectors	1	-	-	3	-
10 non-detect	6	3	-	-	2
Type D.					
8 detectors	1	1	-	5	2
7 non-detect	5	2	-	-	-
TOTAL					
20 detectors	(3)	(1)	(-)	(14)	(2)
40 non-detect	(25)	(11)	(3)	(-)	(6)

TABLE 8.5: Content analysis for answer-types in Experiment 6: bicycle accident scenario.

	Answer-type				
Condition	**Family**	**Home**	**Crash**	**No prob**	**Other**
BICYCLE ACCIDENT SCENARIO					
Type A.					
12 detectors	-	-	-	5	2
3 non-detect	1	1	-	-	-
Type B.					
11 detectors	-	1	-	8	3
4 non-detect	2	-	-	-	2
Type C.					
11 detectors	-	-	-	8	2
4 non-detect	3	-	-	-	2
Type D.					
14 detectors	1	-	-	6	3
1 non-detect	-	-	-	-	-
TOTAL					
48 detectors	(1)	(1)	(-)	(27)	(10)
12 non-detect	(6)	(1)	(-)	(-)	(4)

TABLE 8.6: Content analysis totals for answer-types in Experiment 6.

	Answer-type				
Condition	**Family**	**Home**	**Crash**	**No prob**	**Other**
TOTALS FOR BOTH SCENARIOS					
68 detectors	(4)	(2)	(-)	(41)	(12)
52 non-detect	(31)	(12)	(3)	(-)	(10)

The answer-types corresponded closely to the general pattern that has emerged over most of the studies. There were no obvious differences across the scenario-types, with 69% of detectors asserting that no problem existed. Seventy-seven percent of non-detectors involved the passenger's/cyclist's homes or families.

TABLE 8.7: Content analysis for reference-expressions in Experiment 6: aircrash scenario.

	Reference-expression				
Condition	**Dead**	**Bodies**	**Gener-al**	**Anoma-lous**	**Other**
AIRCRASH SCENARIO					
Type A.					
3 detectors	-	-	-	3	-
12 non-detect	-	-	1	-	3
Type B.					
4 detectors	1	-	-	1	-
11 non-detect	1	1	2	-	4
Type C.					
5 detectors	-	-	-	2	-
10 non-detect	-	-	-	-	3
Type D.					
8 detectors	-	-	-	1	2
7 non-detect	-	-	-	-	2
TOTAL					
20 detectors	(1)	(-)	(-)	(7)	(2)
40 non-detect	(1)	(1)	(3)	(-)	(12)

As in Experiments 4 and 5, the overall number of references to the passengers (or cyclists) decreased on the narrative format. This is again consistent with the view that more note-form answers in the questionnaire are less likely to directly refer to the passengers/cyclists. The number of 'other' references (including pronouns) was high. There were no apparent differences across the scenario-types.

TABLE 8.8: Content analysis for reference-expressions in Experiment 6: bicycle accident scenario.

	Reference-expression				
Condition	**Dead**	**Bodies**	**Gener-al**	**Anoma-lous**	**Other**
BICYCLE ACCIDENT SCENARIO					
Type A.					
12 detectors	-	-	-	2	3
3 non-detect	-	2	-	-	2
Type B.					
11 detectors	-	-	-	4	1
4 non-detect	1	-	-	-	3
Type C.					
11 detectors	-	-	-	6	2
4 non-detect	-	-	-	-	3
Type D.					
14 detectors	-	-	-	1	3
1 non-detect	-	-	-	-	1
TOTAL					
48 detectors	(-)	(-)	(-)	(13)	(9)
12 non-detect	(1)	(2)	(-)	(-)	(9)

TABLE 8.9: Content analysis totals for reference-expressions in Experiment 6.

	Reference-expression				
Condition	**Dead**	**Bodies**	**Gener-al**	**Anoma-lous**	**Other**
TOTALS FOR BOTH SCENARIOS					
68 detectors	(1)	(-)	(-)	(20)	(11)
52 non-detect	(2)	(3)	(3)	(-)	(21)

d: Discussion

That late mention of the scenario-evoking items has not resulted in increased detection suggests that interpretation is not made on a strictly incremental basis. Mention of the anomalous item prior to the scenario description has not given it the 'space' to develop its own inferential fields. Rather, the results are consistent with the view that resolution occurs for the message as a whole, and that this continues to allow the late-signalled contextual information to exert comparable power of expectation. Similarily, with the verb-phrase manipulations, passivising the phrase may not be enough to put focus on the object of the verb if interpretative commitment is not made in an incremental manner.

The issues raised concern the time-course of interpretative commitment and how this interacts with the type of message being resolved. **A text-driven incremental strategy would attempt interpretation of every item as it appeared, mapping it into whatever larger structure was available. A theme-driven strategy might delay full interpretation until a shallower process on all items had indicated which were most likely to evoke the relevant background information**. When the appropriate scenario had been evoked, cohesion for the message could be established <u>relative to the selected background structure</u>. This strategy has the advantage of minimising unnecessary depths of processing on non-central items. Appropriate depths of processing on all items may be easier to orchestrate when the theme or topic had been identified.

The present results support the theme-driven cohesion strategy. The evidence suggests that strongly predicted outcomes (derived from the scenario) have a direct bearing on the depth of processing for relations consistent with predictions. In the aircrash cases, partial evidence seems sufficient to support the mapping. For the bicycle accident conditions, removal of the predictive aspect, relying only on a consistency relation, increases the detection rate for the anomaly. Death from a bicycle accident is more surprising than from an aircrash, the present results suggesting that these levels of goodness-of-fit are reflected in on-line cohesion establishment.

Importantly, any view of text cohesion that proposed the computation of structural relations (e.g. propositions) prior to filtering them through background knolwedge, would not predict any difference in the detection rates across the scenario manipulations. Under such views, the (pragmatically) anomalous relation between burial and survivors should be established irrespective of the context in which it is described. The present results further discredit such views. They suggest that initial processing may be shallow, but of sufficient depth to signal appropriate background information. Subsequent interpretation may be scenario-driven, where the background information controls the depth of processing on proposed mappings. Hence the detection difference across scenario-types. Accomodation of expected references occurs in the aircrash context (where both burial and survival are predicted). In the context of a bicycle accident, neither are expected or naturally relevant: the effort to make an interpretation increases detection of the anomalous relation. This contrasts with proposition-based views where cohesion at the level of the present anomaly is purported to be established prior to fuller interpretation.

Shallow initial processing and delayed thematic commitment will be investigated further in the following chapter. Presently, it is not clear whether the scenario effects are specific to messages where initial cohesion can be shallow, perhaps because message resolution only has to be established over a single sentence. The following experiment returns to the narrative paradigm to test for effects of scenario manipulation under those conditions.

8.3 Experiment 7: A narrative-based scenario manipulation

a: Introduction

The results of Experiment 6 suggest that cohesion patterns are strongly effected by the expectations signalled by a background interpretative scenario. Experiment 7 was carried out using an extended narrative text to test whether the same scenario manipulation would effect anomaly detections under this format. Some of the conditions from Experiments 1, 2 and 3 were retested in a narrative which described a similar problem to the aircrash, but this time situated within the events of a bicycle accident. The original aircrash narrative is presented in vignette (11), and the new bicycle accident narrative is presented in vignette (12).

(11) There was a tourist flight travelling from Vienna to Barcelona. On the last leg of its journey, it developed engine trouble. Over the Pyrenees, the pilot started to lose control. The plane eventually crashed right on the border. The wreckage was equally strewn in France and Spain. The authorities were deciding where to bury the survivors. They couldn't make up their minds.

Question: What would your solution to the problem be?

(12) There was a cycle race being staged from Milan to Geneva. The leaders were approaching the border of Italy and Switzerland. Suddenly, the leading cyclist got a puncture and lost control of his bike. He fell awkwardly in front of twenty other competitors. The whole pack ended up in a heap right on the border. The wreckage was equally strewn in Italy and Switzerland. The authorities were deciding where to bury the survivors. The couldn't make up their minds.

Question: What would your solution to the problem be?

It was expected that the thematic change would make death a less predicted outcome, and neither survivors nor burial would be accommodated as easily. As a result, detection of the anomalous items would increase. There were three conditions tested manipulating the anomalous label. These were **survivors**, **wounded** and the composite expression **surviving dead**.

b: Method

The method was the same as for Experiments 1, 2 and 3, with some changes allowing for the scenario manipulation.

{1} Materials and design

The materials used were three texts each of eight sentences in length. Each of these texts described a problem that an individual or group of individuals was experiencing in a particular situation. Two of the texts were used to familiarise the subjects with the experimental procedure, and the other was the experimental bicycle accident narrative as presented in vignette (11). Neither of the other texts contained anomalies.

There were three independent conditions in this experiment, each condition reflecting a text-based manipulation. The first condition was as printed in (12), and the other two conditions were identical except for the substitution of **wounded** or **surviving dead** in place of **survivors**. Three independent groups of subjects saw only one of these versions. All subjects saw the same first two texts, which are presented in Appendix A (These are the original filler texts with an additional sentence preserving text-length in the present design).

{2} Subjects

These were mostly undergraduates at the University of Glasgow. A total of 75 was tested. After reading, some subjects claimed to have familiarity with the aircrash 'trick', these subjects being removed from the initial analysis.

{3} Apparatus

The materials were presented on an Apple MacIntosh computer screen (MacPlus version) calibrated for self-paced reading with millisecond accuracy. By pressing the space-bar subjects could pace the rate at which successive sentences of the text were presented to them, and the computer measured the dwell-times on each sentence. A pen and paper were provided for the subject to write down their solution after reading the problem.

{4} Procedure

Each subject was informed that the experiment involved answering some problems set on a computer screen. The experimenter displayed the use of the computer equipment by using the space-bar press to move through a test trial (using the sentences of the first text [see Appendix A]). A space-bar press moved the screen-display from "READY" into the first sentence of the text. Subsequent presses moved the reader from sentence to sentence. The current sentence was removed from the screen as the next appeared.

Subjects were asked to read at their normal pace. They were encouraged to imagine that they were reading a newspaper report. They were told that each of the texts would be eight sentences in length, and that following the eighth sentence the computer would set them a question about the text. In each case the computer asked that the subject write down their own solution to the problem that has been described. After writing down a solution, the subject made a space-bar press which moved the program round from the "QUESTION" display to the "READY" display for the next trial. Each subject was asked to inform the experimenter when the computer displayed "FINISH" after the last trial. Each subject provided a solution to two problems, one on a pre-test filler problem, and the other on the bicycle accident narrative.

When each subject had finished the experiment, they were de-briefed in the standard way. A record was kept of each subject's detection status, and whether or not they had previous knowledge of the problem.

c: Results

{1} Detection rates

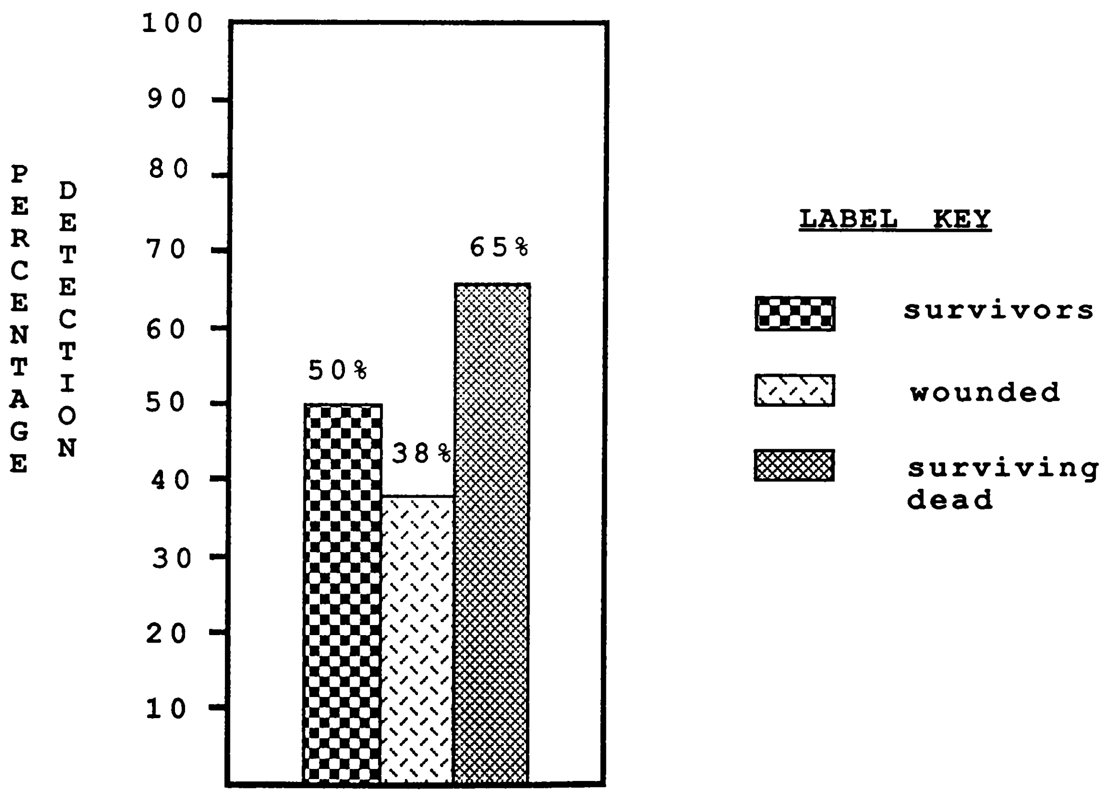
Of the 75 subjects tested, 28 had previous knowledge of the problem and their detection status was excluded from the main results. Despite having encountered the problem before, 8 of this total failed to detect the anomaly. The number of subjects detecting and non-detecting in each condition is presented, with the percentage detection rate, in Table 8.10. The percentage detection rates are also presented in diagrammatic form in Figure 8(b).

At 51%, the detection rate was not as high as had been expected. There was no significant difference in detection rate across conditions (Chi-square=2.45, d.f.=2, p>0.25). The corresponding rate for the same conditions under the aircrash scenario had been 35%, but the present manipulation did not result in a reliable overall increase (Chi-square=2.63, d.f.=1, p>0.1). Separate analyses for each condition are described in the discussion section.

TABLE 8.10: Detection rates for Experiment 7.

Condition	**Detections**	**Non-detections**	**Percentage detection**
survivors	7	7	(50%)
wounded	6	10	(38%)
surv.dead	11	6	(65%)
All conditions	24	23	(51%)
Subjects with previous knowledge			
survivors	7	4	
wounded	6	3	
surv.dead	7	1	
All conditions	20	8	

FIGURE 8(b): Detection rates for Experiment 7.



{2} Anomalous sentence dwell-times and reading speeds

Sentence timing problems occured with the computer equipment for 14 of the 47 subjects contributing to the detection rates. This did not effect the procedure of the experiment in any way, but the dwell-times and reading speeds for these subjects were lost. For the remaining 33 subjects, the mean dwell-times for the anomalous seventh sentence are presented in Table 8.11, and the mean reading speeds (sum of the dwell-times on sentences three, four, five and six) are presented in Table 8.12. For all means, the number of contributing data-points are bracketed.

TABLE 8.11: Anomalous sentence dwell-times for Experiment 7.

Dwell-Times (milliseconds)			
Condition	**Means for Detectors**	**Means for Non-detect**	**Total Means**
survivors	4333 {2}	4715 {7}	4524 {9}
wounded	8681 {4}	4422 {6}	6552 {10}
surv.dead	14810 {9}	5283 {5}	10047 {14}
Total	9275 {15}	4807 {18}	6838 {33}

TABLE 8.12: Reading speeds for Experiment 7.

Reading Speeds (seconds)			
Condition	**Means for Detectors**	**Means for Non-detect**	**Total Means**
survivors	15.5 {2}	13.6 {7}	14.5 {9}
wounded	14.0 {4}	13.4 {6}	13.7 {10}
surv.dead	15.6 {9}	15.5 {5}	15.5 {14}
Total	15.0 {15}	14.2 {18}	14.6 {33}

Separate two-by-three analyses of variance were carried out for the dwell-times and the reading speeds. The results of these tests are shown in Tables 8.13 and 8.14.

TABLE 8.13: Analysis of variance for dwell-times in Experiment 7.

Source of Variation	**Degrees of Freedom**	**F ratios**	**Probability values**
Detection	1	1.661	0.2084
Conditions	2	0.866	0.4321
Interaction	2	0.682	0.5143

No dwell-time or reading speed effects approached significance. The reduced power of these analyses due to missing data makes contrasts with existing effects from the previous narrative-based studies impossible to make.

TABLE 8.14: Analysis of variance for reading speeds in Experiment 7.

Source of Variation	**Degrees of Freedom**	**F ratios**	**Probability values**
Detection	1	0.314	0.5798
Conditions	2	0.480	0.6238
Interaction	2	0.137	0.8725

{3} Content analysis for the written solutions

The same two judges carried out the established evaluation procedure as in the previous experiments. In the present case, references of interest were to cyclists and not to passengers. The content analysis for the solution-types is presented, for each condition, in Table 8.15, and the corresponding analysis for reference-expressions in Table 8.16. A transcript of the solutions for those subjects contributing to the detection rate is included in Appendix H.

The problem solutions displayed the usual pattern of non-detectors concentrating on the cyclist's homes and families. More detectors than usual (just under a half of concordant evaluations) made these attributions also, with the remainder asserting that no problem existed. Perhaps the present text content is more consistent with drawing the inference for subsequent death, though it is not clear why this might be.

TABLE 8.15: Content analysis for solution-types in Experiment 7.

	Solution-type				
Condition	**Family**	**Home**	**Crash**	**No prob**	**Other**
survivors					
7 detectors	2	1	-	4	-
7 non-detect	3	4	-	-	1
wounded					
6 detectors	1	1	-	2	-
10 non-detect	4	2	-	-	3
surv.dead					
11 detectors	3	2	-	5	-
6 non-detect	3	4	-	-	-
Total					
(24 detectors)	(6)	(4)	(-)	(11)	(-)
(23 non-detect)	(10)	(10)	(-)	(-)	(4)

The number of references to the cyclists returned to approximately the rate found for the passengers in the previous narrative studies. There appears to be a greater spread of reference descriptions, particularly for the non-detectors where 10 of the 21 concordant evaluations were in the 'other' category. Eleven of the 23 concordant evaluations for the detectors used the anomalous label.

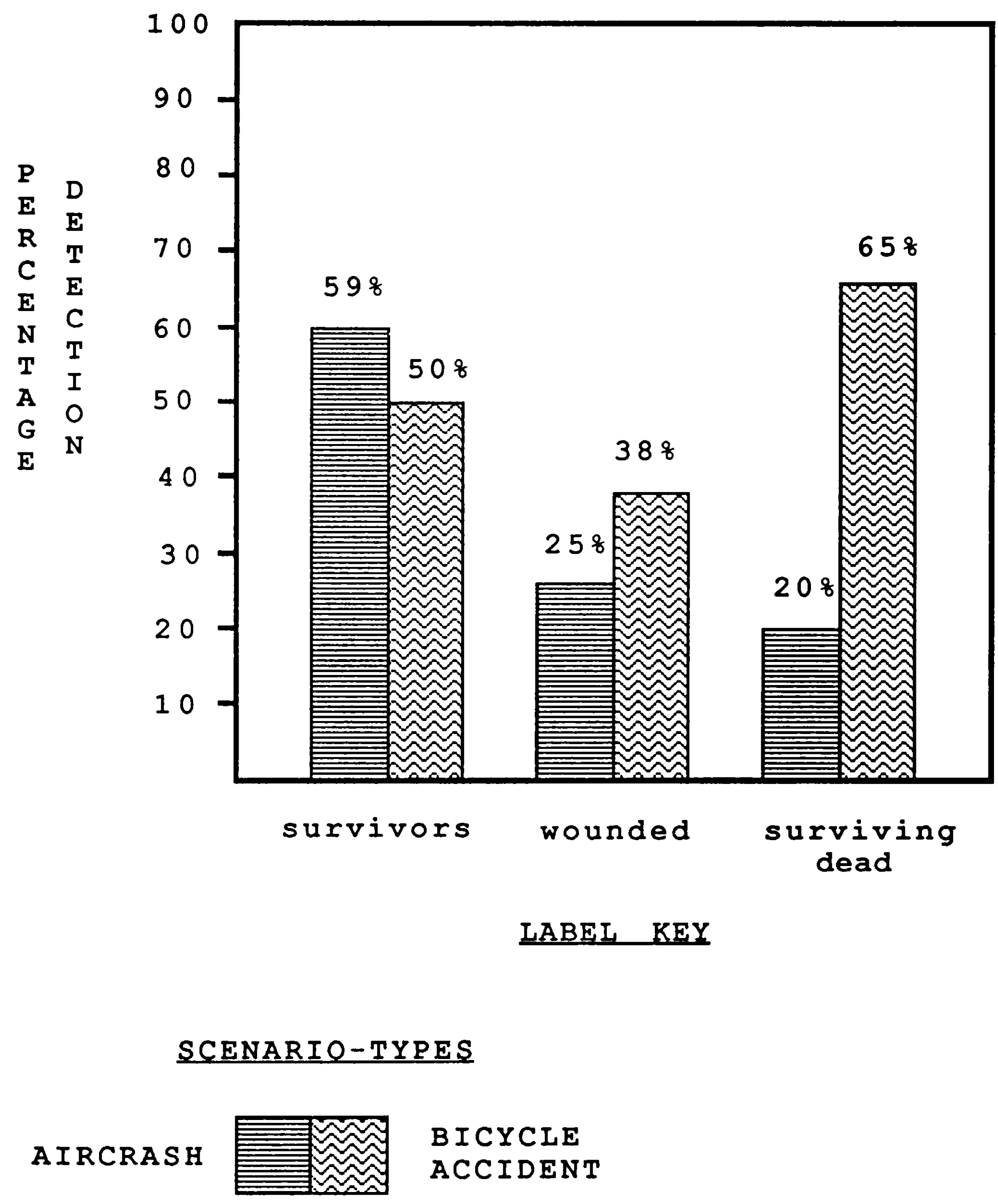
TABLE 8.16: Content analysis for reference-expressions in Experiment 7.

	Reference-expression				
Condition	**Dead**	**Bodies**	**Gener-al**	**Anoma-lous**	**Other**
survivors					
7 detectors	2	-	-	4	1
7 non-detect	1	-	3	-	3
wounded					
6 detectors	2	1	1	2	1
10 non-detect	1	1	1	-	5
surv.dead					
11 detectors	1	-	-	5	3
6 non-detect	-	3	1	-	2
Total					
(24 detectors)	(5)	(1)	(1)	(11)	(5)
(23 non-detect)	(2)	(4)	(5)	(-)	(10)

d: Discussion

Separate analyses on the contrasts between the particular labels showed that only the **surviving dead** label increased in detection under the bicycle accident narrative. This effect was highly reliable (Chi-square=7.62, d.f.=1, $p<0.01$). Neither of the other labels showed any reliable difference across scenario-type (**survivors**, Chi-square=0.239, d.f.=1, $p>0.25$: **wounded**, Chi-square=1.1, d.f.=1, $p>0.25$). These contrasts are illustrated in Figure 8(c).

FIGURE 8(c): Contrasts for narrative manipulations across scenario-types.



The results further support the view that semantic surrogation, and hence shallow processing, is not a simple outcome of top-down testing activity. The bicycle accident topic would seem to be more consistent with damage and injury as an outcome, but the implication of death appears to be consistent enough with the whole situation for the anomalous items to be accommodated. The view presented thus far has suggested that the focusing effects of composite expressions are susceptible to attenuation under top-down processing conditions. The present results support this view. The local anomaly, **surviving dead**, is detected significantly more often under the present scenario, suggesting that expectations of death are more actively constructed from the aircrash theme. The 'expectations' from the present scenario must be more passive. Composite inferential fields are more able to make an impact on the representation, presumably because there is less search for fields particular to one of the items.

With hindsight, it is possible that death was more consistent as an outcome in the present narrative than had been anticipated in the design of the study. Many subjects reported that the text had described an unusual set of circumstances, but, overall, this did not seem to increase the processor's attention on the 'surprising' descriptions. If anything, surprise at the narrative content may have contributed an opposite effect. It is possible that the appearance of implausibility may have had a distracting effect at a level removed from the present level of interest. This suggestion is highly speculative, but some unpublished work by Bell and Grant (1991) using the same paradigm but with an even less death-predicting narrative (a sponsored walk), also failed to find any overall detection increase on the aircrash conditions. Other more global processing factors may be swamping the expected monotonic relationship between degree of prediction and detection rate.

The experiments in the present chapter have shown that on-line cohesion activity is highly influenced by mappings from input descriptions to background information. The degree of surrogation found to occur across scenario-types has supported the view that depth of processing is a partial function of the relation between the scenario-based information and the quasi-mapping. Hence, the relevance of text-based items to the established theme emerges as a strong candidate for control of inferences. In the present case, 'expected' items that fit well into the developing model may be accommodated through a shallower process that might otherwise be the case.

Clearly, further investigation is needed of the conditions under which items are brought into focus and for what reasons. The present chapter has introduced the role of the background scenario in controlling such activity. The following, and final, experimental chapter investigates the role of relevance as a general controlling 'mechanism' for depth of processing.

Chapter Nine

Partial Cohesion and the Role of Relevance: Experiment 8

9.1 Summary

Experiment 8 tested the detection rate for question-based anomalies presented over two sentences. An aircrash with survivors is introduced in sentence 1, and the burial location issue is delayed until sentence 2. Even under these conditions surrogation was found to occur, suggesting that cohesion on the first sentence may have been temporary and contingent on the direction of subsequent information for further processing. Surrogation was most prevalent where a qualification was made to the survivors which was highly relevant to a burial-location problem. This finding supports the view that tight mappings to an existing background (problem) structure can boost global cohesion and relax the tests on other relations.

9.2 Experiment 8: Two-sentence question-constructions investigating partial cohesion and relevance

a: Introduction

This experiment grew out of attempts to find conditions under which all readers would detect the anomaly. No condition tested in the first seven experiments of the case-study had met with 100% detection, so to probe the boundary conditions of depth of processing and semantic surrogation some work was carried out on two-sentence question-constructions.

The idea behind all of the surrogation reported in this thesis has been that an existing representation can have mappings made into it on the basis of partial evidence or subject to global constraints. Detection of mis-mappings would seem to depend upon the anomalous label being allowed sufficient 'space' from the background model in order to establish the growth of its own inferential fields before they are tested for possible mappings. This 'space' could result from greater focusing on some of the properties of the object, or from weaker expectations about the plausibility of the mapping between the two. Where this does not happen, shallow processing can result in un-detected inconsistencies. Vignette (1) is an example of a two-sentence construction where detection should be guaranteed, for two main reasons.

(1) Suppose there was an aircrash with survivors who were mostly unhurt. Where should they be buried?

First, the relative clause qualification 'who were mostly unhurt' places emphasis on the survivor's 'aliveness'. Binding the 'unhurt' property to these individuals seems necessarily to foreground the fact that they are un-damaged. Second, any sentence-completion cohesion-establishment should ensure that the full meaning of **survivors** became manifest without interference from thematic expectations. Although no effect was found for verb position in Experiment 6, one might expect that if it is placed in the <u>following sentence</u>, sufficient 'space' would be created for the **survivors** to be established in their own right. Further, the object for the verb is signalled by the plural pronoun 'they' which refers back to explicitly mentioned entities (i.e. the **survivors**). So, in this format, the cohesion for the whole message demands the match of an anaphoric relation with the anomalous item as antecedent. Notably, the second sentence in the question could just as plausibly be about who should question the survivors before they go home, or such like.

(2) Suppose there is an aircrash with survivors who were mostly unhurt. Who should be responsible for interviewing them before they go home?

In fact, this format raises a whole range of interesting issues. It is the second sentence which signals the question, the first merely setting background information. So it will be the second sentence that will access the problem-domain from which an answer will be selected. The first sentence will provide the information relevant to the problem for creating a solution. It is possible, then, that two related factors could come into play which would perhaps bias against detection of the anomaly.

The first concerns partial cohesion, the second, mapping to problem-structures. If the cohesion that is established for the first sentence is in some way on-hold or partial, waiting to be driven by the subsequent question, then the full inferential fields of the first sentence's items may not, in fact, be given the oppurtunity to grow and establish themselves. It is not meant to be implied that no interpretation is carried out in the first sentence. Rather, the significance of the whole message (eventually signalled by the second sentence) may determine which inferential fields opened by the first sentence are explored in depth. So a full-stop may may not signal the discontinuity in interpretation that it does in sentence structure. As well as this, the type of qualification made to the **survivors** may, independent of its relation to those individuals, be highly relevant to answering the question given in the second sentence (in a manner consistent with the assumption that it is dead people who are to be buried). Consider, then, vignette (3).

(3) Suppose there was an aircrash with survivors who were mostly of no fixed abode. Where should they be buried?

The combined effect of any 'on-hold' cohesion for the first sentence, and the relevance of a homeless person to a burial issue, may distract attention from the anomaly. Global cohesion may be satisfied by the tight mapping of the qualification into the problem structure, and tests on the **survivors** may be relaxed. It is not known whether such constraints would have sufficient force to supress the salience of reference resolution tests. Were surrogation to occur under these conditions, this hypothesis would be supported. Further, we would also expect the content of reader's answers to be dominated with issues pertinent to the qualification in question.

It was with this rationale that Experiment 8 was conceived. A total of eight conditions was tested, each reflecting a manipulation of the qualification in the first sentence of the standard two-sentence format (1). The two qualifications introduced thus far, **mostly unhurt** and **mostly of no fixed abode**, were thought to lie at opposite ends of the 'relevance spectrum'. **Mostly unhurt** was highly irrelevant in the sense that it strongly signalled the inappropriateness of the question (and any burial-type solution). **Mostly of no fixed abode** was thought to be highly relevant, since this information could be integrated into a problem-definition for dead persons and be used to provide a creative burial-place solution. The qualification-types were thought to have an intermediate class, which neither strongly signalled an inappropriate relation, nor provided information that would be particular ly relevant to a death/burial-type problem. Question-forms without qualifications would also come into this class. An example is given in vignette (4).

(4) Suppose there was an aircrash with survivors which happened last week. Where should they be buried?

The hypothesis was that if non-detection of the anomaly were to occur, detection rates across conditions would vary as a function of the qualification-type. Highly relevant qualifications would be most likely to support a death/burial problem domain and therebye distract the processor from the anomaly. This qualification-type would be expected to display the lowest detection rates. Dis-confirming qualifications would be expected to produce the highest rates. The non-disconfirming, but low relevance, qualifications would be expected to lie somewhere between the rates for the other types. The eight manipulations, across the three relevance-levels, are listed in Tables 9.1 and 9.2, together with a brief explanation of the rationale behind each qualification.

TABLE 9.1: Qualifications in Experiment 8: Levels A and B; disconfirming and low relevance qualifications

---

Level A: 1 disconfirming qualification

1. Suppose there was an aircrash with survivors who were mostly **unhurt**. Where should they be buried?

   ['Unhurt' focuses attention on the 'aliveness' of the survivors, hence biasing against surrogation.]

Level B: 4 low-relevance qualifications

2. Suppose there was an aircrash with **survivors**. Where should they be buried?

   [Basic, unqualified condition.]

3. Suppose there was an aircrash with **many survivors**. Where should they be buried?

   [Quantifier added to avoid the possibly stilted (and hence attention-drawing) effect of 'survivors' alone, almost implying that they had not been expected. The referent of 'they' must still be those who survived (c.f. Moxey and Sanford, 1987).]

4. Suppose there was an aircrash with survivors **which happened last week**. Where should they be buried?

   [General situational information of no relevance to burial location issues.]

5. Suppose there was an aircrash with survivors who were mostly **gravediggers**. Where should they be buried?

   [Ascribing a common occupation to the survivors, but unlikely to be of any relevance for burial location issues.]

TABLE 9.2: Qualifications in Experiment 8: Level C; high-relevance qualification

Level C: 3 high-relevance qualifications

6. Suppose there was an aircrash with survivors who were mostly of **no fixed abode**. Where should they be buried?

   [Relevant qualification to burial location decisions: people of no fixed abode would not be able to be sent home for burial.]

7. Suppose there was an aircrash with survivors who were mostly **European**. Where should they be buried?

   [Relevant qualification, because where people's homes are should have an effect on where they are buried.]

8. Suppose there was an aircrash with survivors who were mostly **circus performers**. Where should they be buried?

   [Relevant qualification, because, by inference, circus performers are often nomadic and therefore have no permanent home with natural burial location.]

b: Method

The materials, design, subjects and procedure were the same as for Experiments 4, 5 and 6 using the questionnaire format. Because the question construction in the present experiment was composed out of two sentences, some of the other questions were varied slightly in order to minimise the likelihood of attention being drawn to the fifth question just for that reason. A list of the questions used is given in Table 9.3. The eight conditions in this experiment each reflected a text-based manipulation. These manipulations are shown in Tables 9.1 and 9.2. Eight independent groups of subjects saw only one version of the questionnaire. A total of 205 subjects took part.

TABLE 9.3: A list of the ten questions used in the questionnaire format for Experiment 8.

1. Imagine a group of people who are homeless. Should they have the right to be housed in their home town?
2. Should all citizens have the right to peaceful protest?
3. Should the teacher have ultimate responsibility for a child's education?
4. In the situation of a family death, who officially should be informed?
5. Suppose there was an aircrash with survivors. Where should they be buried?
6. At what age should it be legal to smoke tobacco in Britain?
7. How many MPs should there be in the British parliamentary system?
8. At what age should people be permitted to hold a driver's licence?
9. Should National Service be reinstated?
10. Should students have the right to a state-funded education?

c: Results

(1) Detection rates

Of the 205 subjects tested, 78 had previous knowledge of the problem and their detection status was excluded from the results. Despite having encountered the problem before, 13 of this total failed to detect the anomaly. The number of subjects detecting and non-detecting in each condition is presented, with the percentage detection rate, in Table 9.4 (detection results for subjects with previous knowledge are shown in Table 9.5). The percentage detection rates for the three relevance-levels are presented in diagrammatic form in Figure 9(a), and the rates for all conditions, within their relevance-levels, are presented in Figure 9(b).

Overall there was a reliable influence of condition on detection rate (Chi-square=16.94, d.f.=7, $p<0.05$). The four conditions within Level B (low relevance) were compared and were found not to be reliably different (Chi-square=1.13, d.f.=3, NS). Similarly, there was no reliable difference found between the three high relevance conditions in Level C (Chi-square=0.38, d.f.=2, NS). The detection levels for B and C were compared (with Level B effectively acting as a baseline). The difference was found to be a highly reliable one (Chi-square=8.64, d.f.=1, $p<0.01$). The disconfirming qualification in Level A was tested against both Levels B and C. In each case Level A was found to be reliably better (A * B: Chi-square=4.53, d.f.=1, $p<0.05$. Level A * C: Chi-square=12.5, d.f.=1, $p<0.001$).

TABLE 9.4: Detection rates for Experiment 8.

Condition	**Detections**	**Non-detections**	**Percentage detection**
Level A			
unhurt	**15**	**0**	(100%)
Level B			
survivors	**14**	**5**	(82%)
many survivors	**12**	**5**	(70%)
...last week	**12**	**3**	(80%)
gravediggers	**12**	**3**	(80%)
'B' Total	**50**	**16**	(76%)
Level C			
no fixed abode	**7**	**8**	(47%)
European	**7**	**8**	(47%)
circus performers	**9**	**7**	(56%)
'C' Total	**23**	**23**	(50%)
All conditions	**88**	**39**	(69%)

TABLE 9.5: Detection results for subjects with previous knowledge in Experiment 8.

Condition	Detections	Non-detections
unhurt	10	1
survivors	5	1
many survivors	7	1
...last week	5	0
gravediggers	13	0
no fixed abode	3	3
European	14	6
circus performers	8	1
All conditions	65	13

FIGURE 9(a): Detection rates for the three relevance-levels in Experiment 8.

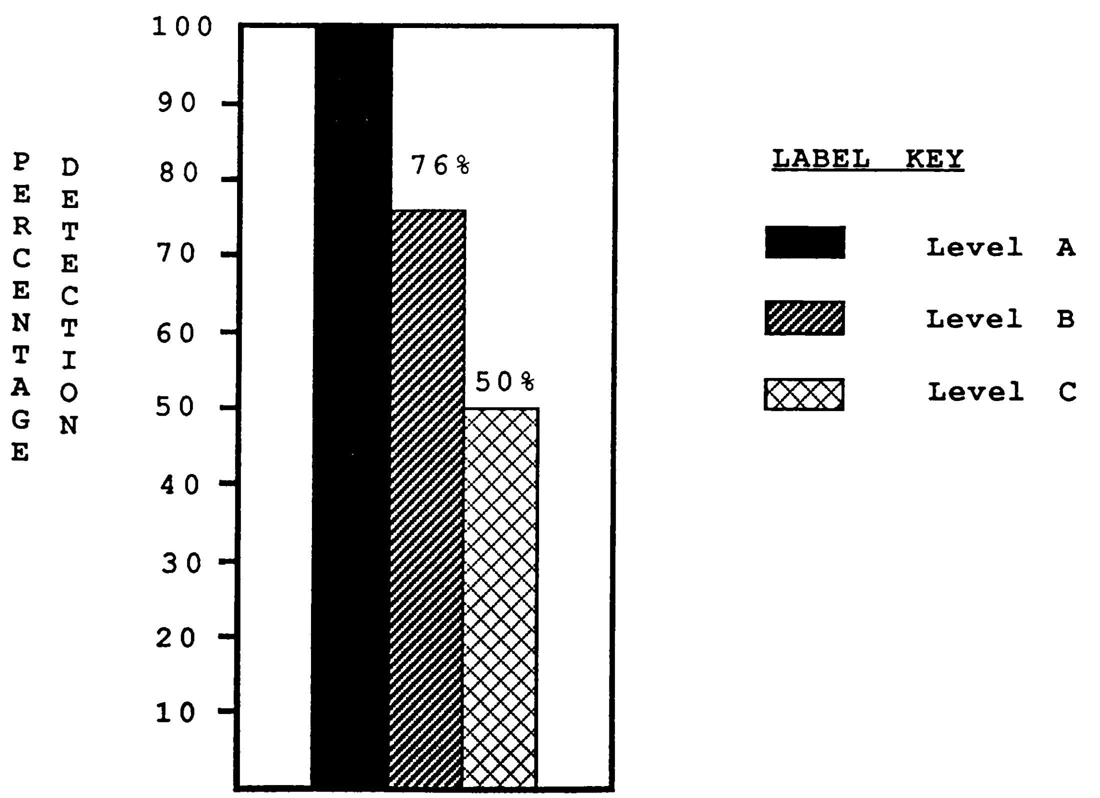
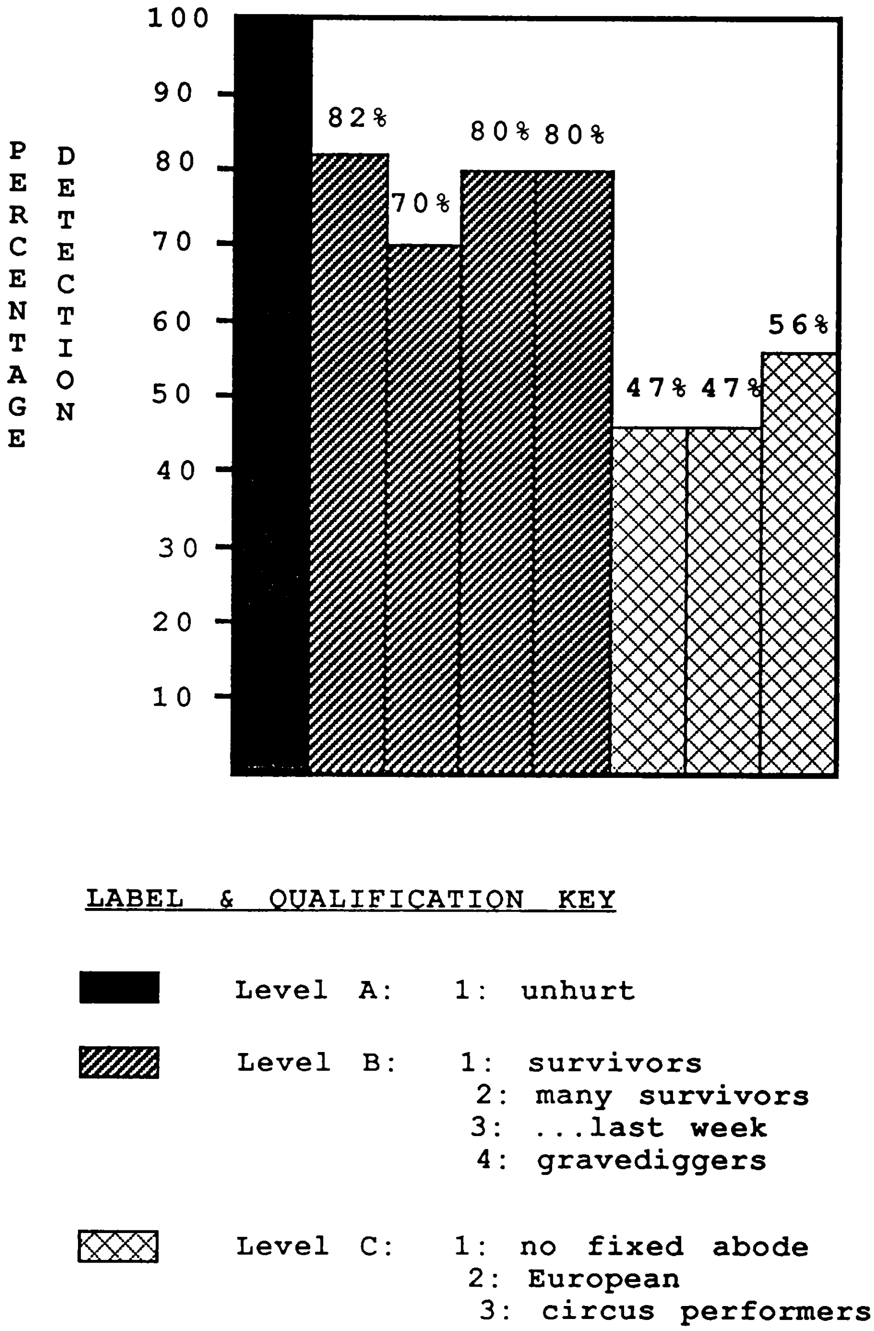


FIGURE 9(b): Detection rates for all conditions in Experiment 8.



{2} Content analysis for the written answers

The same two aspects of the answers (solution-type and reference-expression) were evaluated. The concordant evaluations were achieved from the ratings of the same two judges as before, operating under the same instructions.

The content analysis for the solution-types is presented, for each condition, in Tables 9.6, 9.7, 9.8 and 9.9, and the corresponding analysis for reference-expressions in Table 9.10, 9.11, 9.12 and 9.13. In each case the first tabulates the evaluations for the disconfirming qualification, the second low relevance, and the third high relevance conditions. The fourth provides the totals for all levels. A transcript of the solutions for those subjects contributing to the detection rate is included in Appendix I.

The answer-types for the combined levels appear consistent with the emerging pattern across the studies. Overall 66% of detectors asserted that no problem existed. However, this was as low as 51% in the evaluations for the low relevance conditions (Level B), suggesting that inferences for subsequent death may have been stronger in these manipulations. It is not clear why this should have emerged there but not in the other levels. The discussion section will look in more detail at some of the answer content.

TABLE 9.6: Content analysis for answer-types in Experiment 8: Level A (disconfirming)

	Answer-type				
Condition	**Family**	**Home**	**Crash**	**No prob**	**Other**
unhurt					
<u>15 detectors</u>	-	1	-	11	-
<u>0 non-detect</u>	-	-	-	-	-

TABLE 9.7: Content analysis for answer-types in Experiment 8: Level B (low relevance)

	Answer-type				
Condition	**Family**	**Home**	**Crash**	**No prob**	**Other**
survivors					
14 detectors	3	1	2	6	1
5 non-detect	1	3	1	-	1
many surv.					
12 detectors	3	2	1	5	-
5 non-detect	1	2	2	-	-
...last week					
12 detectors	3	1	1	6	2
3 non-detect	1	1	-	-	2
gravediggers					
12 detectors	-	-	-	7	3
3 non-detect	-	-	-	-	3
Total					
(50 detectors)	(9)	(4)	(4)	(24)	(6)
(16 non-detect)	(3)	(6)	(3)	(-)	(6)

TABLE 9.8: Content analysis for answer-types in Experiment 8: Level C (high relevance)

	Answer-type				
Condition	**Family**	**Home**	**Crash**	**No prob**	**Other**
no fixed abode					
7 detectors	-	-	-	7	-
8 non-detect	1	2	2	-	3
European					
7 detectors	-	1	-	3	1
8 non-detect	1	8	-	-	-
circ.perform.					
9 detectors	-	1	-	7	-
7 non-detect	2	1	2	-	3
Total					
(23 detectors)	(-)	(2)	(-)	(17)	(1)
(23 non-detect)	(4)	(11)	(4)	(-)	(6)

TABLE 9.9: Content analysis for answer-types in Experiment 8: Totals for all levels

Condition	Family	Home	Crash	No prob	Other
	Answer-type				
Totals for all levels					
(88 detectors)	(9)	(7)	(4)	(52)	(7)
(39 non-detect)	(7)	(17)	(7)	(-)	(12)

The reference expressions conformed to the emerging pattern for the questionnaire-based manipulations. The number of references were again far fewer than in the narrative format with a large number in the 'other' category. Most detectors referred to the passengers using the anomalous label. The were no apparent differences across conditions or levels.

TABLE 9.11: Content analysis for reference-expressions in Experiment 8: Level B (low relevance)

Condition	Dead	Bodies	Gener-al	Anoma-lous	Other
	Reference-expression				
survivors					
14 detector	1	-	-	4	4
5 non-detect	-	-	-	-	1
many surv.					
12 detectors	-	-	-	6	1
5 non-detect	-	1	-	-	2
...last week					
12 detectors	5	-	-	4	3
3 non-detect	1	-	-	-	1
gravediggers					
12 detectors	-	-	-	3	4
3 non-detect	-	-	-	-	1
Total					
(50 detectors)	(6)	(-)	(-)	(17)	(12)
(16 non-detect)	(1)	(1)	(-)	(-)	(5)

TABLE 9.10: Content analysis for reference-expressions in Experiment 8: Level A (disconfirming)

Condition	Dead	Bodies	General	Anomalous	Other
			Reference-expression		
unhurt					
15 detectors	1	-	-	2	4
0 non-detect	-	-	-	-	-

TABLE 9.12: Content analysis for reference-expressions in Experiment 8: Level C (high relevance)

Condition	Dead	Bodies	General	Anomalous	Other
			Reference-expression		
no fixed abode					
7 detectors	-	-	-	3	3
8 non-detect	-	-	-	-	3
European					
7 detectors	-	-	-	2	1
8 non-detect	-	-	-	-	1
circ.perform.					
9 detectors	-	-	1	2	3
7 non-detect	-	-	-	-	3
Total					
(23 detectors)	(-)	(-)	(1)	(7)	(7)
(23 non-detect)	(-)	(-)	(-)	(-)	(7)

TABLE 9.13: Content analysis for reference-expressions in Experiment 8: Totals for all levels

	Reference-expression				
Condition	**Dead**	**Bodies**	**Gener-al**	**Anoma-lous**	**Other**
Totals for all levels					
(88 detectors)	(7)	(-)	(1)	(26)	(23)
(39 non-detect)	(1)	(1)	(-)	(-)	(12)

d: Discussion

An overall detection rate of under 70% indicates that surrogation under this format occurred for over 3 out of every 10 subjects. Considering the structure of the two-sentence question, with the anomalous object and verb in separate sentences, this is a remarkable finding. Even processes of anaphoric resolution appear susceptible to incompleteness, or at least are subject to global cohesion-establishment that can effectively reduce the tightness of the resolution test.

The initial hypothesis was that partial cohesion and relevance-to-problem factors may shape global cohesion-establishment, with the possible outcome of biasing against anomaly detection in the high-relevance conditions. The results confirm this hypothesis with the astonishingly low detection rate of 50% in Level C. Detection rates on the three levels differed reliably from each other, and the ordering showed the anticipated trend. The disconfirming qualification in Level A resulted in the first 100% detection rate in the case-study, and the low relevance conditions settled at the intermediate rate of 76%.

The detection rate data confirm that the contribution which the meaning of **survivors** makes on the representation is a function of the global cohesion constraints. This in turn is shaped by the other information presented that is relevant to answering the question. Since providing relevant information to the death/burial problem-domain results in lower detection rates, it is concluded that this effectively supresses the contribution that the **survivors** item would otherwise make. These findings are taken to support the view that initial cohesion can be partial or shallow, on-hold for the significance of subsequent information to be established. Further, information that is relevant to the commited theme or topic (when it is established) may be brought into focus and processed in-depth over other items.

The role of relevance is further confirmed when close inspection is made of the content of subject's answers. Notice in particular that in the high relevance **European** condition eight concordant evaluations were made to the 'home' answer-type, suggesting that all eight non-detectors focused on this aspect in their solutions. The written answers, taken from Appendix I, confirm this to be the case.

1. Where they came from originally.
2. In their own towns with memorial on site of crash.
3. Their home town if they later died in hospital.
4. In their home countries.
5. At home if possible.
6. In their home town if possible. If not at a special burial ground near the site.
7. Where their families desire, probably back home.
8. Their own home town.

As anticipated, a qualification like **European** was sufficiently relevant to a burial location problem for the anomaly to be detected less than would otherwise have been expected. Further, the introduction of nationality-type has focused subject's attention onto the passenger's homes as a viable solution to the problem. Although the other high relevance qualifications do not show the same blanket effect, it is clear from inspection of the non-detector's answers in those conditions that the qualification has had a direct bearing on the devised solutions. The solutions of the eight non-detecting subjects in the **no fixed abode** condition and the seven in the **circus performers** condition are shown below. In particular, notice that 'home' is only mentioned once in the **circus performers** manipulation, and even then only on the condition that they have a permanent home.

'No fixed abode' non-detectors:

1. Nearest place.
2. The country where they lived for most of their lives, or where their relatives are.
3. They should be cremated.
4. In a graveyard! -everyone is entitled to respect.
5. Buried or burned. Memorialised at the place of the crash.
6. State cemetries -- perhaps even set a new one up near disaster site.
7. Where their nearest families wished them to be buried.
8. In their country of origin - i.e. if British, buried in UK.

'Circus performers' non-detectors

1. Where their friends and family would like them to be buried.
2. In the country that the plane crashed in, or in the country of origin of the plane.
3. Anywhere.
4. In the country where the plane crashed.
5. Wherever they had closest ties, otherwise as close as possible to crash site all together.
6. If leave a will - where they choose. If have a home of own, there. If parental home only, there. If none, within jurisdiction where they died.
7. Depends on what their will says, or family say.

Similarily, the solutions of non-detectors in the other levels (though fewer in number) could reflect the content of the lower-relevance qualification. The best, and most amusing, example of this comes from the **gravediggers** condition where of the three non-detectors, two attempt to integrate the qualification material into their solution.

1. In a graveyard.
2. Where they worked.

Part Four

Chapter Ten

Conclusions

10.1 Summary

The attractive properties of the case-study materials are discussed and arguments are presented for the generality of the findings. The major findings are summarised with respect to local, scenario and pragmatic influences on processing completeness. Implications for real-time processing are discussed, and sensitivity to the time-course of interpretation emerges as a strong theme, particular ly in relation to on-line measurement. A constraint satisfaction implementation of the case-study materials is described. Properties of this network accurately model the detection rate results of the experiments, and confirm the computational tractability of the major claims. Remaining questions and broader issues are considered.

10.2 Task-related issues

The fact that anomalies can pass undetected in texts indicates some kind of shallowness, or incompleteness in processing. The anomaly selected for the case-study has proved to be a fruitful example. It concerns items which are bound to inappropriate roles (survivors as objects for burial). The semantics of the items used in examples such as this are often a close enough match to appropriate fillers for interpretation to occur as though a proper filler had been written. The full meaning of the real term frequently fails to make an impact on the representation, and hence anomalous descriptions can pass undetected. Indeed, it was argued that 'full meaning' may be a useless concept, and consequently the idea of a complete match is not algorithmically definable.

The frequency of failures to detect the anomaly is an index of the <u>contribution</u> the relevant descriptions make to overall coherence. Hence, low detection rates signal conditions where there has been only a minimal impact. The strength of the error detection paradigm is that text-based manipulations can be easily incorporated and tested. Comparisons of detection rates proved to be informative about factors controlling the impact afforded to the anomalous items.

TABLE 10.1: Possible background expectations from the case-study materials.

Aircrash:

Evokes scenario information which provides expectations about the sort of objects that will be referred to and possible outcomes of the events:

planes, **pilots**, **stewardesses**, **passengers**, **panic**, **fear**, **damage**, **death**, **wreckage**, **rescue services**, etc.

Bury:

Evokes an underlying schema representing the primitive actions of burial.

(1) Hole made in ground with instrument(s) by agent(s).
(2) Object(s) placed in hole by agent(s).
(3) Object(s) covered with the material removed to make the hole.

In broader contexts, mention of burial may fine-tune expectations concerning the inferable objects, agents and instruments in question.

(a) 'Dogs':	agent: dog(s) object: bone instrument: paws
(b) 'Smugglers':	agent: smuggler(s) object: treasure instrument: spade
(c) 'Aircrash':	agent: **families/authorities** object: **dead bodies** instrument: (**funeral**)

The items in the present case-study have provided a particular ly flexible research tool because the magnitudes of non-detection have permitted meaningful cross-condition comparison. Had the items been easier or more difficult to detect, less could have been concluded from such rates. Mention of burial in the context of an aircrash may be sufficient for predictions of funerals and dead passengers, and these expectations will have contributed to the levels of non-detection. Table 10.1 provides a rough illustration of the sort of expectations assumed to result from the main case-study materials.

The generality of the task

There are two reasons supporting the view that the case-study reveals general properties of interpretation. First, the items were selected from a range of possible materials, all of which are established anecdotal 'tricks', and some of which have been subject to experimental investigation (Wason and Reich, 1979; Erickson and Mattson, 1981; Ehrlich and Loridant, 1990). The relevance of these anomalies goes beyond their status as tricks. What makes them effective tricks is an important processing issue. The property of interest for all the anomalies introduced, is that background constraints prevent an input-string from obtaining bottom-up interpretation. Conditions which are consistent or expected become part of 'preferred' interpretations. Items in the text can be prevented from making the impact that would be expected of them.

(1) This book fills a much needed gap.

(2) How many animals of each kind did Moses take on the ark?

(3) Can a man marry his widow's sister?

(4) How much soil is there in a hole that is two metres deep, six metres long and three metres wide?

(5) When an aircraft crashes where should the survivors be buried?

Clearly, the survivors trick is not an isolated example. The general phenomenon is well-known.

Second, there has been far more non-detection in the case-study than was anticipated. Such levels of non-detection must be dependent on a considerable degree of shallow processing. It would be surprising if corresponding effects were not obtainable with materials forcing slightly weaker expectations. Experiment 6 provided a case in point. Non-detection occurred in the bicycle-accident materials, but weaker predictions for death and burial in this context increased detection overall. Consequently, detection rate analysis may be a less sensitive measure of incompleteness under weaker conditions. Manipulations within the bicycle accident domain, for example, are less likely to provide numerically informative results. All conditions are likely to display reasonably high detection rates.

Further, it is not consistent with the arguments presented in Chapter 2 that incompleteness is specific to occasions of anomaly non-detection. **There can be no such thing as an exhaustive analysis on a mapping test, so incompleteness is inherent to all cohesion-based interpretation**. The following section summarises the main effects of background constraints found to have an influence on the case-study materials.

10.3 Major findings

In all the major findings, high levels of expectation have had an effect on detectibility. Expectations of how the discourse 'should' cohere can bias cohesion-establishment. Evidence for these effects will be summarised under three headings: local, scenario and pragmatic influences.

a: Local effects

These include the surface structure and semantic-based emphasis manipulations tested in Experiments 2, 3 and 5. The most extreme result came from the **surviving dead** condition in Experiment 3. The low rate of detection indicates that the impact of the qualifier is attenuated despite its presentation <u>prior</u> to the **dead** noun. **The time-course of interpretation for some items can be sufficiently extended for subsequent items to dominate their processing**. It is known that delayed effects are present where subsequent material helps to disambiguate competing interpretations. Bierwisch (1983) describes this as a cesation of on-line interpretation,

'if there are intervening operations that require previous results to be retained for later processing.'

He presents the following example, (6).

(6) In preparing the lecture, he badly missed the book, as it was full of important (notes/insights).

Full interpretation of the role of the book depends on the status of the final word. Hence, there is a delay in processing. The case-study results are more surprising than this because later material <u>swamps rather than disambiguates</u> the contribution of earlier items. It is not a simple case of helping to make interpretation more complete.

We know that parsing systems have been developed to optimise incremental processing (e.g. Mellish, 1981). This type of parser could not handle the delayed interpretation effects. Any system of this sort would not predict such strong

processing effects for later items over prior ones.

b: Scenario level effects

The results of Experiment 6 confirm a selective influence on inference at the scenario level. The degree of compatibility between references and the scenario has a controlling effect on cohesion. Highly predicted references are 'trusted' more to map appropriately at lower levels of representation (e.g. verb schema). If a reference is not expected in the scenario then more processing is required to establish its relevance than would otherwise be the case. As a result, where anomaly is present, detection is more likely. Since interpretation at the level of the present anomaly must depend on scenario effects, local structures, such as phrasal components or propositions, are not necessarily established prior to interpretation. Hence, any view of processing that builds local structures prior to global evaluation is a massive oversimplification.

c: Pragmatic level effects

The major finding of the case-study came from Experiment 8. Cohesion patterns depend upon pragmatic factors. The sort of question that is being asked, and the type of representation built to answer it, has a direct effect on cohesion for the message. Processing depth is allocated selectively on the basis of the **relevance** of available material to the pragmatic factors (e.g. the 'point' of a question). Since the pragmatic status of a question may not be established early in a message, inference allocation on previous items may be delayed. Importantly, these effects can be su pra-sentential. In summary, full incremental interpretation is not a hard processing constraint. **Pragmatic factors exert influence over the contribution a word makes to overall coherence.** The impact on interpretation of this level of influence may be delayed until pragmatic status is established.

10.4 Implications for real-time processing

All of these findings have corresponding implications for real-time processing. What is absolutely clear is that anomaly detection is not fixed to the point of anomaly presentation. Full interpretation can have an extended time-course. Hence, the contribution of a word to the coherence of a message is a function of the status of prior and subsequent items, including items in the next sentence. This level of influence is not typically recognised in word-processing studies.

In relation to anomalous materials, the general implication is this. **There is less inferential activity around anomalous items in conditions of low detection.** Consequently, the 'details' of items in those conditions make a weaker contribution to overall coherence, and will have less impact on related concepts testable via probes. There is a variety of on-line measures that were considered for the present case-study. These included eye-movement recording and probe techniques (e.g. word naming, lexical decision). However, the quality of detection rate data was so good, particular ly in the later experiments, that this paradigm was extended as far as possible within the constraints of the thesis. Importantly, there are wider implications for real-time evaluation that are not specific to the present materials or to anomaly detection. These are summarised below.

Since local structures are not necessarily completed before 'higher level' interpretation, a principle emerges concerning on-line mapping.

PRINCIPLE: If sufficient match is obtained between two active patterns at a processing site, then the system does not continue to seek extra evidence for cohesion.

As we have seen, 'sufficient' may be quite a small amount. Hence, (possibly disconfirming) information <u>available</u> to the system may not be attended to if cohesion demands have been met.

This principle can be extended to 'sufficient matches' at the scenario level. Scenario-evoking items make such a strong contribution to coherence, that the contribution of other items may be low in comparison. Real-time processing will be a function of the background structures signalled by items, and not necessarily a direct result of local manipulation (e.g. word-length). The structures signalled by items could effect 'sufficient' matches such that some relevant information is not fully attended to. This may be particular ly true for information from small structures when being mapped into larger ones.

An example from the case-study is Experiment 6. The time-course of impact for 'survivors' and 'aircrash' must overlap sufficiently for the contribution of former to be weakened.

(6) Where should the survivors be buried after an aircrash?

In probe studies, these influences have to be accounted for. Interpretation of a whole message, and interactions between text-based items, may not be fully established at the onset of a probe. Hence, a probe may only provide a snap-shot of activity to that point.

This observation is extended to the influence of pragmatic factors. Since communicative variables have been found to effect interpretation at the cohesion-level, comparable effects (e.g. lexical access) may be susceptible to fluctuation across utterance-type. Generalisation of local effects to more global situations may not be a simple matter.

10.5 Realisations

We have seen in the Sentence Gestalt network of St.John and McClelland (1990), that multiple constraint satisfaction models of comprehension can afford greater impact to contextual constraints than input constraints. On some occasions these graded constraints are such that items introduced explicitly are swamped and inappropriate mapping takes place. A similar model could be invisaged for the effects of background constraints in this case-study. One such model has been developed by Paterson (1991) and is part of work in progress extending the case-study topic (Barton, Moxey, Paterson and Sanford, in preparation: Barton and Sanford, in press).

Paterson's network represents concepts as nodes, and degree of association between concepts as the weighting on the arcs between nodes. Individual units represent hypotheses about the meaning of an input, and the degree of activation reflects the strength of evidence in support of particular hypotheses. The top level of the network is presented in Figure 10(a).

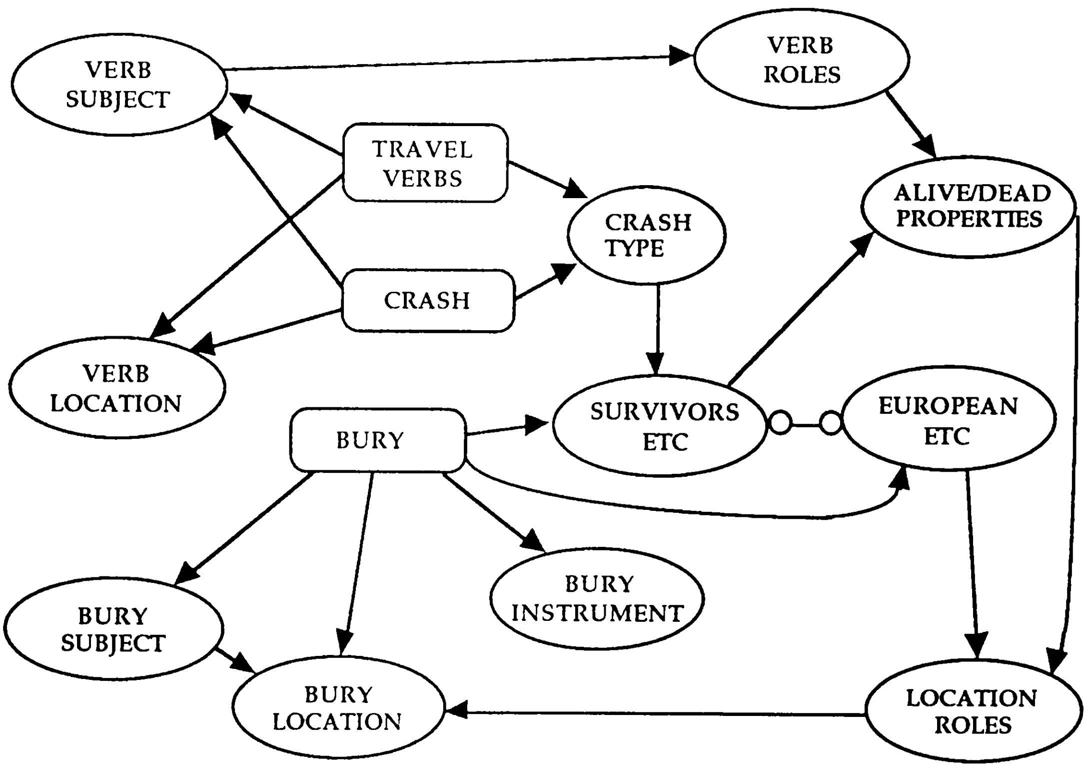
Each of the labels represents a pool of weighted units. The network represents input as the external activation given to input-nodes, and top-down expectations are the pre-set, and assumed to be pre-learned, associations between network concepts. An input configuration corresponds to the principal lexical items in the relevant condition of the case-study. The key aspects of the input are activated in the network simultaneously, and this results in the net settling on an interpretation of the input. The interpretation is the set of concepts that are active when the net is fully relaxed.

One example would be to model interpretation of (7) by activating the FLY, CRASH, SURVIVORS, and BURY units in the network (some of these are embedded as part of the processing pools).

(7) When an aircraft crashes where should the survivors be buried?

The network is then cycled until fully relaxed. Detection rate of the anomaly is modelled by the number of cycles required for the net to relax on a 'dead' interpretation for the object of burial (i.e. full activation of the 'dead' concept node).

FIGURE 10(a): Top level description of a constraint satisfaction model for the 'survivors effect' (Adapted from Paterson, 1991).



Without going into the details of the network, the patterns of activation over the processing units settle to the point where detection rate is accurately reflected by the number of relevant cycles. This is true for both label variation (injured, maimed etc.) and scenario variation (aircrash, bicycle accident).

Like any model, only aspects of the 'thing-in-itself' are represented. Consequently, it has important limitations. Since input is evaluated simultaneously, the time-course of impact for different items is not modelled. Hence, the slippage effects for strong over weak items are not represented in time. However, the important point is this. Incomplete processing has been modelled in a principled way in a constraint satisfaction network, and this model has implemented the content of the case-study materials. Background constraints are found to influence processing depth (e.g. scenario-type) and can exert sufficient impact over input activity for 'mis-interpretation' to occur. Paterson's model shows that the results and claims of the case-study are coherent with constraint satisfaction realisations.

Whatever form of implementation is invisaged for the range of investigated influences, there are two main demands on any such model. First, global interpretation has to rely on interactions between different levels of constraints (e.g. semantic, thematic, scenario, etc.). Such constraint satisfaction must be capable of manifesting graded impact across constraint-types. Second, the impact of any particular constraint cannot be a simple function of only its own properties, nor can its full impact be described adequately from observation of initial effects. Subsequent impact from other constraints can shape overall outcomes. Hence, stepwise decision-making mechanisms are incongruous with the case-study findings.

10.6 Questions

a: Extensions to the task and general error detection

Detection rate analysis with the present task can be extended with other manipulations. One example would involve thematic variation in order to manipulate the role-status of the (dead) passengers. The anomaly in the case-study is anchored to a relatively subordinate role (burial object). Increasing status may bring potential fillers further into focus. As a consequence, more inferential activity would be expected around such items, even if other expectation levels remained constant. An example using the narrative materials is presented in vignette (8).

(8) There was a private plane flying from Milan to Barcelona. It was carrying the Inter Milan football team to the Europan Cup final. On the last leg of its journey, the pilot started to lose control. The plane eventually crashed on the border of France and Spain. The wreckage was equally strewn over the boundary. The authorities were deciding where to bury the survivors. They couldn't make up their minds.

Question: What would your solution to the problem be?

Another extension concerns the swamping of other types of discourse constraint. One example would be the influence of background expectations over syntactic information. Consider the following as a possible manipulation in the questionnaire paradigm.

(9) Imagine an aircraft crashes in a remote part of the world. Should the dead be allowed to bury the survivors?

Significant non-detection of the anomaly would indicate that subject and object fillers for a verb can be interchanged from their text position into more plausible slots. This example is reminicent of the 'much needed gap'. To the author's knowledge, no empirical research has established the boundary conditions for such effects.

Although anomaly detection has been employed in this thesis as an 'inferential probe', considerable interest exists in error and error detection in its own right (e.g. speech errors, Fromkin, 1980; Cutler, 1982: cognitive errors, Reason, 1984). The drive behind this type of work is to obtain a better understanding of the factors contributing to human error. Of course, failure to detect anomaly in texts is a kind of error itself. Consequently, the mis-interpretations of the non-detectorsare also relevant to error-based perspectives on human cognition.

In relation to language, relationships between production error and comprehension error may be particularily intriguing. The extent of the 'Colemanballs' phenomena in every day communication is not well-known, and their production not well understood. Profit could be made from focusing on influences effecting the translation of coherent thought into an incoherent message. Also, why some of these phenomena arouse amusement (e.g. mixed metaphor) may indicate important properties of interpretative stability. There is a dearth of serious inquiry into these sorts of issues.

b: Focus-based cohesion and embellishment

The general findings suggest that as a theme, role or description comes more into focus, it is subject to more inferential activity. Hence, a close correspondence is posited between focus and inference. Indeed, one way of defining focus may be in terms of inference patterns. An attractive property of focus-based inferential control is that inferential explosion is prevented as a natural outcome.

Exploring in depth only those fields that are central or relevant to the theme prevents over-processing of plausible inferences for less important relations.

Text-based processing models developed as recently as 1988 are still advocating the establishment of local structures and generation of inferences before background information is accessed. As Kintsch (1988) writes,

'In contrast to expectation-based, predictive views of discourse comprehension, a model is developed where initial processing is strictly bottom-up. Word meanings are activated, propositions are formed, and inferences and elaborations are produced without regard to the discourse context'.

The inaccuracy of such views has to receive wider appreciation. We have seen that cohesion processes demand considerable elaboration. Background information provides the expectations for discourse-objects to be bound together in a coherent way. One outcome of this is that cohesion-based activity may be closely related to embellishment inference. For example, in the case-study materials, shallow processing on the object slot for the 'bury' verb might imply that embellishments are less likely on fillers for that role. The reader is less likely to embellish on a relation established by a shallow process in the first instance. This hypothesis would predict that embellishments are more likely to be made about the authorities than the survivors/dead, and this is testable via probes for related concepts (e.g. word naming).

Should it be the case that cohesion, coherence and embellishment activity are under unified control, this would be a major breakthrough in understanding inferential activity. At present, the different types tend to be treated in separate analyses. This is an eradicable weakness in inference research.

Appendices

Appendix A: Filler texts for the narrative-based experiments

Experiments 1, 2, 3 and 7. The filler texts for Experiment 7 have an additional (bracketed) sentence to match text length.

Text 1: presented by the experimenter displaying the procedure

A football team is travelling by coach to play in a match. The coach breaks down with 5 miles to go. There is no other available transport. The team would forfeit the points by arriving late. The players could run the rest of the way. If they did this, they would be very tired and would surely lose. (Promotion depended on getting a win). They couldn't decide what to do.

Question: What would your solution to the problem be?

Text 2: first experimental text: solution provided by subject

A man and wife decide to emigrate to Australia. The wife has managed to secure employment. The husband has no job to go to, so the authorities wont let him in. The wife could try to find the husband a job when she gets there. If she failed to do this, she would have to pay her own way back. The couple couldn't afford the expense of a return flight. (Finance was a problem at the best of times). They were trying to make up there minds what to do.

Question: What would your solution to the problem be?

Appendix B: Written solutions for Experiment 1

Condition: '**survivors**'

<u>10 detecting subjects</u>

1. The <u>survivors</u> would not need to be buried!
2. Why bury survivors?
3. There would be no need to bury the survivors as they are not dead.
4. Bury the survivors? Did I read too quickly? Bury the dead in their home towns.
5. You do not bury survivors.
6. You don't bury survivors.
7. Why are the authorities burying survivors?
8. They should bury them in whatever country they came from, if they can be identified.
9. They didn't.
10. Survivors don't need to be buried.

<u>7 non-detecting subjects</u>

1. Ask the relatives where they wished their dead to be buried, or if you can't find out the nationality of most of the people on the plane, bury them there.
2. The survivors relatives should be contacted and their remains flown home for burial.
3. Ask the relatives of the deceased where they would prefer them to be buried.
4. Bury the dead where their relatives decided .i.e. some might want them transported home. Otherwise bury them in the nearest town.
5. Survivors should be flown back and buried in their home countries.
6. The survivors should be buried wherever their relatives wanted them to be buried.
7. All the passengers should be buried in their home countries and if desired a monument be built on the border to please both countries.

Condition: '**injured**'

1 detecting subject

1. I wouldn't bury the injured because they might be dead.

14 non-detecting subjects

1. Find the country of origin of the dead & bury them in whatever country is closest to their homeland - Either France or Spain.
2. Find identity of passengers, contact relatives and see if they wanted the bodies buried.
3. Bury the people in the geographically nearest cemetary (?) if the relatives had no objections - else fly the bodies home.
4. Ask their families opinion. If can't get their families then investigate the availabilities of the burial ground to decide.
5. Send the bodies home - cost to be met jointly by both authorities.
6. If people could be identified they could be taken by families or families decide where to bury them. If they are not identified they should be buried in the country they were found.
7. The bodies should be buried in Spain because that was the destination of the plane.
8. Send them to their own homes for burial & split the cost of doing so between the two countries.
9. Contact the **relatives** wherever possible and let them decide. (Bury them according to place of birth on passport.) If no passport fly them back to where the plane flew from and bury them there.
10. They should bury the dead in their own countries of origin, or wherever they lived.
11. Authorities should find out who was on the plane and ask the surviving relatives where they would prefer to have their relations buried since they would have the best idea. Not really a decision for the authorities.
12. Bury the dead passengers in their respective countries of origin.
13. The authorities should contact the relations of the

people who were killed and ask them if they want their deceased buried in their place of origin (-where they used to live) or if they want them buried where they were killed (allowing them to make a choice between the two countries).

14. Find the relatives and ask what they would to be done with the bodies.

Condition: **'maimed'**

<u>3 detecting subjects</u>

1. The dead should be buried in their home country, which would probably be Austria; the cost of returning the bodies should be met by the airline company.
2. Which ever country they were found.
3. You don't bury the maimed, if they're, presumably, still alive.

<u>10 non-detecting subjects</u>

1. Fly the identifiable bodies home, and depending in which country the bodies are kept bury them there.
2. Dependent on nationality - express wishes of bereaved.
3. Bury the bodies all in one place - France.
4. Bury them where their relatives wanted them buried.
5. The deceased should be buried in their country of origin, or if this isn't possible if any of the victims had relations etc. in France then they should be buried there. The same should also happen with those buried in Spain.
6. They should bury the remains in the nearest plausible site to the crash area.
7. Ask the families of the dead people where they would like them to be buried.
8. Ask their families.
9. They should send them home for burial.
10. The people should be buried where their family & friends think suitable.

Condition: '**wounded**'

<u>3 detecting subjects</u>

1. The wounded are not dead.
2. The wounded do not need to be buried as they are obviously not dead yet.
3. I would bury the wounded in neither Spain nor France as it would be unfair to take people's lives like this.

<u>9 non-detecting subjects</u>

1. Bury people in the country they came from.
2. They should both be equally involved in the burying of the bodies.
3. Authorities from which the countries were dealing with the problem? - Bury in authoriites own countries. Nationalities of passengers taken into consideration (unless all Austrian) i.e. Spanish buried in Spain. Return bodies to original country for relatives.
4. They should decide where to bury them by finding out where the majority of passengers came from.
5. Let the relatives decide where they want the bodies to be buried.
6. This does not say what nationality the tourist was. Surely he would be buried in his native country.
7. In Vienna, and Barcelona, because that's where most of the dead would have come from.
8. To ask the relatives of the people who had been killed, where they would like the people to be buried.
9. Bury the Spanish ones in Spain and the French ones in France.

Appendix C: Written solutions for Experiment 2

Condition: **'surviving passengers'**

<u>9 detecting subjects</u>

1. They wouldn't bury the surviving passengers as they are still alive.
2. Bury half one side, half on the other.
3. Contact the families ans see what they decide. Otherwise which ever country has the better facilities.
4. You don't bury surviving passengers because they're alive.
5. You don't bury surviving passengers.
6. If possible contact relatives and let them decide.
7. Simplest way is to both pay for a joint memorial at site of crash - on the border.
8. If passengers survived you wouldn't need to bury them.
9. Can't bury surviving passengers.

<u>10 non-detecting subjects</u>

1. The relatives of the dead should take over. They could decide where to bury them.
2. The dead should be buried in their native country.
3. Where someone is buried is fairly menial. More importantly, are their family and friends going to be financed to be there.
4. can't remember.
5. Establish the nationality of each tourist - bury them in either of the two countries nearest their homeland.
6. Return all the passengers to their country of origin. If not known, bury them in France where the plane left.
7. I would contact the relatives and let them decide.
8. Return their bodies to the countries of origin. e.g. Spanish people buried in Spain, English in England etc.
9. The authorities should share the responsibilities.
10. Contact the families of dead, as they might want to have them buried in their own countries.

Condition: '**surviving injured**'

<u>11 detecting subjects</u>

1. You don't bury surviving injured people.
2. In France.
3. Surviving victims? Don't need burial.
4. Wouldn't have to bury surviving injured - take them to hospital.
5. The surviving injured shouldn't be buried (at least not until they are dead).
6. They would not want to bury survivors.
7. If they were only injured you shouldn't bury them - if they survived!
8. Don't bury the surviving injured.
9. I don't need to bury the survivors.
10. Don't need to bury survivors.
11. If they were surviving injured you wouldn't bury them.

<u>6 detecting subjects</u>

1. Bury the people where the families suggested.
2. If possible to identify the bodies, ship the bodies back in casks to their respective countries.
3. Find out about relatives and find out from them where they would like the deceased to be buried.
4. The dead should be buried in their country/place of birth.
5. The dead should be buried in Spain due to the fact that this was their destination and the majority on the plane would have been Spanish or Austrian. Therefore France would have been a neutral country with little of the dead coming from France.
6. Return the dead to their home country.

Condition: '**surviving maimed**'

<u>8 detecting subjects</u>

1. The surviving maimed wouldn't need to be buried.
2. If there was survivors they wouldn't need to be buried.
3. Don't bury the surviving maimed.

4. To bury them in the country they originally came from.
5. Why would they want to bury the survivors?
6. Do not bury the survivors, allow them to go to hospital for treatment - then home.
7. Make the airline pay to have the bodies returned to their country of origin.
8. Depending where each individual person landed - the French take responsibility for those who landed in France, Spanish take responsibility for those in Spain. (Perhaps let any relatives have a say).

<u>6 non-detecting subjects</u>

1. If the bodies were identifiable their families should be contacted and asked where they would like the member of the family to be buried. If the bodies were unidentifiable then they should be buried (with the others) in either country depending on the agreement of the authorities and availability of land etc.
2. Solution would be to bury the bodies in the places of their home. i.e. some in Austria some in Spain, France, rest of World etc.
3. Bury the survivors in both territories depending on where each landed - to be fair.
4. Either to bury the people in the country they had died in or to send them (the bodies) to their country of origin for burial.
5. To bury where the families requested them to be buried.
6. Ship all the bodies home. If some are unable to be traced, set up some memorial where the plane crashed - or nearby to the site.

Condition: **'surviving wounded'**

<u>10 detecting subjects</u>

1. You really don't need to bury the surviving wounded.
2. Why bury the surviving wounded.
3. The surviving wounded should not be buried for if they survived they are still alive.
4. Don't bury them at all.

5. You don't bury surviviors or the wounded.
6. The injured should go to the nearest hospital as soon as possible while the dead should be brought back to their own countries. Arrangements have to be made by the airlines company and the France and Spain governments.
7. Allow the families to decide.
8. Where there will be the nearest place for this person. Either Spain or France. It makes no difference.
9. If they were burying survivors surely they should be locked up. May be read it wrong. If they were burying dead, it should be in home countries.
10. Find out where they come from and send them home to be buried.

<u>3 non-detecting subjects</u>

1. Contact the next of kin of all the dead and ask them where they would like their relation(s) buried. If the dead has no next of kin at all he should be buried in whichever country he/she crashed in.
2. Fly all the bodies home to relatives etc. instead of burying them in either of the countries.
3. The bodies should be flown back to the country of their nationality.

Appendix D: Written solutions for Experiment 3

Condition: **'passengers who survived'**

<u>9 detecting subjects</u>

1. You don't bury survivors.
2. Ask the families of the dead what their preference would be.
3. You don't bury live people.
4. The passengers who survived would hopefully not want to be buried - if they did, I would let them choose their own location!
5. When the passengers that survived eventually died allbeit of natural causes or otherwise, their families could decide where to bury them.
6. Any dead passengers should be buried in their own country. They cannot bury passengers who survived.
7. They were alive so why bury them?
8. <u>Don't</u> bury passengers who surviv ed!
9. Why do they want to bury survived passengers? Let them go home.

<u>12 non-detecting subjects</u>

1. Bury them in their own home town.
2. Find out where the passengers come from (if possible) and bury them in the country of their birth.
3. Send the bodies back to their families.
4. On finding out the names of the passengers bury each of them in their rightful country where they lived or were born.
5. Bury the identifiable bodies in their home countries and the others as close to the border as possible - or scatter their ashes at the border.
6. Send the dead passengers back to their country of origin to be buried.
7. Bury people in their country of origin. Otherwise, where the 'nicest' graves would be.
8. Each person should be buried where their relatives wish.

9. The passengers should be buried in their own countries wherever they were travelling from.
10. They should send the bodies back home.
11. Toss a coin.
12. The passengers should be buried in their home country i.e. wherever they originated from.

Condition: **'surviving dead'**

<u>4 detecting subjects</u>

1. You don't get surviving dead. They're either alive or dead! The dead people should be returned to their country or relatives.
2. Don't bury the surviving dead at all since they are not dead.
3. Give the relatives charge of the deceased.
4. All bodies to be taken into Spain, then if no information could be found about them or their relatives bury them in Spain.

<u>16 non-detecting subjects</u>

1. Have the relatives decide where they wanted the burials to be.
2. Call up the families and ask them where they want to have their chunks of meat put in.
3. Identify the bodies if possible and let the relatives decide or bury them anywhere i.e. France or Spain if the authorities there will comply.
4. It should not be up to the authorities where the dead should be buried, but instead by the relations of the deceased.
5. Bury the people in which ever country had the most 'burying' space.
6. Inform relatives and ask their opinion.
7. Bury them in country of their own origin. e.g. if French, bury in France etc.
8. Bring the dead back to their own country.
9. The bodies found in France should be buried there and the bodies found in Spain buried there. If relatives

could be traced, they should also be allowed to choose.

10. Bury them in the country they were found.
11. Take them to their own countries of origin.
12. Ask the families where they would like their relation to be buried.
13. Bury dead in countries they came from. If not possible, equal amounts in France and Spain.
14. Find out where the individual passengers came from - arrange for them to be buried in their home town.
15. France and Spain would find nationalities of the dead, then proceed with burial arrangements.
16. In their country of origin.

Appendix E: Written answers for Experiment 4

Condition: **'survivors'**

3 detecting subjects

1. Why bury the survivors, if they've survived. I don't think they'd let you.
2. You don't bury survivors.
3. Survivors are not buried.

12 non-detecting subjects

1. Where the aircraft crashed unless they have left specific instructions to the contrary.
2. Wherever their families wish.
3. Homeland if possible/where relatives wish.
4. At home i.e. as per wishes of deceased - or bereaved.
5. Depends on how remote the area is. But in their home country if possible.
6. Where their relatives want them to be buried.
7. Hometown.
8. Where their relatives want.
9. Where their family wants.
10. Wherever their family/friends feel they should be buried, or if the person had expressed his/her own preference.
11. In their home town or where they wish if stated in their will.
12. In home country if possible.

Condition: **'injured'**

10 detecting subjects

1. Where their relatives or friends request.
2. Injured people should not be buried.
3. In their native country.
4. It would be better to wait till they're dead then bury them where next of kin advise.
5. The injured should not be buried.

6. Injured dont get buried - only dead people.
7. If they have concerned families/friends or have expressed a wish for a particular funeral then, in line with their own wishes or those of family/friends. If not then at nearest convenient place.
8. ?
9. Are they dead yet? Changing word to 'dead'. Relatives wishes should be respected, if possible, and remains returned to them. If it is not possible, in nearest public cemetary.
10. Where their relatives want them to be.

<u>15 non-detecting subjects</u>

1. Where the relatives want.
2. At home.
3. At home.
4. In their home town.
5. Where the relatives wish, or in their place of birth.
6. Doesn't matter.
7. In their home town, if this has been their wish.
8. Buried where their families feel is appropriate.
9. If possible in their home town - however this may not be practical! -esp. if lost at sea, etc.
10. Cemetary?
11. Wherever their next of kin or immediate family wish them to be buried (assuming cremation isn't prefered).
12. Where their family decides.
13. Where they or relatives requested.
14. According to the wishes of close family.
15. This is a matter for the next of kin to decide or if it has been previously specified by the deceased, it should be their decision.

Condition: **'maimed'**

<u>15 detecting subjects</u>

1.
2. Curious question.
3. In their native country.

4. At their own homes.
5. Do you mean this or should it be deceased? Depends on wishes of deceased's relatives.
6. Wherever their relatives or their will says.
7. I think only the dead should be buried!
8. In home town.
9. At the location.
10. Where the relatives wish.
11. The maimed are not usually buried.
12. Wherever convenient in line with relatives wishes.
13. Wherever the family want.
14. Nowhere since they are still alive.
15. In their home town or a previously stated place by them.

10 non-detecting subjects

1. Choice of themselves (will) or their nearest relatives.
2. If prior arrangements made by deceased then abide by these, otherwise it is the wish of the family.
3. Wherever the closest relative wishes.
4. Wherever the relatives wish.
5. Where their relatives wish.
6. At their home towns.
7. In their local graveyard.
8. In the place (i.e. town) that they felt happiest in.
9. Where the relatives wish.
10. Wherever the next of kin wish.

Condition: **'wounded'**

9 detecting subjects

1. Wounded should be taken to hospital.
2. Wherever they would have wished to be buried.
3. Where the relatives wish.
4. i.e. dead?!? Decision of immediate family.
5. Depending on left instructions or families wishes.
6. I would hope they would firstly be taken to a hospital to be 'unwounded'. If dead then where the relatives or a will stipulates.

7. Where their families desire.
8. They should be taken to hospital if wounded.
9. Where the relatives wish.

<u>11 non-detecting subjects</u>

1. Home if possible.
2. As near their homes as possible (or where they wished).
3. Their home town.
4. Depends on relatives. If none, near to site.
5. If relatives are able to identify the body they should decide.
6. Wherever the relatives want.
7. Where the relatives request / will requests. Will should have priority.
8. In town where their family stays.
9. Wherever they or their surviving partner or relatives want them to be buried.
10. In the country they were born in.
11. Local cemetary.

Appendix F: Written answers for Experiment 5

condition: **'surviving passengers'**

<u>10 detecting subjects</u>

1. Where they had asked to be buried if they had mentioned it some time before.
2. Stupid question. 'Surviving passengers' are alive and will only be buried when they die of other causes later in life!!
3. Wherever family wants or if no family wherever authorities pay for.
4. The surviving passengers are not dead so should not be buried.
5. Surviving passengers are still alive.
6. Survivors should be allowed to live! <u>or</u> if dead dependent on relatives/persons involved wishes.
7. It would be more sensible only to bury the dead ones!
8. Ha Ha.
9. Surviving passengers <u>don't</u> get buried.
10. !

<u>5 non-detecting subjects</u>

1. Wherever relatives/deceased wish.
2. Wherever they or their relatives would like.
3. Wherever their family decide.
4. Where next of kin desire.
5. If possible in their home town/village.

condition: **'surviving injured'**

<u>12 detecting subjects</u>

1. Wherever they choose or wherever their family chooses.
2. Surviving injured - buried. I like it!
3. The surviving injured should not be buried, as this would constitute a serious breach of the criminal law of most countries.

4. Its not a wise move to bury the survivors.
5. Question seems odd.
6. /
7. Hopefully nowhere until they are dead.
8. The surviving injured are alive and therefore should not be buried at all. Those who don't surviv e should be buried according to their relatives wishes.
9. They are still alive so may object.
10. ???
11. Not sure of I would like to be injured and then buried - however assuming death occurs following crash & injury, surviving relatives should decide.
12. Why should injured survivors be buried?

<u>3 non-detecting subjects</u>

1. This will depend on any wishes of the deceased and the relatives.
2. Where their relatives want them to be buried.
3. Wherever they or their relatives want them to be buried.

condition: '**surviving dead**'

<u>13 detecting subjects</u>

1. Surviving dead? Wherever the individuals relatives wish them to be buried, if they are dead. (Do you mean that their bodies were found intact?)
2.
3. ?
4. Dead people rarely survive an aircrash.
5. ?
6. Can't bury the surviving dead. They would surely protest.
7. ?
8.
9. Does not make sense.
10. Surviving dead? Where their relatives wish them to be buried or where they wished to be buried, if they made this clear in their will.

11. You can't bury surviving dead?!
12. Silly question! (No not silly, trick more appropriate.)
13. According to the wishes of the surviving relatives.

<u>2 non-detecting subjects</u>

1. I don't understand the importance of the problem.
2. In the place where originally they came from.

Appendix G: Written answers for Experiment 6

Condition: **Question-form A: aircrash**

See Appendix E, Experiment 4, 'survivors' condition.

Condition: **Question form A: bicycle accident**

12 detecting subjects

1.
2. Survivors aren't buried. Ha!
3. ----
4. Hopefully a bicycle accident would not entail any deaths but if so in a cemetary as normal.
5. At their home - when they die.
6. Don't understand...
7. ?
8. eh?
9. Where they will receive the kind of treatment that will help them lead a normal life again.
10.
11. A ridiculous question, presumably the survivors won't require burial.
12. They shouldn't.

3 non-detecting subjects

1. Wherever their family want them to be.
2. They should be buried in their home towns.
3. In a cemetary as usual.

Condition: **Question-form B: aircrash**

4 detecting subjects

1. Ha! Ha!.
2. Where the relatives of the dead choose.
3. You don't bury survivors!
4. The survivors don't need to be buried!

11 non-detecting subjects

1. To me it is not a matter of any great consequence - wherever the crash took place; unless the deceased had left precise instructions to cover this rather remote eventuality.
2. Wherever their next of kin would like them to be buried.
3. Given that there is identifiable bodies - where their families wish.
4. Where next of kin would like them to be buried.
5. The choice should lie with relatives, otherwise in the home town of each person if no relations are known.
6. The wishes of the family and those of the individual should be followed.
7. Where they or their families want them buried.
8. Wherever next of kin wishes.
9. At scene.
10. Wherever the relatives wish it.
11. In their hometowns.

Condition: **Question-form B: bicycle accident**

11 detecting subjects

1. You don't bury the survivors.
2. In a graveyard.
3. Where they belong.
4. Does this question refer to the fact that bicycle accident is always the fault of motorists - if so - agreement that bicycle accident shows more negligence by motorist!
5. Trick question, I fancy! But what the heck just bury them anyway.
6. Don't bury survivors.
7. Don't bury survivors.
8. You shouldn't - they're alive.
9. Graveyard. /Just caught it!
10. What?! You shouldn't!.
11. If I was a survivor I wouldn't want to be buried!

4 non-detecting subjects

1. This would depend on the deceased's wishes and his/her family's wishes.
2. Where their family wants.
3. Where they would have wished to have been buried.
4. In a grave/ or cremate them.

Condition: **Question-form C: aircrash**

5 detecting subjects

1. Survivors are not dead and so can't be buried.
2. Wherever you bury live people (Ha ha).
3. They would be alive!
4. What!!
5. The dead should be buried where their relatives wish them to be buried. If there are no relatives then the dead **should** be buried near where the crash happened (i.e. in a graveyard) or buried in their home town.

10 non-detecting subjects

1. Where their family wants.
2. Decision should be left to the bereaved relatives.
3. Home town.
4. In their own hometown.
5. Where ever the survivors family wishes (i.e. at home town or crash site, or as specified in a last will and testament.)
6. Where they would have wanted to have been.
7. Wherever their families/relatives wish.
8. In the area their family come from or the place they have requested to be buried (if they have made this known).
9. Where their relatives would want them to be buried.
10. Where the next of kin would wish.

Condition: **Question-form C: bicycle accident**

<u>11 detecting subjects</u>

1. Shouldn't be buried at all.
2. Survivors wouldn't be buried.
3. Where they have wished to be buried.
4. Medically unethical to bury survivors, I think.
5. Burying survivors is a bad idea.
6. Survivors do not need to be buried.
7. Uh!!!! I think that <u>survivors</u> would object.
8. I don't fully understand this question!
9. After they eventually die (preferrably of natural causes at a ripe old age) in their local cemetary!?!
10. ?
11. Survivors aren't buried.

<u>4 non-detecting subjects</u>

1. If they have made a will, or whatever: in the grave of their choice. If not, where the next of kin decides - a family grave for example.
2. Whatever their family or they (if they specified) want.
3. Whatever the relatives wish.
4. Wherever they want.

Condition: **Question-form D: aircrash**

<u>8 detecting subjects</u>

1. Their own local cemetary/wherever you want to.
2. What?
3. Where the family or friends would have wanted it. (Hang on is this a trick question?)
4. Where their relatives choose.
5. Why would you want to bury the survivors?
6. Very funny! I think that they might object to being buried!
7. Where they <u>eventually</u> died!
8. Where their families want them to be buried if they are dead - survivors normally are alive and therefore shouldn't be buried.

<u>7 non-detecting subjects</u>

1. Whatever is in the will, or according to the closest relation (next of kin).
2. Where their families/next of kin wish: whatever helps them to come to terms with the crash. An attempt to find bodies should be made if reasonable and if it doesn't put others at risk.
3. Country of origin.
4. Where the relatives wish them to be buried.
5. Where the family involved wishes.
6. In their home town or where they have requested in a will e.g. a special graveyard.
7. Where the relatives prefer.

Condition: **Question-form D: bicycle accident**

<u>14 detecting subjects</u>

1. What!
2. Strange to bury survivors - but seeing as they are alive it should be their choice.
3. In a cemetary with everyone else.
4. What!!
5. ?
6. In a jam sandwich.
7. Nowhere if still alive.
8. Shouldn't bury them since they're still alive.
9. You shouldn't bury survivors - they are still alive!
10. Trick question!
11. -murder!
12. Is this sentence not a contradiction!
13.
14. According to wishes of relatives.

<u>1 non-detecting subject</u>

1. Where they wish - with everyone else.

Appendix H: Written solutions for Experiment 7

Condition: '**survivors**'

7 detecting subjects

1. Find out the nationalities of those dead and send them to their country of origin (Did that say where to bury the survivors!).
2. The survivors are alive and don't need buried.
3. If the question hinges upon the crash happening on the border, I feel that nationality and where the person lives to be more important than where the person dies. Is this some kind of trick question by the way?...burying survivors!
4. You don't bury survivors.
5. You don't bury survivors!
6. Let the families decide. (Did they actually die?)
7. I would ask the families where they wanted their dead relations buried.

7 non-detecting subjects

1. Either take the people to be buried home to their native land or bury them where they landed - even on the border.
2. Bury them all where they came from in the first place.
3. Consult the relatives & try to bury them where their next of kin wished.
4. Ask the relatives where they would want the person buried, do it accordingly.
5. Find out the nationality of the cyclists and bury them in their own countries.
6. Bury each of the dead in their own home towns & place a monument on the border where the crash took place.
7. Send the people hom e to their families who will deal with the funeral in their home country.

Condition: '**wounded**'

6 detecting subjects

1. Home countries (& you bury dead not wounded!!!)
2. "Bury the wounded?" If they were dead, bury them in the home countries.
3. Fly the bodies home to the respective countries of domicile.
4.
5. I would ask the relatives where they wanted the person to be buried. If they had no preference, I would bury half in Italy and half in Switzerland.
6. Bury them in a line on the border with the authorities being responsible for half of each.

10 non-detecting subjects

1. They should bury them in the country which organised the race, or in which it started.
2. Send the bodies back home to their home countries and bury them there.
3. Bury on border.
4. Forget political boundaries. After checking their wills for other means of burial, bury them anywhere.
5. Put memorials up in the country in which the race started.
6. Where the cyclists actually died is of no relevance to their burial place - they should each be flown back home and buried according to their families wishes.
7. Find out where the families of the cyclists want them to be buried, & act accordingly.
8. Build a graveyard over the border - bury the dead on the border of the two countries.
9. They should be buried according to their nationality.
10. Firstly it would depend on their families. Ask the families where they would prefer them to be buried and erect some kind of monument on the spot of the accident in their memory.

Condition: **'surviving dead'**

11 detecting subjects

1. Contact the families of the dead and let them decide where they want their relation buried.
2. Don't bury survivors, if there were any deaths then they would be sent home for burial.
3. Phone and contact relatives. Ask them where they would like them to be buried. If no relatives then bury them nearest their home (i.e. in the country they were born in).
4. Don't bury the surviving dead.
5.
6. They should discover their nationality (either Italian or Swiss) and bury them in their respective home countries or if they are of another nationality, they should consult the next of kin on where to bury the deceased.
7. There wouldn't be any surviving dead.
8. You don't bury survivors.
9. Let their families decide - they would probably want to make the decision.
10. Surviving dead!
11. I don't understand what it means by 'the authorities couldn't make up their minds where to bury the surviving dead' - I can't give a solution because I can't make sense of the problem.

<u>6 non-detecting subjects</u>

1. I would find out where each individual came from, contact their respective relations and let them decide on the place of burial.
2. Bury them in country of birth.
3. Bury them in countries that they lived in/had come from.
4. Return their bodies to their own home countries for burial.
5. Send the bodies home to the parents for them to decide what to do with them.
6. The relatives of the cyclists should be contacted and be asked if they wanted the bodies to be buried at their homes, or if not which country they would prefer their relative to be buried in.

Appendix I: Written answers for Experiment 8

Level A: disconfirming qualification

Condition: '**unhurt**'

15 detecting subjects

1. In home country.
2. Why should they be buried if they are mostly unhurt.
3. They should go to hospital.
4.
5. Survivors should not be buried.
6. Survivors are not normally buried.
7. ?
8. They should not be buried at all if they survived.
9. If they were survivors, they shouldn't be buried!
10. Where they wanted to be, or wher their family want them to be (assuming they have one). (I take it you mean those that were killed?)
11. Wherever the family (if any) want them to be buried - the dead that is. I hope you're not planning on burying the survivors.
12. I wouldn't bury survivors!
13. What!
14. I wouldn't bury survivors.
15. If they were mostly unhurt then they wouldn't need to be buried! i.e. they wouldn't be dead!

Level B: low relevance qualification

Condition: '**survivors**'

14 detecting subjects

1. Wherever they wish
2. Survivors can't be buried.
3. Survivors? Wait till they die first!
4. Sounds nasty.

5. Wherever their families choose.
6. Where their families would want them to be buried.
7. Nearest place to where crash occured.
8. Where they would have been buried had they died at home.
9. In the place of their relatives choice.
10. Perhaps you should wait till they die before asking this question.
11. Nowhere.
12. Survivors are not buried (I hope).
13. The survivors should not be buried:- any victims should be buried in their home town/country.
14. Wherever the crash took place.

5 non-detecting subjects

1. Home.
2. Where they crashed.
3. Their home country.
4. In place of their own choice or that of their relatives.
5. Their home town.

Condition: **'many survivors'**

12 detecting subjects

1. Wherever their relatives (or they themselves) wish, if possible.
2. Either near the site of crash or in their home locality - decision to be made by next of kin.
3. In their home towns - if possible.
4. No point in burying survivors.
5. Dont bury survivors.
6. Depends on their relatives requests.
7. I didn't think we buried survivors!
8. In their home towns but why do you want to bury the survivors?
9. Bury survivors??
10.
11. In their home town.

12. Survivors don't get buried.

<u>5 non-detecting subjects</u>

1. Find out their wishes if possible.
2. If they have expressed no wishes previously themselves on the matter then wherever their families wish them to be.
3. If bodies identifiable: at home cemetary. If not, at one nearest or on site of crash.
4. Where the plane crashed.
5. a. Per instructions in will. b. Country of domicile (home). c. If in doubt according to wishes of closest relations.

Condition: '...**last week**'

<u>12 detecting subjects</u>

1. Survivors shouldn't be buried, but any dead should be buried where their relatives want.
2. Can't bury survivors! Those who died should be buried where their families wish.
3. The site of the crash unless desired otherwise by relatives.
4. You wouldn't bury the survivors!
5. Where the victims wanted to be buried.
6. Where the families see fit. Not mass burial.
7. Who? The survivors? Those who actually died (if any) should be buried where their family wish them to be buried.
8. I take it you don't mean the survivors. The dead should be flown home.
9. In their home context, with relatives.
10. They wouldn't be buried if they were survivors!
11. Wherever their wills or families specify.
12. Shouldn't be buried waste of land. Should be cremated.

<u>3 non-detecting subjects</u>

1. Where they specicfied they wanted buried.
2. Where the next of kin wants/ or where the deceased's

will states.

3. In their home towns, with their families.

Condition: **'gravediggers'**

12 detecting subjects

1. When they die they should be buried where they wish to be buried.
2. Aren't they alive.
3. This is irrelevant. Why should occupation matter?!!! In a cemetary I suppose, with full honours laid on by the airline company & compensation.
4. They're not dead: where they wanted.
5. Aren't survivors usually alive?
6.
7. Survivors should not be buried.
8. Sorry - don't understand this question.
9. In a cemetary.
10. Don't understand question
11. If they survived they wouldn't need to be buried.
12. If they were survivors of the aircrash, they wouldn't need to be buried!

3 non-detecting subjects

1. ?
2. In the ground.
3. Where they worked.

Level C: High relevance qualification

Condition: **'no fixed abode'**

7 detecting subjects

1. Survivors would not be buried.
2. Wherever they die at some future time.
3. Burying survivors?
4. There is no need to bury survivors!!

5. In a burial ground near where they were found. It doesn't matter where. (Unless someone - oh. Wait a minute - <u>you don't bury survivors</u> unless you want a manslaughter case on your hands.)
6. They're not dead.
7. Shouldn't bury survivors!

<u>8 non-detecting subjects</u>

1. Nearest place.
2. The country where they lived for most of their lives, or where their relatives are.
3. They should be cremated.
4. In a graveyard! -everyone is entitled to respect.
5. Buried or burned. Memorialised at the place of the crash.
6. State cemetries -- perhaps even set a new one up near disaster site.
7. Where their nearest families wished them to be buried.
8. In their country of origin - i.e. if British, buried in UK.

Condition: **'European'**

<u>7 detecting subjects</u>

1. Survivors don't get buried.
2.
3. ?
4. Don't need to be buried.
5. Wherever they would have been buried had they died naturally.
6. I didn't think that you buried the survivors?
7. In their home country.

<u>8 non-detecting subjects</u>

1. Where they came from originally.
2. In their own towns with memorial on site of crash.
3. Their home town if they later died in hospital.
4. In their home countries.
5. At home if possible.

6. In their home town if possible. If not at a special burial ground near the site.
7. Where their families desire, probably back home.
8. Their own home town.

Condition: '**circus performers**'

<u>9 detecting subjects</u>

1. Survivors would not be buried.
2. Preferably nowhere!
3. They shouldn't be buried if they survived.
4. The circus performers are not dead, so they shouldn't be buried.
5. They don't need to be buried if they are survivors.
6. Nowhere they are not dead.
7. Original hometowns if possible.
8. Where they died.
9. The survivors shouldn't be buried! What difference would there being circus performers or not have to do with where they should be buried anyway?

<u>7 non-detecting subjects</u>

1. Where their friends and family would like them to be buried.
2. In the country that the plane crashed in, or in the country of origin of the plane.
3. Anywhere.
4. In the country where the plane crashed.
5. Wherever they had closest ties, otherwise as close as possible to crash site all together.
6. If leave a will - where they choose. If have a home of own, there. If parental home only, there. If none, within jurisdiction where they died.
7. Depends on what their will says, or family say.

Appendix J: Instructions for the questionnaire-based studies

Investigation into Social Attitudes

Instructions:

Thank you for agreeing to take part in this investigation. It simply involves answering a series of 10 short questions on the adjoining sheet.

These questions form part of an investigation into people's views on various social and related issues. Some of the questions may seem very general or even vague, but please answer as best you can without seeking further information. It is your own opinion that we are interested in.

Please work through the questions reasonably quickly: in-depth or complex answers would go beyond what we are really looking for, so be as brief (and as un-grammatical) as you like. There should be enough space under each question for your answer.

No record of your name or any personal details is required.

Thank you for taking part!

List of References

Altmann, G.T.M. and Steedman, M.J. (1988), 'Interaction with context during human sentence processing', Cognition, **30**, 191-238.

Anderson, A., Garrod, S.C., and Sanford, A.J. (1983), 'The accessibility of pronominal antecedents as a function of episode shifts in narrative text', Quarterly Journal of Experimental Psychology. **15**.

Anderson, R.C. and Pichert, J.W. (1978), 'Recall of previously unrecallable information following a shift in perspective', Journal of Verbal Learning and Verbal Behaviour, **17**, 1-12.

Bartlett, F.C. (1932), Remembering, Cambridge University Press.

Barton, S.B., Moxey, L.M., Paterson, K. and Sanford, A.J. (in preparation), 'Understanding incompleteness in discourse cohesion: recent advances in anomaly detection and connectionist modelling'

Beeman, M. and Gernsbacher, M.A. (1990), 'Structure building and coherence inferencing during comprehension', unpublished manuscript.

Bell, M. and Grant, G. (1991), 'Anomaly detection as a function of scenario-type: an extension to the 'survivors problem'', unpublished undergraduate manuscript, Department of Psychology, University of Glasgow.

Bierwish, M. (1970), 'Semantics', in J.Lyons (ed.), New Horizons in Linguistics, Penguin, Baltimore.

Bierwich, M. (1983), 'How on-line is language processing?', in G.B.Flores d'Arcais and R.J.Jarvella (eds.), The process of language understanding, Wiley.

Blutner, R and Sumner, R (1988), 'Sentence processing and lexical access: The influence of the focus-identifying task', Journal of Memory and Language, **27**, 359-367.

Bobrow, S.A. and Bower, G.H. (1969), 'Comprehension and recall of sentences', Journal of Experimental Psychology, **80**, 455-461.

Bobrow, R.J. and Webber, B.L. (1980), 'Knowledge representation for syntactic/semantic processing', In Proceedings of the First Annual National Conference on Artificial Intelligence, 316-323., Stanford, California.

Bower, G.H., Black, J.B. and Turner, T.J. (1979), 'Scripts in memory for text', Cognitive Psychology, **3**, 193-209.

Bransford, J.D. and Johnson, M.K. (1973), 'Consideration of some problems of comprehension', In Chase.N. (ed.) Visual Information Processing, New York; Academic Press.

Bransford, J.D. and McCarrell, N.S. (1974), 'A sketch of a cognitive approach to comprehension: some thoughts about understanding what it means to comprehend', In W.B.Weimer and D.S.Palermo (eds.), Cognition and the symbolic process, Lawrence Erlbaum: Hillsdale.

Carnap, R. (1956), 'Meaning and necessity: a study in semantics and modal logic', Chicago Press.

Chase, W.G. and Ericsson, K.A. (1981), 'Skilled memory', In J.R.Anderson (ed.), Cognitive Skills and their Acquisition, Erlbaum, N.J.

'Colemanballs 2' (1984), Private Eye Productions Ltd.

'Colemanballs 3' (1986), Private Eye Productions Ltd.

'Colemanballs 4' (1988), Private Eye Productions Ltd.

'Colemanballs 5' (1990), Private Eye Productions Ltd.

Corbett, A.T. and Dosher, B.A. (1978), 'Instrument inferences

in sentence encoding', Journal of Verbal Learning and Verbal Behaviour, **17**, 479-491.

Cotter, C.A. (1984), 'Inferring indirect objects in sentences: some implications for the semantics of verbs', Language and Speech, **27**, Part 1, 25-45.

Crain, S. and Steedman, M.J. (1985), 'On not being led up the garden path: The use of context by the psychological parser', In D.Dowty, L.Karttunen, and A.Zwicky (eds.), Natural language parsing: Psychological, computational, and theoretical perspectives, Cambridge: Cambridge University Press.

Cutler, A. (1982), 'Slips of the tongue and language production', Mouton.

Dooling, D.J. and Lachman, R. (1971), 'Effects of comprehension on retention of prose', Journal of Experimental Psychology, **88**, (no.2), 216-222.

Dooling, D.J. and Mullett, R. (1973), 'Locus of thematic effects in retention of prose', Journal of Experimental Psychology, **97**, 404-6.

Ehrlich, M-F. and Loridant, C. (1990), 'Metacognitive control in the resolution of anaphora in skilled and less-skilled comprehenders', Paper presented at the Fourth Conference of the European Society for Cognitive Psychology, Como, Italy.

Erickson, T.D. and Mattson, M.E. (1981), 'From words to meaning: a semantic illusion', Journal of Verbal Learning and Verbal Behaviour, **20**, 540-551.

Fillmore, C.J. (1968), 'The case for case', in E.Bach and R.T.Harm (eds.), Universals of Linguistic Theory, New York, Holt, Rinehart and Winston, 1-90.

Fodor, J.A. (1975), 'The Language of Thought', Hassocks, Sussex: Harvester Press

Fodor, J.A. and Pylyshyn, Z.W. (1988), 'Connectionism and cognitive architecture: A critical analysis', Cognition, **28**, 3-71.

Frege, G. (1892), 'On sense and reference', Zeitschrift fur Philosophie and philosophische Kritik, **100**, 25-50. Translation in P.Geach and M.Black (eds.), 'Philosophical Writings of Gottlob Frege', (1977).

Fromkin, V.A. (1980), 'Errors in linguistic performance: slips of the tongue, ear, pen and hand', Academic Press.

Garnham, A. (1982), 'Testing psychological theories about inference making', Memory and Cognition, **10**, 341-349.

Garrod, S.C. and Sanford, A.J. (1977), 'Interpreting anaphoric relations: the integration of semantic information while reading', Journal of Verbal Learning and Verbal Behaviour, **16**, 77-90.

Garrod, S.C. and Sanford, A.J. (1978), 'Anaphora: a problem in text comprehension', in R.N.Campbell and P.T.Smith (eds.), Recent Advances in the Psychology of Language, New York, Plenum Press.

Garrod, S.C. and Sanford, A.J. (1983), 'Topic dependent effects in language processing', in G.B.Flores d'Arcais and R.J.Jarvella (eds.), The Process of Language Understanding, Wiley.

Garrod, S.C. and Sanford, A.J. (1985), 'On the real-time characteristics of interpretation during reading', Language and Cognitive Processes, **1**, 43-61.

Garrod, S.C. and Sanford, A.J. (1988), 'Discourse models as interfaces between language and the spatial world', Journal of Semantics, **6**, 147-160.

Gernsbacher, M.A., Varner, K.R. and Faust, M.E. (1990), 'Investigating differences in general comprehension skill', Journal of Experimental Psychology: Learning, Memory and Cognition.

Grice, P. (1975), 'Logic and conversation', in P.Cole and J.L.Morgan (eds.), Studies in Syntax, vol.3: Speech acts. New York: Academic Press.

Haddock, N.J. (1989), 'Computational models of incremental semantic interpretation', Language and Cognitive Processes, **4**, 337-368.

Halliday, M.A.K. (1967), 'Notes on transitivity and theme in English', Journal of Linguistics, **3**, 199-244.

Haviland, S.E. and Clark, H.H. (1974), 'What's new? Acquiring new information as a process in comprehension', Journal of Verbal Learning and Verbal Behaviour, **13**, 512-521.

Hilton, D.J. (1988), 'Logic and Causal Attribution', in D.J.Hilton (ed.), Contemporary Science and Natural Explanation: Commonsense Conceptions of Causality, Harvester, Sussex.

Johnson, M.K., Bransford, J.D. and Solomon, S. (1973), 'Memory for tacit implicactions of sentences', Journal of Experimental Psychology, **98**, 203-205.

Johnson-Laird, P.N. (1983), 'Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness', Cambridge University Press.

Katz, J.J. and Fodor, J.A. (1963), 'The structure of a semantic theory', Language, **39**, 170-210.

Kintsch, W. (1974), 'The Representation of Meaning in Memory', Potomac, Md., Erlbaum.

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

Kintsch, W. and Mross, E. (1985), 'Context effects in word identification', Journal of Memory and Language, **24**, 336-349.

Kintsch, W. and van Dijk, T.A. (1978), 'Towards a model of text comprehension and production', Psychological Review, **85**, 363-394.

Lakoff, G. (1972), 'Hedges: a study in meaning criteria and the logic of fuzzy concepts', in Papers from the Eighth Regional Meeting of the Chicago Linguistics Society, Chicago: Chicago Linguistics Society.

Lakoff, G. and Johnson, M. (1980), 'Metaphors We Live By', Chicago Press.

Marks, L.E. and Miller, G.A. (1964), 'The role of semantic and syntactic constraints in the memorisation of English sentences', Journal of Verbal Learning and Verbal Behaviour, **3**, 1-5.

McClelland, J.L., St.John, M.F. and Taraban, R. (1989), 'Sentence comprehension: a parallel distributed processing approach', Language and Cognitive Processes, **4**, 287-335.

McKoon, G. and Ratcliff, R. (1981), 'The comprehension processes and memory structures involved in instrument inference', Journal of Verbal Learning and Verbal Behaviour, **20**, 671-682.

McKoon, G. and Ratcliff, R. (1986), 'Inferences about predictable events', Journal of Experimental Psychology: Learning, Memory and Cognition, **12**, 82-91.

McKoon, G. and Ratcliff, R. (1990), 'Textual inferences: models and measures', in D.A.Balota, G.B.Flores d'Arcais and K.Rayner (eds.), Comprehension Processes in Reading, Erlbaum, Hillsdale, N.J.

Mellish, C.S. (1981), 'Coping with uncertainty: Noun phrase interpretation and early semantic analysis', Ph.D Thesis, University of Edinburgh.

Mellish, C.S. (1982), 'Incremental evaluation: An approach to the semantic interpretation of noun phrases', Cognitive Science Research Paper No.1, Cognitive Studies Programme, University of Sussex, Brighton.

Mellish, C.S. (1983), 'Incremental semantic interpretation', In K.Sparck-Jones and Y.A.Wilks (eds.), Automatic natural language parsing, Chichester/New York: Ellis Horwood/Wiley.

Mellish, C.S. (1985), 'Computer interpretation of natural language descriptions', Chichester: Ellis Horwood.

Miller, G.A. (1956), 'The magical number seven, plus or minus two: some limits on our capacity for processing information', Psychological Review, **63**, 81-87.

Miller, G.A. (1979), 'Images and models, similies and metaphors', in A.Ortony (ed.), Metaphor and Thought, Cambridge University Press.

Minsky, M. (1975), 'A framework for representing knowledge', in P.H.Winston (ed.), The Psychology of Computer Vision, McGraw-Hill, New York.

Minsky, M. (1988), 'The Society of Mind', Heineman.

Montague, R. (1970), 'The proper treatment of quantification in ordinary English', in J.Hintikka, J.Moravcsik, and P.Suppes (eds.), Approaches to Natural Language: Proceedings of the 1970 Stanford workshop on Grammar and Semantics, Reidel: Dordrecht.

Moxey, L.M. and Sanford, A.J. (1987), 'Quantifiers and focus', Journal of Semantics, **5**, 189-206.

Moxey, L.M. and Sanford, A.J. (1991), 'Context effects and the communicative function of quantifiers', in N.Schwartz and S.Sudman (eds.), Context effects in social and psychological research, Springer-Verlag, New York.

Moxey, L.M., Sanford, A.J. and Barton, S.B. (1990), 'Control of attentional focus by quantifiers', in K.J.Gilhooly, M.T.G.Keane, R.H.Logie and G.Erdos, 'Lines of Thinking: Reflections on the Psychology of Thought: Volume 1', Wiley.

Murray, H.G. (1970), 'Stimulus intensity and reaction time: evaluation of a decision-theory model', Journal of Experimental Psychology, **84**, 383-391.

Paris, S.G. and Landauer, T.K. (1976), 'The role of inferences in children's comprehension and memory for sentences', Cognitive Psychology, **8**, 217-227.

Paterson, K. (1991), 'A connectionist model of the 'survivors problem'', unpublished M.Sc. thesis, University of Glasgow.

Pereira, F.C.N. and Shieber, S.M. (1987), 'Prolog and natural language analysis', CSLI, Stanford.

Potts, G.R., Keenan, J.M. and Golding, J.M. (1988), 'Assessing the occurrence of elaborative inferences: lexical decision versus naming', Journal of Memory and Language, **27**, 399-415.

Rayner, K. (1978), 'Eye movements in reading and information processing', Psychological Bulletin, **85**, 618-660.

Reason, J. (1984), 'Absent-mindedness and cognitive control', in J.Harris and P.Morris (eds.), 'Everyday memory, actions and absent-mindedness', London: Academic Press.

Rosch, E. (1973), 'On the internal structure of perceptual and semantic categories', in T.E.Moore (ed.), Cognitive Development and the Acquisition of Language, New York, Academic Press.

Rubin, D.C. (1977), 'Very long term memory for prose and verse', Journal of Verbal Learning and Verbal Behaviour, 16, 611-621.

Rumelhart, D.E. (1975), 'Notes on a schema for stories', in D.G.Bobrow and A.Collins (eds.), Representing and Understanding: Studies in Cognitive Science, Academic Press, New York.

Rumelhart, D.E. and Ortony, A. (1977), 'The representation of knowledge in memory', in R.C.Anderson, R.J.Spiro, and W.E.Montague (eds.), Schooling and the Acquisition of Knowledge, Hillsdale, N.J., Erlbaum.

Rumelhart, D.E., Smolensky, P., McClelland, J.L. and Hinton, G.E. (1987), 'Schemata and sequential thought processes in PDP models', in J.L.McClelland, D.E.Rumelhart, and the PDP research group (eds.), 'Parallel Distrbuted Processing: Explorations in the microstructure of cognition: Volume 2, Psychological and Biological models', 7-57, Cambridge MA: MIT Press.

Ryan, J. (1969), Temporal grouping, rehearsal and short term memory', Quarterly Journal of Experimental Psychology, **21**, 148-55.

Sachs, J.S. (1967), 'Recognition memory for syntactic and semantic aspects of connected discourse', Perception and Psychophysics, Volume **2**, 437-442.

Sanford, A.J. (1972), 'Loudness and simple reaction time', Sound, **6**, 92-96.

Sanford, A.J. (1990a), 'On the nature of text driven inference', in D.A.Balota, G.B.Flores d'Arcais and K.Rayner (eds.), Comprehension Processes in Reading, Erlbaum, Hillsdale, N.J.

Sanford, A.J. (1990b), 'Component processes of reference resolution in discourse', in N.Sharkey (ed.), Models of

Cognition: A review of Cognitive Science, Volume 1, 113-140, Norwood, NJ: Ablex.

Sanford, A.J. and Garrod, S.C. (1980), 'Memory and attention in text comprehension: the problem of reference', in R.S.Nickerson (ed), Attention and Performance VIII, Hillsdale, N.J.: Lawrence Erlbaum.

Sanford, A.J. and Garrod, S.C. (1981), 'Understanding Written Language: Explorations Beyond the Sentence', Wiley.

Sanford, A.J. and Garrod, S.C. (1989), 'What, when, and how?: Questions of immediacy in anaphoric reference resolution', Language and Cognitive Processes, **4**, 235-262

Sanford, A.J. and McGinley, M.T. (1990), 'In search of a richer model of written language comprehension: Three fragments of evidence', in O.Dahl and K.Fraurud (eds.), Papers from the Second Nordic Conference on Text Comprehension in Man and Machine.

Searle, J.R. (1969), 'Speech acts: an essay in the philosophy of language', Cambridge University Press.

Seidenberg, M.S., Tanenhaus, M.J., Leiman, J.M. and Bienkowski, M. (1982), 'Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing', Cognitive Psychology, **14**, 538-559.

Shank, R. and Abelson, R. (1977), 'Scripts, Plans, Goals and Understanding: an Enquiry into Human Knowledge Structures', Erlbaum, Hillsdale, N.J.

Singer, M. (1981), 'Verifying the assertions and implications of language', Journal of Verbal Learning and Verbal Behaviour, **20**, 46-60.

Singer, M. and Ferreira, F. (1983), 'Inferring consequences in story comprehension', Journal of Verbal Learning and

Verbal Behaviour, **22**, 437-448.

Sperber, D. and Wilson, D. (1986), 'Relevance, Communication and Cognition', Basil Blackwell: Oxford.

St.John, M.F. (1991), 'Hitting the right pitch: A meta-analysis of the effect of sentence context on lexical analysis', unpublished manuscript, Depertment of Cognitive Science, University of California.

St.John, M.F. and McClelland, J.L. (1990), 'Learning and applying contextual constraints in sentence comprehension', Artificial Intelligence.

Sutcliffe, J.P. (1957), 'A general method of analysis of frequency data for multiple classification designs', Psychological Bulletin, **54**, 134-137.

Thorndyke, P.W. (1976), 'The role of inferences in discourse comprehension', Journal of Verbal Learning and Verbal Behaviour, **15**, 437-446.

Tulving, E. (1962), 'Subjective organisation in free recall of 'unrelated' words', Psychological Review, **69**, 344-354.

Tulving, E. (1968), 'When recall is higher than recognition', Psychonomic Science, **10**, 53-4.

Walker, C.H. and Meyer, B.J.F. (1980), 'Integrating different types of information in text', Journal of Verbal Learning and Verbal Behaviour, **19**, 263-275.

Wason, P.C. and Reich, S.S. (1979), 'A verbal illusion', Quarterly Journal of Experimental Psychology, **31**, 591-597.

Weinreich, U. (1966), 'Explorations in semantic theory', In T.A.Sebeok (ed.), Current Trends in Linguistics, Volume 3. The Hague: Mouton.

Wilkins, A.J. (1971), 'Conjoint frequency, category size and categorisation time', Journal of Verbal Learning and Verbal Behaviour, **10**, 383-5.

Winer, B.J. (1971), 'Statistical Principles in Experimental Design', 2nd edition, McGraw-Hill.

Winograd, T. (1972), 'Understanding Natural Language', Edinburgh: Edinburgh University Press.