Towards a Psychological Basis for a Theory of Anaphora

William Marslen-Wilson & Lorraine K. Tyler
Max-Planck-Institut für Psycholinguistik
Nijmegen, The Netherlands

1. Introduction

Those of us who have studied videotapes of conversational interactions may have come across a class of gestures that can be called "abstract deictics". That is, the speaker makes a small and indeterminate gesture which, while evidently having a deictic quality to it, does not point to anything in particular in physical space. Rather, it indicates some nebulous region in some ill-defined conceptual space. It is along these lines that the word "towards" in the title of this talk should be interpreted. Our purpose here is not to try to provide a highly structured theory of anaphora, or of the mental processes involved in the use of such devices, but rather to indicate what we think such a theory should be about in the first place; that is, to indicate what kind of psychological object it is that is being studied.

We should also comment on the other terms in the title: "a psychological basis for a theory of anaphora". The location of the modifier "psychological" may appear somewhat presumptuous, since it implies that the only proper basis for a theory of anaphora is indeed psychological. While this claim need not hold for all aspects of the study of anaphora, we believe that it has to hold for any approach to anaphora which is intended to make claims about human mental processes and about the properties of the mental representations underlying these processes. The reason we stress this is because there seems to be considerable pressure at the moment to put certain branches of the study of language in the mind, or even in the brain - and to call the results "cognitive science". Unfortunately, this often seems to be done without any very clear conception of what it means to make interestingly testable statements about mental processes.

The point here is not that a scientific solution to problems in human cognition should not involve some confluence of experimental psychological, linguistic, and artificial intelligence approaches to these problems. It is, rather, that this confluence has to be more rigorously constrained than it generally seems to be. We believe, furthermore, that a particularly essential constraint must be basic psychological evidence about the dynamic properties of the mental
phenomena that one is trying to explain. If such constraints tend to be missing, this is at least partly due to the failure of psychologists to provide the appropriate experimental and theoretical basis for them. Our goal here, therefore, is to point out some of the currently available processing constraints that bear on the study of anaphora.

To do this, we have to start by embedding the study of anaphora in a wider context. Given our own particular psycholinguistic interests, the wider context we will define here is the general process of spoken language understanding; that is, the ways in which human listeners perceive and interpret natural spoken language. It is, above all, important to define the characteristics of the processing environment in which one can suppose that anaphoric interpretation processes are operating. We begin, then, by presenting a view of the speech understanding process.

2. On-line Speech Understanding

The view of speech understanding we will present here is one that we have labelled as an "on-line interactive" approach. Both these terms carry a number of contrasts with other approaches. The term "on-line" reflects, first, our emphasis on the mental processes that take place, word-by-word, millisecond-by-millisecond, as the listener hears an utterance, and, second, the types of experimental tasks we use to investigate these processes. Typically, these are fast reaction-time tasks, where the subject responds, during the perception of an utterance, to some experimentally manipulated attribute of that utterance.

The reason for this methodological emphasis is that our goals are to determine what happens when during speech understanding; that is, when certain forms of analysis of the speech input can become available to the listener. The speech input is necessarily ordered in time; thus, at any chosen moment, one can determine exactly how much of the input the listener could have heard. Given a response tapping his internal representation of the input at some given moment, we can then hope to capture the properties of the forms of internal representation that develop as the input is being processed. In particular, we can determine to what extent the listener's on-going processing reflects just a "bottom-up" analysis of the signal, and to what extent internally generated "top-down" inputs are also involved in on-line speech processing.

The importance of using fast reaction-time tasks, in this connection, is, first, that this enables a close temporal relationship to be specified between the response and the portion of the input heard when the
response occurred. Second, fast responses prevent the subject from deliberating for too long about what he is doing. The types of mental processes that we assume underly speech understanding are highly automatised and obligatory; once the listener has had time to reflect upon the final products of these processes, then all sorts of "off-line" variations can start to be introduced. We want to avoid these as much as possible, and to concentrate instead on what Jerry Fodor once called the "brute cognitive reflexes" that constitute the heart of the speech understanding process.

The second contrast carried by the label "on-line interactive" is, of course, the assumption that speech processing is interactive in character. Technically, this means that the system allows for two-way communication between the different knowledge sources involved in the process. This is to be contrasted with the quite wide-spread assumption that these processing components are autonomous; namely, that the operations within each component cannot be affected by inputs further along in the processing sequence (e.g., Forster, 1979). In such a system, for example, operations within the word-recognition component could not be affected by contextual influences derived from the syntactic and semantic environment in which the word being recognised was occurring.

To legitimise the view we are presenting of speech understanding, and to make explicit just what we are claiming, it is necessary to examine a couple of experiments in detail. The first of these experiments depends critically on the role of word-recognition in speech understanding. In particular, it depends on the fact that a word in a natural utterance context can usually be recognised before all of it has been heard. Before, in fact, enough of the word could have been heard to allow a unique identification of the word on the basis of the sensory input available at the point of recognition. This property of word-recognition performance means that other sources of information must be being brought to bear upon the recognition process - to replace, as it were, the sensory information missing at the moment of recognition. The important point, for present purposes, is the nature of these sources of contextual constraint; that they include not only the structural syntactic environment in which the word is occurring, but also its discourse context.

This claim is based on the following sort of experiment. We used stimulus sentences which systematically varied the availability of different forms of analysis of these sentences, and looked to see whether the presence or absence of these forms of analysis affected performance in a task that was sensitive to word-recognition processes. The contrasts
we were making are illustrated by the following set of sentences:

1a) The church was broken into last night. Some thieves stole most of the *lead* off the roof.

1b) The power was located into great water. No buns puzzle some in the *lead* off the text.

1c) Into was power water the great located. Some the no puzzle buns in *lead* text the off.

The major variable here is the distinction between prose types. In (1a) - Normal Prose - not only is a syntactic analysis of the string possible, but also what we will call an interpretative analysis. By this we mean the interpretation of an utterance in its discourse context, and in terms of the listener's general knowledge of the world - in short, what is generally thought of as "understanding" [1]. In (1b) - Anomalous Prose - only a form of syntactic structuring of the input is possible, while in (1c) - Scrambled Prose - the subject hears an unstructured list of words.

The second important variable was the position of the target-word in the test-sequence. In the examples given above, the target-word ("lead") appears in seventh position in the second sentence. Over the 81 test-sequences, with their associated 81 target-words, the location of the target-word varied from the second to the tenth word-position in the second test-sentence. The purpose of this was as follows. We assume that the availability of an interpretative and syntactic analysis in Normal Prose can facilitate the recognition of the target word. The word-position variable enables us to track the availability of these sources of contextual constraint as they develop across the sentence. As such constraints become available - that is, as the appropriate form of representation of the input sequence is developed by the listener - then we should begin to find differences in performance in Normal Prose as opposed to Anomalous Prose, and in Anomalous Prose as opposed to Scrambled Prose.

The third main variable was the presence or absence of a lead-in sentence - that is, of the first sentence in each test-sequence. First, we will examine the results for the case where the lead-in sentence is present. The tasks we used were a variety of word-monitoring tasks, where the subjects were told in advance of each sequence what the target was, under various forms of description of the input. The subjects simply pressed a response-key as soon as the target-word in question had been identified (for further details, see Marslen-Wilson & Tyler, 1980).
We found, as expected, that responses were faster in Normal Prose than in Anomalous Prose, and faster in Anomalous Prose than in Scrambled Prose -- indicating that the interpretative and the syntactic constraints available in Normal Prose, and the syntactic constraints available in Anomalous Prose, were able to speed on-line word-recognition decisions. More important, however, was the distribution of these differences across the test-sequences. In Scrambled Prose, where no structural constraints can build up across the string, reaction-times stayed the same from the first to the tenth word-positions. In Normal Prose and Anomalous Prose, we found, first, that reaction-times decreased linearly across the sentence, and, second, that the advantage of Normal Prose over Anomalous Prose was just as great at the beginning of the sentence as it was ten words further along.

On the logic of this type of experiment, any difference between Normal and Anomalous Prose must be due to the availability, in Normal Prose, of a form of analysis of the speech input which is not available in Anomalous Prose. The patterning of the facilitation of Normal Prose across the utterance shows that this facilitation is just as great very early in the utterance as it is later on. Since the advantage in Normal Prose is assumed to derive from the interpretative analysis of the utterance, and since the first one or two words of an utterance cannot be sufficient by themselves to provide the same degree of facilitation as would be obtained later in the utterance, then the early advantage of Normal over Anomalous Prose must in large part have derived from the preceding context (the lead-in sentence). This has the important implication that the discourse context in which an utterance is occurring can affect the analysis of even the first one or two words of that utterance.

To check this interpretation of the results, we ran a second experiment, using exactly the same stimuli and tasks, but with the first (lead-in) sentence omitted in each test-sequence. If the early advantage of Normal Prose was indeed due to the presence of a lead-in sentence, then it is clear what effects its absence should have on the pattern of reaction-time differences. And this is indeed what was obtained. The effects in Scrambled and Anomalous Prose were unchanged, while the advantage of Normal Prose early in the sentence essentially disappeared. This result confirms, then, our claims for the early involvement of discourse context in the on-line processing of an utterance.

It is important to realise that these interactions with word-recognition processes are not only taking place early in the utterance, but also very early within the word itself. For example, in one of the tasks
(Identical Monitoring) the mean reaction-time in Normal Prose was 275 milliseconds. Subtracting 50-75 milliseconds for actual response execution, this indicates that the subjects were able to start responding within about 200 milliseconds of the onset of a word. And this is for words whose mean total duration was 370 msec (Marslen- Wilson & Tyler, 1980). Thus we are not only obtaining very fast responses, but also very early responses — in fact, at a point in a word where sufficient sensory input could not possibly have accumulated to allow the word to be correctly identified and responded to on that basis alone.

This type of result (see also Marslen-Wilson & Welsh, 1978) appears to force a model of word- recognition in which contextual constraints interact directly with the selection of the correct word-candidate from among the set of candidates that are still compatible with the sensory input at the point at which these early recognition decisions are being made. This has two main implications. The first is that what is being accessed early in the recognition of a word cannot be just the set of phonological forms associated with different word-candidates, but also the syntactic and semantic properties associated with them. Given this information, the single correct word-candidate can be selected, on-line, as a function both of the accumulating sensory input and of the compatibility and incompatibility of the competing candidates with the contextual environment in which they are occurring. Thus, by the time a word has been recognised — and, usually, before all of it will have been heard — the syntactic and semantic properties of that word will already have begun to influence the development and propagation of the on-line interpretation of the utterance.

The second implication is that the processing system is capable of analysing several competing word-hypotheses simultaneously. Syntactic and interpretative criteria need to be applied to potentially quite large sets of word-candidates; the rapid time-course of recognition processes, and evidence from experiments which manipulate the size of the set of potential candidates (c.f., Marslen-Wilson, 1978; 1980), indicate that these contextual assessments are carried out in parallel, rather than sequentially and one at a time.

To complete this sketch of the speech understanding process, we will briefly describe a second experiment that illustrates in a different way the interactions between knowledge sources that the process seems to permit. This experiment involved stimuli of the following types:
2a) As a traditional way of gaining votes, shaking hands....
2b) If you're trying to thread a needle, shaking hands....
3a) If you walk too near the runway, landing planes....
3b) If you've been trained as a pilot, landing planes....

In each pair there is a structurally ambiguous fragment (such as "shaking hands") preceded by a biasing context clause. In written pre-tests we had established that each context did indeed bias the preferred interpretation of the ambiguous fragment in the desired direction. The purpose of the actual experiment (Tyler & Marslen-Wilson, 1977) was to determine whether these biasing effects could be obtained in an on-line task; that is, whether by the time the ambiguous fragment has been heard, the listener had already established a preference for one reading rather than another.

To do this, we presented the context clauses and the ambiguous fragments auditorily, over headphones. Immediately at the offset of the last word in the fragment (e.g., "hands") we presented visually a continuation word, which was compatible with one of the readings of the fragment. Thus, following the context clause and the ambiguous fragment in, for example, sequence (2a), the subject would immediately see either the word "IS" or the word "ARE". His task was simply to say the word as rapidly as possible. Our assumption was that naming latency would be slower when the continuation word was inconsistent with the preferred analysis of the fragment - but only if the subject had already computed this preferred analysis by the time he saw the continuation word.

This assumption was confirmed by the results. Naming latency to inappropriate continuations was significantly slower than to appropriate continuations. Thus the latency to name "IS" was slower in contexts (2b) and (3a) than in contexts (2a) and (3b), while latency to name "ARE" was slower in the (2a) and (3b) contexts than in the (2b) and (3a) contexts. The speed of the subjects' responses (500-600 msec) meant that they had little time available for post-stimulus analyses. The effects of context could only have derived from analyses that were already available to them when the fragment ended.

These results mean, then, that on-line speech processing involves the rapid and sophisticated integration of at least three distinguishable sources of information - lexical, structural, and interpretative. No single one of these sources could have been
sufficient by itself to produce the obtained on-line preferences. Consider, for example, the context clause "If you walk too near the runway". This context leads to a certain preferred reading of the fragment "landing planes". But the context clause by itself cannot be the only source of the preference. If, for example, the fragment had been "avoiding planes", then the alternative structural reading would naturally have been adopted, even though the context stayed the same. The preferences obtained in this experiment must involve the on-line analysis of the meanings of the words and their structural properties, relative to the most plausible interpretation of this information given the scenario set up by the context clause.

The descriptions of this and the previous experiment give the general flavor of the approach we have been developing to the speech understanding problem - that is, to the characterization of the basic properties of the processes involved, as they take place over time. We emphasise here the finding that the fundamental goal of the speech understanding process - the interpretation of the message - is not only in progress immediately an utterance begins, but is also intimately connected to analysis processes throughout the system. We now turn to the implications of this for the study of anaphora - in particular, for the types of "discourse" or "inter-sentential" anaphora involved in the experiments we will shortly be reporting.

3. Some Implications

Our research indicates that all speech processing involves an immediate attempt at a mapping onto a mental representation of the previous and current discourse interpretation. This means, then, that no lexical item is ever analysed in isolation from the consequences it should have for its interpretative context. Now, the special feature of anaphora is supposed to be precisely this - that their interpretation requires a mapping onto an antecedent [2]; onto a location in the representation of the prior discourse.

But from our present perspective we can see that anaphora are not really special in this respect. It is true that for lexical items which are not in some way semantically attenuated in their specification - as Bolinger (1977; 1979) puts it - one can more readily maintain the illusion that they are processed just on their own terms. We suggest, however, that this is just an illusion. In the case of anaphora, of course, this illusion never became sustainable in the first place.

This is not, however, to minimise the importance of the problem of anaphora in speech understanding. They are indeed attenuated indicators of paths through the
mental representation, and they must, evidently, pose special problems for the speech understander. Our particular hope in studying them, furthermore, is that they represent a simpler, more manageable case of the general mapping problem. This is because anaphors can be clearly located in the local and global structure of a text, and because one can determine quite specifically what the listener has to be able to do in order to make a successful link with the appropriate antecedent location. They might function, then, as a sort of "mother-in-law language" [3] - throwing a simplifying net over the intolerable complexities of the general language understanding process.

A further point is that, given the general interactive nature of speech processing, we should expect to find the same sorts of cooperative interactions going on in anaphoric mapping processes. One can, for example, discriminate a number of apparently distinct sources of information that a listener might use in appropriately mapping an anaphoric pronoun. These would include at least the following: (i) The "lexical" properties of the pronoun itself - such as gender, number, and case. (ii) The location of the pronoun in some syntactic configuration - an example of this would be the sorts of constraints embodied in Lasnik's (1976) NonCoreference Rule [4]. (iii) The relationship of the pronoun to the larger scale configuration of the discourse - here one would invoke notions such as "topic", "focus", "foregrounding", etc. (iv) Inference ("pragmatic" or otherwise) - this would involve, for example, the role of inference in determining whether the properties predicaterv of the pronoun in the utterance are consistent with the properties associated with its proposed antecedent in the discourse model. In certain cases, furthermore, pragmatic inference would be the major determinant of how a pronoun could be mapped - as the following contrast (derived from Sidner, 1979) shows:

1) I took my dog to the vet yesterday
2a) He bit him on the shoulder
2b) He injected him in the shoulder

The only way to determine the appropriate antecedents for "he" and "him" in (2a) and (2b) is on the basis of one's general knowledge about dogs, vets, biting, and injecting.

A basic issue in the psychological study of anaphora should clearly be to investigate how these different potential sources of mapping information operate in the on-line resolution of anaphoric expressions, and to determine the relative salience of each type of source in the mapping process. Is there
evidence, for example, that simple heuristics based on recency [5] or on gender and number matching are preferred for those cases where they are sufficient to ensure correct mapping? From the perspective of our research on general speech understanding, we would expect this not to be the case. If all processing is conducted with regard to the discourse-level interpretability of the consequences of these processes, then this should also apply to anaphor resolution. Heuristics based on gender and number matching may well play a role in the mapping process, but we would expect them to run concurrently with other procedures that assess the overall intelligibility of possible antecedents relative both to the discourse context and to the listener's knowledge of the world. This expectation might be contrasted with the conclusion, in some of the AI literature on anaphor resolution, that inference should be kept out of the resolution process until as late as possible, because inferencing is expensive in time and resources, and is liable to combinatorial explosions unless the inferencing search space is highly constrained.

We will return to this and other issues in the next section of the paper, which presents some of our recent experimental work on anaphoric mapping processes during speech understanding — focusing in particular on the roles of lexical and inferential variables in pronoun resolution. We should stress that each of the three experiments we will describe are to be taken as strictly preliminary; in each case, they represent only the first step in a series. We report them here to illustrate some directions one could take in extending the psychological basis for the study of anaphora.

4. Some Recent Experimental Studies of Anaphora

4.1. The first experiment reported here follows directly from the issues raised above; it is designed to examine the time-course of anaphor resolution, given the availability or not of different sources of information, and to evaluate the relative salience of these different sources. The major contrasts are reflected in the following stimulus set:

1) As Philip was walking back from the shop, he saw an old woman trip and fall flat on her face.

2a) He only hesitated for a moment.
2b) She seemed unable to get up again.

3a) Philip ran towards....
3b) He ran towards....
3c) Running towards....
This stimulus set is made up of a context, or scene-setting sentence, followed by six possible continuations; sentence (1) can be followed either by (2a) or by (2b), and each of these in turn can be followed by either (3a), (3b), or (3c). Each of these sequences was presented auditorily (to different subjects). Immediately at the offset of the incomplete continuation fragment (3a, 3b, or 3c), a visual probe was presented to the subject, whose task was to name this probe word as quickly as possible, and then to write down whether or not the probe was an appropriate continuation of the fragment. For the examples given above, the probe word would be either "HIM" or "HER", and in each case "HIM" would be the inappropriate continuation. Just as in the "shaking hands" experiment described earlier, we expected naming latencies to be slower to inappropriate probes.

The major variable here was the nature of the anaphoric device linking the fragment to its preceding context. In (3a) the device is simply the repetition of the name of the antecedent individual; in (3b) it is an (unambiguous) personal pronoun; while in (3c) there are no explicit linguistic cues at all. It is clear that rather different procedures need to be invoked to cope with these three cases. In (3a) a simple matching of the name could be sufficient. To resolve (3b), a straightforward heuristic could be used which matched the pronoun against the set of possible antecedents, and found the one which was male, animate, and singular. To resolve (3c), however, pragmatic inference has to be invoked. It is necessary to infer, given the scenario set up by the preceding context sentences, who is most likely to be running towards whom [6].

If it is indeed the case - for the sorts of reasons we mentioned in the previous section - that anaphor resolution which depends entirely on inference is more difficult and more time-consuming for the processor, then one should predict a smaller difference between the appropriate and inappropriate probe in (3c) than in (3a) or (3b). Note that for there to be a difference at all - in any continuation condition - the listener has to have determined who is the agent of the verb - who is running towards whom. Otherwise both "HIM" and "HER" would be equally acceptable as continuations.

We also included two other variables in the experiment. The first involved the length of the incomplete fragment, which varied from one to four words in length. For 12 of the 48 test-sequences, the fragment (in its 3c form) consisted of a single word, such as "Punishing" or "Opening"; for another 12 the fragment was a verb plus a preposition or particle - as in the example given above; for another 12 it was a verb plus an adverb, such as "Politely removing" or
"Tearfully welcoming"; or, finally, it could be a verb plus particle or preposition and an adverb, as in "Skillfully probing inside" or "Carefully getting out of". The motivation for this variation was that if the effect in (3c) was smaller - that is, the difference between appropriate and inappropriate probes was smaller - then this should interact with the amount of time the listener had available from the onset of the fragment. For the longer fragments, the effect in (3c) might be larger, and closer to that obtained for (3a) and (3b).

The other variable was a manipulation of the structure of the discourse - namely, whether or not the antecedent for the anaphor in each fragment had been "foregrounded" in the intervening sentence (2a or 2b). Again, one might expect an interaction between this variable and the type of anaphoric linkage. The assignment of Philip as the agent in (3c) might be easier when it was preceded by (2a) - "He only hesitated for a moment" - than when it was preceded by (2b) - "She seemed unable to get up again".

The actual results surprised us, and support a view of anaphor resolution which stresses the dominant role of pragmatic inference. The results, given in Table 1, show an equally large difference between appropriate and inappropriate probes in all conditions. In an analysis of variance, the only significant effect was a large main effect for probe type. Whether the probe came after a more or less explicit anaphor, and independent of the foregrounding manipulation and of the length of the continuation fragment, the subjects had an equally strong preference for one rather than the other of the two probes [7]. The experiment was in fact run two times, on two very different populations of native English speakers, and we found exactly the same results in both cases.

The conclusion to be drawn from this outcome is that purely inference-based anaphor resolution does not necessarily cost the listener significantly more in

<table>
<thead>
<tr>
<th>Continuation Type</th>
<th>2a App.</th>
<th>Inapp.</th>
<th>Diff</th>
<th>2b App.</th>
<th>Inapp.</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>3a. Full NP</td>
<td>379</td>
<td>427</td>
<td>48</td>
<td>378</td>
<td>431</td>
<td>53</td>
</tr>
<tr>
<td>3b. Pronoun</td>
<td>382</td>
<td>432</td>
<td>54</td>
<td>388</td>
<td>436</td>
<td>48</td>
</tr>
<tr>
<td>3c. Zero</td>
<td>381</td>
<td>417</td>
<td>36</td>
<td>387</td>
<td>423</td>
<td>35</td>
</tr>
</tbody>
</table>
terms of either time or success. The different sources of information available in the three types of continuation seem to be functionally equivalent when one tests at the end of the verb immediately following the anaphor. This is consistent with a strong version of the on-line interactive view we have been presenting.

Note, however, that the results do not mean that only inference is used to resolve anaphors. It certainly must make a difference whether one hears a case like (3a), (3b), or (3c). But what the results do suggest is that, whether other sources of information are available or not, the resolution process does not go through without "pragmatic checking" (c.f., Sidner, 1979). Even though, in case (3b), the lexical information carried by the pronoun is, strictly speaking, sufficient to uniquely determine the correct antecedent, the assignment process may remain incomplete without the further information, carried here by the verb, which allows an inferential check on the interpretative plausibility of the proposed antecedent. In the example given here, the acceptability of Philip as the antecedent is confirmed by the likelihood that it is he (rather than the old woman) who is performing the action of running.

This would mean that all three types of continuation tested here have in common the use of inference as an essential (and normal) part of the resolution process. What the results for the (3c) case show is that inferencing without the constraints provided in (3a) and (3b) need not create special problems for the human speech understander.

Finally, the failure here of the discourse structure manipulation - the contrast between (2a) and (2b) - to affect performance should not be taken to mean that discourse variables do not interact with anaphor resolution! The particular manipulation used here was evidently too weak to set up pronounced discourse structural constraints on future likely anaphors.

4.2. The second experiment [8] we will describe here is a developmental one; the reason we are citing it is because it suggests that the developmentally basic strategies for establishing links between utterances are pragmatic and inferential in nature. As Karmiloff-Smith (this volume) has demonstrated, children below the age of six, while able to use pronouns appropriately as far as agreement in gender and number is concerned, do not seem to do so in a way which ensures their unambiguity in a discourse context where there is more than one potential antecedent of the appropriate sex and gender. The listener has to rely on the non-linguistic aspects of the communicative situation to determine who or what the child is referring to. The present experiment
suggests that five year-olds also have problems with pronoun anaphors in the comprehension of utterances.

The experiment used stimuli of the following sort:

1) Jan was trying to catch some butterflies

2a) The butterflies had beautiful wints (wings) and...

2b) They had beautiful wints (wings) and...

The children heard short texts, containing these two-sentence pairs - i.e. (1) followed by either (2a) or (2b) - and their task was to press a response key whenever they heard a "mispronounced" word. In the example here, the word "wings" is mispronounced as "wints" (the actual experiment was run in Dutch, and the above example is a loose translation of one of the stimulus sets). Earlier experiments by Cole and Jakimik (1980) had used this task successfully with first-grade children.

The reason for using the mispronunciation detection task here was that reaction-times in this task are facilitated when the original word - the source for the mispronounced word - is more contextually predictable. To determine that a word has been mispronounced, one needs to establish what the source word was, and this process is apparently facilitated when the contextual constraints on the source word are stronger. In the present example, the word "wings" (the source for "wints") is only predictable if one knows that it is "butterflies" that are the subject of the sentence. But in sentences like (2b) one can only know this if one has established that the antecedent of "They" is indeed "butterflies". Thus, if there are any developmental differences in children's ability to resolve pronoun anaphors on-line, then this should produce a slower response in (2b) relative to (2a), where "butterflies" are explicitly mentioned in the sentence.

<table>
<thead>
<tr>
<th>Age-group</th>
<th>Continuation Type</th>
<th>Nounphrase</th>
<th>Pronoun</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Years</td>
<td></td>
<td>709</td>
<td>769</td>
<td>60</td>
</tr>
<tr>
<td>7 Years</td>
<td></td>
<td>560</td>
<td>561</td>
<td>1</td>
</tr>
<tr>
<td>10 Years</td>
<td></td>
<td>495</td>
<td>500</td>
<td>5</td>
</tr>
</tbody>
</table>

The results are given in Table 2, for the three age-groups we tested. The seven and the ten year-old groups show no differences between the pronoun and the
full nounphrase case, whereas the five year-olds are significantly slower in the pronoun case (note that the absolute differences in reaction-time between age-groups are not important here; what matters are the relative differences between the two continuation types).

This means that the predictability of "wings" - the source for "wints" - was the same for the older children independent of whether the continuation had an anaphoric repeated noun or an anaphoric pronoun in subject position. The anaphoric pronoun was apparently functioning to bring the properties of its antecedent into play in the subsequent processing of the utterance. The five year-olds, in contrast, did not seem to be able to benefit to the same extent from the presence of the pronoun. This suggests that children of this age have not yet fully developed the processing skills needed to exploit properly the information carried by the lexical properties of a pronoun.

We now have in progress further experiments, investigating in more detail just what the problem is here for the five year-olds. But in an earlier experiment with five year-olds, using the monitoring task mentioned in section 2 above, we found the same pattern of immediate discourse mapping early in the sentence as we did for adults (Tyler & Marslen-Wilson, 1980). This shows that young children, just as much as adults, are conducting their on-line analysis of an utterance within the framework of its discourse context. Thus their problems in the present experiment are unlikely to be caused by some general failure to relate utterances adequately to the context in which they are being heard (see also Tyler & Marslen-Wilson, 1978).

We suggest instead - and this is consistent with Karmiloff-Smith's analysis of her results - that young children's construction of a discourse-level interpretation of an utterance is primarily guided by pragmatic and inferential processes, and that there is a developmental lag in the ability to use other cues for discourse mapping. The present results, together with those of Karmiloff-Smith's (see also Karmiloff-Smith, 1979), suggest that there are at least two components to this lag. The first involves the child's sensitivity to constraints on pronominalisation provided by the structure of the discourse. Karmiloff-Smith shows, for example, that five year-olds regularly fail to observe what she calls the "thematic subject" constraint on pronominalisation when they produce a discourse. This implies that they should also have difficulties in making use of discourse configurational constraints in the comprehension of pronouns.

The present experiment does not bear directly on this claim, since it did not involve strong discourse constraints - in the example given above, both "Jan" and
"the butterflies" could equally well have been pronominalised. What the experiment appears to reflect instead is a second type of difficulty for young children — namely, in exploiting the information carried by the lexical properties of an anaphoric pronoun. Karmiloff-Smith (1979) has reported related effects in an experiment on story comprehension in children aged between four and ten years. She used, for example, a story about a child who took the only apple from a bowl of fruit. In this story the information that there was just one apple was carried solely by the (lexical) contrast between the definite and the indefinite article. Up until the age of about eight, the children typically failed to draw the proper implications from this lexical cue. When asked, at the end of the story, how many apples were left in the bowl, they would tend to reply that there were several. They justified this claim on the basis of their knowledge of normal situations in the world — bowls of fruit usually have lots of apples in them, and so on. Since these were children who could correctly interpret a definite article when heard in an isolated sentence, Karmiloff-Smith argues that their failure to do so in a story context reflects the dominance of pragmatic variables in their interpretation of a story.

These two areas of difficulty — with the lexical and the configurational properties of pronoun anaphors — suggest that, for the five year old, the presence of a pronoun in an utterance may do little more than to indicate that something or someone that has already been mentioned will now be talked about again. Determining exactly what is now being talked about will then depend on pragmatic inferences based on what is said elsewhere in the string. If this view of the developmental sequence is correct, then it is certainly consistent with an emphasis on the dominant role of pragmatic inference in the interpretation of anaphors. The additional discourse processing skills that develop later in childhood should therefore be seen as a sort of overlay on a primary process that operates in terms of the plausibility of antecedent assignments in a given discourse context.

4.3. The third experiment [9] represents an attempt to track the precise time-course of the mapping processes involved in the resolution of an anaphoric pronoun. We were hoping to pinpoint the particular instant at which, as it were, an anaphoric pronoun homed in on its antecedent. In the light of the other experiments we have described, and of the results we will now present, this enterprise was based on too simple a conception of the anaphor resolution process.
The experiment used stimulus sentences of the following sort (the actual stimuli were in Dutch):

1) The sailor * tried to save the cat *,
   (a) but * it * fell * overboard * instead.
   (b) but * he * fell * overboard * instead.

There were two versions of each sentence, which were identical except for the pronoun at the beginning of the second clause (1a and 1b). This meant that the critical cue for pronoun resolution was the lexical properties of the pronoun; in (1a) "it" cannot refer to "the sailor", while in (1b) "he" cannot refer to "the cat". The other potential inputs to anaphor resolution - syntactic and discourse configurations, and pragmatic inference - could not distinguish between "the sailor" and "the cat" as potential antecedents for an anaphoric pronoun in subject position in the second clause. This held true for all the test-sentences in this experiment.

Given this simplified situation, we used the cross-modal lexical decision task developed by Swinney (1979) to try to determine when an unambiguous anaphoric pronoun contacted its antecedent. The stimulus sentences were presented to the subjects auditorily, and at different points during the sentence - marked with a "*" in the example - a word would be flashed up on a screen in front of the subjects. Their task, apart from listening to the sentence, was to make a word/non-word decision about the visual probe. Swinney's research shows that this type of lexical decision is facilitated when the visual probe occurs immediately after a word in the auditory input which is associatively related to the probe. In pre-testing we obtained associates to the subject and object NPs in the first clauses of our test-sentences (e.g., for "sailor" and "cat" in the sample above). We then established that visual lexical decisions to these associates ("boat" and "dog" in the present case) were facilitated when they were presented immediately after the NP with which they were associated - thus "boat" would be presented immediately following "sailor", at the first location marked with a "*" above. Decision-times averaged 640 msec for these cases, relative to decisions for the same probes in control sentences, where decision-times averaged 680 msec.

These results provided the basis for the main experiment, in which we presented the same visual probes at the four probe positions indicated for the second clauses of the test-sequences. Thus, across groups of subjects, the probes "boat" or "dog" (for example) could appear either immediately before or after the pronoun, or at one of the two probe positions further downstream. Our assumption was that, at the point at which the
pronoun made contact with its antecedent, there should be a facilitation of a lexical decision about the probe related to that antecedent. If, for example, mapping was achieved at the probe position immediately following the pronoun, then, for sentence (la) above, responses to "dog" should be facilitated relative to responses to "dog" in the same probe position in (lb), where the pronoun indexes the sailor rather than the cat. This would give us an estimate of the time-course of a basic anaphoric mapping process.

The results provided no evidence for a differential facilitation of the type we had hypothesized. Across the four probe positions in the second clause, decision-times varied between 620 and 640 msec. This reflects some degree of facilitation of the probes, since they are being responded to faster than in the control conditions cited above. But there was no sign of any differential facilitation as a function of the relationship of a probe to the pronoun, at any of the four positions. To the extent that facilitation reflects the activation of the semantic regions associated with the potential antecedent NPs, then these results suggest that both antecedents were being activated during the analysis of the second clause, irrespective of which antecedent was actually indexed by the pronoun occurring at the beginning of that clause.

This interpretation of the results is of course preliminary, since a number of additional control studies need to be run. Nonetheless, it is clear that there is no support here for the view of pronoun resolution which had motivated us to run the experiment in the first place. Namely, the assumption that the resolution of a lexically unambiguous anaphoric pronoun is mediated by a unitary, all-or-none process, which sorts through the set of possible antecedents until the only matching antecedent is contacted, and that the properties of that antecedent - and of that antecedent only - are then made available for the further processing of the utterance. If there were some discrete choice-point of this type, then we should have been able to detect a differential facilitation of the probe related to the pronoun at at least one of the probe locations.

The analysis that fits the results best - that both antecedents are activated early in the second clause, and remain activated for at least the next few words - is consistent with the picture of anaphor resolution suggested by the other two experiments presented in this section. According to this view, the constraints on anaphor resolution that derive from the lexical properties of a pronoun are a relatively subsidiary aspect of the resolution process. The more dominant factor is what we have been calling pragmatic inference.
- those processes that assess the interpretative plausibility of potential antecedents. But for these processes to function rapidly and effectively, we need to assume - just as in the case of spoken word-recognition - that they can assess more than one interpretative possibility simultaneously. They should not necessarily be restricted to just a single antecedent selected on the basis of other criteria, but rather should be able to operate on the wider set of potential antecedents offered by the discourse at that particular moment (this implies, of course, that the configurational properties of the discourse may themselves be capable of restricting the set of alternatives that will be considered; however, this is not a problem we have addressed experimentally here).

This type of parallel interpretative assessment has already proved necessary to account for the general properties of spoken word recognition in discourse contexts. There seems no reason not to extend it to the analysis of anaphoric pronouns, since these differ only in degree of semantic specificity from other, less explicitly anaphoric lexical items (see sections 2 and 3).

If we do make this extension to the processing of anaphors, then this provides a plausible framework for the results of the experiment reported in section 4.1. Take, for example, cases (3b) - "He ran towards...", and (3c) - "Running towards...". The presence of the pronoun in one case, and the absence of a surface subject in the other, would alert the processor to the existence of an anaphor in the speech stream. Given the two antecedents previously mentioned - and given that the discourse constraints here were not strong enough to pre-select an antecedent - then both of these would be assessed for their interpretative plausibility, in the light of the properties required of the correct antecedent by the following verb. In this framework, the function of the pronoun would not be to select, in an all-or-none manner, a single unique antecedent, but rather to given a positive weighting to the antecedent that best matched its lexical (and configurational) properties. The absence of such a weighting, as in the (3c) cases, would not, therefore, change the basic nature of the resolution process. And since the experimental stimuli were constructed so that pragmatic inference was always sufficient to unambiguously select one antecedent rather than another, then the outcome of the resolution process - testing at the end of the verb - would be the same whether a pronoun was present or not.

This tentative framework is also, of course, very close to the account we proposed (in 4.2) of young children's processing of anaphoric pronouns. As such,
it makes it clear that the differences between older and younger children in this domain do not reflect qualitative differences in processing strategies. The five year-olds' poorer performance in the pronoun condition simply reflects a developmental lag in the ability to exploit all of the weighting information carried by a pronoun.

We note, finally, that the present account also accommodates a number of otherwise puzzling observations we have made about the on-line interpretability of anaphoric pronouns in natural conversational discourse (Marslen-Wilson, Levy, & Tyler, in preparation). Although adult speakers produce a high proportion of pronoun anaphors in such a way that they are, in principle, immediately resolvable (given their lexical and configurational properties), there are also numerous exceptions. First, and consistent with our proposals above, there are pronouns which are lexically and configurationally ambiguous between more than one potential antecedent, and where the properties predicated of the pronouns by the rest of the utterance provide inferential disambiguation. Second, and most significant for the present argument, there are some cases where the lexical properties of the pronoun clearly imply a particular antecedent, but where the subsequent properties predicated of the pronoun are pragmatically incompatible with this antecedent, and point to a quite different one [10].

If lexical variables were operating in a context of all-or-none selection, then one would expect cases of this type to produce "garden-path" effects in the listener. But, in informal observation, we have noticed no such effects. In fact, we only detected these cases in the first place as the result of a systematic analysis of the on-line interpretability of each pronoun anaphor in the discourse. If one listens to the discourse normally, then these "violations" slip by without causing any apparent perturbations. This is difficult to explain unless one assumes that the listener can indeed carry along multiple interpretative hypotheses, and may delay commitment to any single one of these until pragmatic inferencing, relative to the discourse and the world, confirms one particular assignment and rules out any others.

If this analysis is correct, then the results of the present lexical decision experiment can be taken at their face value - namely, as reflecting the multiple activation of potential antecedents. The lexical constraints carried by the pronoun here did not lead to an immediate all-or-none choice of a single antecedent, but only to a weighting in favour of one of the two antecedents being actively considered. It remains to be seen in future experiments whether discourse structural
constraints, unlike local lexical constraints, can have stronger effects on the selection of antecedents. The following sort of sequence suggests that they can:

1) The little puppy trod on a wasp.
2) The puppy was very upset.
3) It started to buzz furiously.

Most listeners experience a form of garden-path in the third sentence, indicating that by the time they get to "buzz" they have firmly committed themselves to "puppy" as the antecedent of "It". Given that the sequence "It started to" is by itself compatible with both "wasp" and "puppy", this commitment to "puppy" must be due to discourse structural constraints.

5. Concluding Remarks

In the first sections of this paper we set up a general view of how we think human speech understanding operates in real time. The results of the experiments we have just described show that this general perspective leads to a plausible preliminary account of some aspects of anaphoric mapping processes. This account appears, furthermore, to be broadly compatible with the other relevant psychological and artificial intelligence work on anaphora.

Most psychological studies of anaphora, it should be noted, have tended to use written rather than spoken language, which leads to some major differences from the kind of research we have been describing here. First, in cases where a complete text is displayed to the subject - a situation similar to normal reading - there is the opportunity, as Carpenter & Just (1978) show in some eye-movement studies, for the reader to look back at earlier sections of the text, to check, for example, on the properties of some potential antecedent. This is evidently a different situation from listening to a spoken discourse, where the previous input cannot be available for inspection in the same way, so that different requirements are placed both on the speaker's output and on the listener's memory representation of this output.

Second, in those experiments which have used the successive presentation of a sequence of sentences (e.g., Garrod & Sanford, 1977; Haviland & Clark, 1974), the standard measurement is either total reading-time for the critical sentence, or "time to comprehend" - i.e., the delay between the presentation of the sentence and the point at which the subject presses a response-key to indicate that he or she has understood the utterance. This type of measure turns out to be sensitive to a variety of variables affecting anaphor
resolution, but it cannot, nonetheless, probe the internal structure of the mapping process in the same way that on-line probes during the presentation of an (auditory) input are able to.

These reservations aside, the research on anaphor resolution in written texts seems consistent with the approach we have taken here. The series of studies, for example, by Caramazza and his colleagues (e.g., Caramazza, Grober, Garvey, & Yates, 1977; Caramazza & Gupta, 1979) confirm that a variety of different forms of constraints are brought together in the resolution of a lexically ambiguous anaphoric expression. These studies (see also Cowan, this volume) typically use isolated sentences of the form "The criminal attacked the policeman because he was stupid", where the pronoun "he" is ambiguous. The experiments show that a variety of factors - including, among others, the semantics of the verb in the first clause, the nature of the connectives used ("and" vs. "because"), and the ordering of the main and subordinate clauses - can affect subjects' judgements about the antecedent of the pronoun. One clear result of these studies is that sentential configurational factors - as reflected, for example, in the "parallel function" constraint (c.f., Cowan, this volume) - are not by themselves sufficient to determine co-reference. A dominant consideration is always the pragmatic plausibility of an antecedent assignment, in the light of the scenario implied in the first clause and of the properties predicated of the pronoun elsewhere in the second clause.

Another type of experiment, exemplified by the work of Garrod and Sanford and of Clark and Haviland, has used short texts to examine the role of inference in the linkage of anaphoric elements to the preceding context. Haviland & Clark (1974) and Clark & Haviland (1977) focussed on what they call "bridging inferences", using contrasts like the following:

1a) Mary unpacked the picnic things.
1b) Mary unpacked the beer.

2) The beer was warm.

Overall "comprehension time" for sentence (2) is slower when it follows (1a) than when it follows (1b). This is taken as evidence, first, that the comprehension of the anaphoric nounphrase "The beer" (its anaphoricity being indexed by the presence of the definite article) involves its linkage to the prior text, and, second, that this linkage requires a bridging inference (in this case that the picnic things must have included some beer) when the noun has not been explicitly mentioned before. Such inferences, it is assumed, take time,
which slows down comprehension in the (1a) cases. Recent work by Garrod & Sanford (1978) shows, however, that not all bridging inferences necessarily cost more in terms of reading time. They examined contrasts like the following:

1a) Mary put the baby's clothes on.
1b) Mary dressed the baby.

2) The clothes were made of pink wool.

Garrod and Sanford assume, following AI researchers such as Schank (1973) and Minsky (1975), that the mental representation of (1b) would include the information that what Mary was doing was to put clothes on the baby, since that is what it means to dress someone. The experimental hypothesis was that, under these conditions, reading time for (2) would be the same following both (1a) and (1b). In both cases, the mental representation of the discourse would include clothes - in (1a) because they were actually mentioned, and in (1b) because they were directly presupposed by the action of dressing. Thus any linkages that needed to be constructed between (2) and the preceding sentence would be operating under essentially identical informational conditions. The results confirmed this, in that reading times did not differ in the two context conditions (1a and 1b).

Such studies, of course, differ from ours in their focus on the processing of anaphoric definite nounphrases, rather than of anaphoric pronouns. But, like ours, they emphasize the role of inference in anaphor resolution; in addition, they suggest limitations on the kinds of inference that - in the spoken language situation - the speaker might reasonably expect the listener to draw without disrupting the flow of on-line interpretation.

In contrast to the relatively scattered psychological studies, there has been extensive research on anaphora in the field of artificial intelligence - if only because anaphora present a serious practical problem for natural language understanding systems. We cannot attempt to do justice here to this large body of work, and, in any case, it remains difficult to define the proper relationship between AI treatments of anaphor resolution, expressed in the form of explicit algorithms, and the still rather indistinct generalities which are all that the on-line psychological data allow us to express. The available psychological data are simply not adequate to assess the psychological relevance of the details of the AI proposals.
What the AI literature does provide the psychologist, at present, is with a series of clearly expressed observations about the kinds of informational variables that a processing system capable of interpreting anaphoric expressions needs to be able to handle. The most relevant to our present arguments is the use of inference in AI systems, going back at least to Charniak's (1972) observations. In fact, much of the subsequent history of AI research on anaphora can be seen as an attempt to develop techniques for involving inference in the resolution process that were less explosively unconstrained than the methods that Charniak (1972) and Rieger (1974) originally proposed (for some recent treatments, see Hobbs, 1976; 1979; Nash-Webber, 1978; Sidner, 1979).

Sidner (1979), for example, presents an interesting set of algorithms centered around the notion of "focus" (c.f., Chafe, 1976; Grosz, 1978; Halliday, 1967) as a way of selecting and ordering the set of potential antecedents for an anaphoric element. As the program analyzes a discourse, it carries along with it a primary focus element, an ordered list of alternative focus elements, and a stack of old foci. This provides the system with a means of determining in what order potential antecedents should be checked when an anaphoric element is encountered. Thus inferencing procedures need not be applied indiscriminately to all potential antecedents, but rather, in a predetermined order, to those antecedents (and the knowledge associated with them in the discourse model) that have been most salient in the prior discourse.

While one would not expect to translate Sidner's procedures directly into a psychological model, they do illustrate the potential importance of discourse focussing factors in anaphor resolution. As we have mentioned elsewhere in this paper, it is likely that such discourse factors, unlike local lexical factors, do function psychologically to pre-select potential anaphors for on-line consideration, even in a device apparently capable of multiple parallel assessment processes. Sidner's formulation of these discourse factors would clearly form a useful background to psychological experiments on this issue.

A somewhat different perspective is provided by the work of Hobbs (1976; 1979). His approach is based on the role of "coherence" relations in the understanding of discourse. Speakers produce utterances which cohere in various ways (c.f., Halliday & Hasan, 1976) to what has been said previously, and the primary goal of the listener is to discover how this coherence is realised in the utterance - how what the speaker is saying is relevant to what has just been said. Hobbs' argument is that many problems in anaphor resolution will be solved,
without any extra processing, as a natural by-product of this primary operation of coherence recognition.

Take, for example, the following sequence (from Hobbs, 1976):

1) The boy walked into the bank.
2) Moments later he was seen on its roof.

Hobbs' claim here is that the resolution of the anaphoric "its", in sentence (2), does not require a special-purpose anaphor resolution algorithm to be invoked. It derives, instead, from the operation of coherence assumptions which merge the common elements in the representations of (1) and (2). The interpretation of "walked into" in (1) requires the program to place in the representation the fact that "the bank" is "a building" (and, therefore, something which can be walked into). Then, in the interpretation of (2), the presence of "roof" also brings "building" into the representation. Given the assumption of coherence, the occurrences of "building" - and their associated arguments - in (1) and (2) are merged into a single entity in the discourse model. The by-product of this merging operation is that the antecedent of "its" has thereby been identified as "the bank".

The point here is not the details of Hobbs' particular procedures, but rather that anaphor resolution need not be anaphor resolution as some special act, but instead may fall out of the general process of mapping from the utterance onto the interpretative model of the discourse. This, of course, is very close to the observations we made earlier in the paper, where we emphasised the on-line interpretative goals of the listener, and suggested that anaphor resolution did not place demands on the processor that were different in principle from those required in the proper interpretation of any lexical item.

We conclude, then, by pointing out that the deepest problems posed in the explanation of anaphor resolution are identical to those posed in the explanation of on-line speech understanding in general. Namely, to find some psychologically intelligible theory of the basic categories and the basic configurations of the cognitive space within which listeners construct their on-line interpretation of a discourse. These categories and configurations must determine how utterances are interpreted, how anaphoric connections can be made, and, indeed, what can be indexed anaphorically in the first place. The research we have described here is intended to constrain a psychological theory of this space, by laying out the dynamic properties of the mental processes for which it provides the substrate.
Footnotes.

1. Normal Prose could also allow, in principle, for the construction of a semantic interpretation of the classical type-based, that is, on the meanings of the words and the grammatical relations between them. There is no on-line evidence, however, to suggest that such a form of analysis functions as an independent processing operation in speech understanding; word-senses appear never to be analysed without reference to the discourse and world context in which they occur (c.f., Marslen-Wilson & Tyler, 1980).

2. Our use of the term "antecedent" should not be taken to imply that anaphora always "refer back" to some discrete entity already mentioned in the text.

3. We apologise to R.M.W.Dixon for any discomfort this simile might cause!

4. For a different perspective on some of the traditional syntactic configurational constraints, see the discussion in Bolinger (1977; 1979).

5. One example of a recency heuristic is given by Sidner (1979) - if a pronoun occurs in subject position, then first check to see if it matches the last constituent in the previous sentence.

6. The sequence "Running towards" in the given contexts is in principle still ambiguous, since there could have been some third party who was now running towards either of the two individuals already mentioned. However, in the written pre-tests of these materials, "Philip" was always chosen as the agent.

7. There is, admittedly, a trend in the data for a smaller difference in the (3c) cases. But not only does this trend not approach statistical significance, but also, as Table 1 shows, it derives from a slight decrease in the effects for the inappropriate probe, rather than from a change in the effects for the appropriate probe. This means that the correct continuation was equally strongly acceptable in all three conditions.

8. This experiment was carried out with the help of Peter Jeurisson.

9. This experiment was carried out in collaboration with Tony Ades, and with the help of Colin Brown and Pie nie Zwitserlood.

10. An example of this would be the use of the generic "they" in a context where plural antecedents are already present in the discourse, and are also being indexed by "they".
References.


Cowan, J.R. (this volume) The significance of parallel function in the assignment of intrasentential anaphora.


Garrod, S. & Sanford, A.J. (1978) "When do we need to make bridging inferences?" Manuscript, Glasgow University, Scotland.

Karmiloff-Smith, A. (this volume) Pronominalisation and non-pronominalisation in connected discourse.


Acknowledgements.

This paper was written while the authors were Postdoctoral Fellows at the Center for Cognitive Science at MIT. We thank, in particular, Ed Walker for his help with many aspects of this paper's preparation.