# Temporal Coherence and Defeasible Knowledge*

Alex Lascarides Jon Oberlander

Human Communication Research Centre
University of Edinburgh
Draft: October 10, 1997

**Address for Correspondence**

**Contact:** Alex Lascarides

**Address:** Human Communication Research Centre, University of Edinburgh, 2 Buccleuch Place, Edinburgh, EH8 9LW, Scotland.

**Phone:** +44 31 650 4448

**Email:** `alex@uk.ac.ed.cogsci`

---

*Earlier versions of this paper were presented at the University of Cambridge, in January 1991, and at the Workshop on Discourse Coherence, held in Edinburgh, in April 1991, under the auspices of the European Community Basic Research Action in Dialogue and Discourse. We thank Nick Asher, Mimo Caenepeel, Herb Clark, Dick Crouch, Robert Dale and Victor Poznanski for helpful discussions on the ideas reported here. The authors gratefully acknowledge the support of UK Science and Engineering Research Council through project grant GR/G22077. HCRC is supported by the UK Economic and Social Research Council.

## List of symbols

**Greek:** $\alpha$ $\beta$ $\phi$ $\psi$

**Logic:** $\wedge$ $\vee$ $\neg$ $\forall$ $>$ $\rightarrow$ $\Box$

**Other:** $\langle$ $\rangle$ $\prec$ $\emptyset$

## Statistics

**Number of pages:** 37 (including this pre-amble, figures etc.)

**Number of tables:** 0

**Number of figures:** 8

**Running head:** Temporal Coherence

## Abstract

We discuss data involving the temporal structure of connected discourse. Questions are raised about the relation between clause order in discourse and causal order in the world, and about the coherence of certain discourses. We maintain that interpretation is contextually influenced by knowledge of the world and of pragmatics, and that the role of this knowledge should be formalised via a defeasible logic. It transpires that a constrained set of reasoning patterns underlies the retrieval of certain temporal structures. Not all defeasible logics capture the set; the data help choose between candidate logics. We demonstrate that an adequate logic characterises when a text is temporally coherent, reliable and unambiguous relative to the context. We also discuss defeasible reasoning in language generation, and some consequences for the semantics-pragmatics interface.

# 1 Introduction

An account of text interpretation must answer the following question. Given the way events are described in the text, what are their causal and temporal relations in interpretation? And similarly, an account of text generation must answer an analogous question. Given the causal and temporal relations between events in a knowledge base, what are the ways they can be described in text?

This paper has five basic goals. First, we wish to indicate a range of data involving the temporal structure of connected discourse; there, the main issues concern the relation between the order of clauses in a discourse and the order of eventualities in the world, and the coherence (or otherwise) of certain discourses. Secondly, we wish to show how a system (human or artificial) can exploit knowledge of the world, and of pragmatics, to resolve problems in the data. Thirdly, we aim to formalise the role of that knowledge via a defeasible logic. Fourthly, we aim to choose between different possible logics for defeasible reasoning; not all candidates can capture the range of reasoning patterns we require. Lastly, we wish to discuss the role of defeasible reasoning in generating connected discourse, and its effect on semantic structure. A general consequence of our discussion is the following. Defeasible reasoning can be used to model the pragmatic interpretation of text, over and above its semantic interpretation; or it can be used to choose among several possible semantic representations of text; or it can be used dynamically to construct the semantic structure of text.

# 2 Data, definitions and intended coverage

In this section, we will first outline the three basic types of temporal data we wish to discuss, then introduce some terminology intended to clarify the discussion, and finally note various limitations on the coverage of the approach to be elaborated in this paper.

## 2.1 Reversals, incoherence and ambiguity

The three pairs of texts that follow exemplify the first of the three basic categories of data that we wish to address.

(1) Max stood up. John greeted him.

(2) Max fell. John pushed him.

(3) Max opened the door. The room was pitch dark.

(4) Max switched off the light. The room was pitch dark.

(5) Max took an aspirin. He was sick.

(6) Max took an overdose of aspirin. He was sick.

These examples expose a puzzle concerning the relation between the description of events and their temporal order in the world: the syntactic structure of the sentences cannot fully determine this relation. The sentences in texts (1) and (2), for example, have similar syntax. The natural interpretation of (1) has the descriptive order of events match their temporal order. However, in introducing this problem to the literature, Moens (1987:92) notes that in cases like (2), descriptive order *mismatches* temporal order. Indeed, if (1) were meant to describe a situation where the greeting occurred before Max stood up, then at best it would be misleading. At least, it would be misleading unless the reader *already* knows, from the extra-linguistic or linguistic context in which (1) is uttered, that the greeting occurred before Max stood up. Texts (3) and (4) also have similar syntax. But the natural interpretation of (3) has the state of darkness temporally *overlap* the event of Max opening the door; whereas in (4) the event of Max switching off the light *precedes* the darkness.

Existing treatments of tense, such as Kamp & Rohrer (1983), Partee (1984) and Hinrichs (1986) account for the forward movement of time in (1) by encoding in logical form the temporal order between the events concerned. This logical form is built solely from the syntactic structure of the text's sentences. None of these theories attempt to account for the apparent backward movement of time in (2), since they are concerned only with narrative texts like (1), (3), (4) and (6). But even within the narrative domain, the account isn't quite complete, in that they don't predict that the preferred reading of (3) is different from (4). In extending the theories to account for the above texts, it would be necessary to assign (1) a distinct logical form from (2), in spite of the similar syntax of the sentences. This would involve extensive revisions to the way logical form is constructed.

Dowty's (1986) treatment of tense also deals only with narrative texts. The only basis for distinguishing temporal structures in this theory is the event/state distinction. To extend Dowty's account to explain the difference between (1) and (2) would require additional structures, because both texts refer to events alone. Webber (1988) can account for the backward movement of time in (2). However, her theory is unable to predict that, given a context, mismatching the descriptive order and temporal order of events is misleading in some cases (e.g. (1)) but not in others (e.g. (2)).

It might be thought that a simple solution is to say that there is some "default" principle working along the following lines. Unless there is causal information about the eventualities, descriptive order matches temporal order. When there is causal information which somehow "overrides" this principle, descriptive order mismatches temporal order. But the existence of pairs such as (5) and (6) indicates that causal information need not have this effect. In (5), we have the intuition that the sickness leads to the taking of aspirin; but in (6), we have the intuition that it is the overdose that causes the sickness, rather than vice versa. To handle such cases, the simple solution would need to be made considerably more precise.

The second main category of data we wish to consider involves, in one way or another, the idea of incoherence.

(7) ?Max won the race. He was home with the cup.

(8) Max won the race. He arrived home with the cup.

The query in (7) is meant to indicate a question about its acceptability. We believe that it is less good than (8); and that its relative unacceptability is not to be traced to the grammatical or semantic unacceptability of either member of the pair. Rather, the problem lies in the juxtaposition of the two sentences: we have the intuition that the state mentioned in the second sentence of (7) cannot really overlap with the event mentioned in the preceding sentence. And yet the overlap is somehow implied by the fact that the event and state are described consecutively in the particular way featured in (7). The problem goes away when we replace the state-description with an event-description.

The third category of data involves the idea of ambiguity. Given the way thermostats work, the events described in (9) could conceivably be related in one of two ways: either the strip changing shape caused the temperature fall; or the temperature fall caused the strip to change shape.

(9) The bimetallic strip changed shape. The temperature fell.

It has been suggested, however, that there may be empirical evidence to support the following conjecture. Readers don't interpret texts like (9) as ambiguous, even if they have no prior knowledge of the temporal order of events described.

On a more general level, however, there is a sense in which ambiguity does arise. A text can correspond to different temporal structures with respect to different contexts of interpretation, as defined by the reader's KB. For example, the reader can infer that the linguistic order and temporal order match in (2) if she *already* knows that the falling preceded the pushing; but she infers that the linguistic order and temporal order don't match in (2) if she doesn't have this prior knowledge. Conversely, in the context where the reader *already* knows the greeting preceded the standing up, the descriptive order of (1) mismatches temporal order; but without this knowledge, the descriptive order of (1) matches temporal order. Ambiguity thus arises in the sense that the same text can have different temporal structures with respect to different contexts.

With regard to ambiguity, then, we will want a theory which can do justice both to the evidence that interpreters don't detect ambiguity in cases like (9); and to the general intuition that the context of interpretation will strongly affect choice among possible temporal structures.

## 2.2 Some definitions

To describe what is going on in these cases, and in particular, to make explicit what we mean by coherence, we now introduce some terminology. Throughout, the term "eventualities" is used as in Bach (1986) to cover both events and states. We define *temporal coherence*, *temporal reliability* and *temporal precision* in terms of a set $C$ of relations between eventualities. This set intuitively describes when two eventualities are connected. The relations in $C$ are: causation, the part/whole relation,[1] temporal overlap, and the immediately precedes relation (where "$e_1$ immediately precedes $e_2$" means that $e_1$ and $e_2$ stand in a causal or part/whole relation that is compatible with $e_1$ temporally preceding $e_2$).[2] The definitions are as follows:

- **Temporal Coherence**
  A text is *temporally coherent* if the reader can infer that at least one of the relations in $C$ holds between the eventualities described in the sentences.

- **Temporal Reliability**
  A text is *temporally reliable* if one of the relations in $C$ which the reader infers to hold does in fact hold between the eventualities described in the sentences.

- **Temporal Precision**
  A text is *temporally precise* if whenever the reader infers that one of a proper subset of the relations in $C$ holds between the eventualities described in the sentences, then she is also able to infer *which* of these two relations holds.

Clearly these definitions bear on the idea of "textual truth" (cf. Kamp 1981). Arguably, for a text to be true, it's not enough for its component sentences to be true; the text must be coherent and reliable as well. There are some obvious corollaries to the definitions. A text is temporally *incoherent* if the natural interpretation of the text is such that there are no inferrable relations between the events. A text is temporally misleading, or as we shall say, *unreliable* if the natural interpretation of the text is such that the inferred relations between the events differ from their actual relations in the world. Roughly, the reader can only infer the wrong member(s) of the set $C$, given what she knows. In addition, a text is temporally imprecise, or as we shall say, *ambiguous*, if the natural interpretation of the text is such that the reader knows that one of a proper subset of relations in $C$ holds between the eventualities, but the reader can't infer *which* of this proper subset holds.

It follows from the above definitions that a text can be coherent but unreliable. For example, if John greets Max and then Max stands up, then (1) is coherent but unreliable. At least, it's unreliable unless the reader *already* knows John greeted Max and then Max stood up, before she calculates the temporal structure of (1). On the other hand, in an example like (7), there may be no question about reliability because we cannot establish a temporal or causal relation between the two eventualities.

### 2.3 Caveats about coverage

With these notions in place, we can say that we wish to provide an account for data which involve temporal coherence, reliability and ambiguity. By showing how to calculate when a text has these properties, we hope to show how to solve the puzzles involving the texts in (1)–(9).

However, before proceding any further, we should note some provisos concerning the intended coverage of this paper. First, we discuss only *local* coherence, not *global* coherence. That is to say: we consider the constraints on temporal coherence holding between two consecutive sentences in a text.[3]We shall therefore not be concerned with constraints on the relations between sentences separated from one another by intervening material. To address these, we would have to have available some notion of hierarchical discourse structure, as in Hobbs (1985), Grosz and Sidner (1986), Mann and Thompson (1987), Scha and Polanyi (1988), or Webber (1991).

Secondly, we discuss only *temporal/causal* relations, not *discourse-structural* relations. This is not because we believe that temporal coherence is sufficient for discourse coherence; we don't. It is because we have chosen to present the argument for defeasible logic in a framework without full discourse structure. As a result, there will be many cases where a pair of sentences appears to be temporally coherent, but fails to achieve discourse coherence, and the current account cannot explain what is wrong with them (cf. Caenepeel 1991 for a discussion of such cases). On the one hand, our discussion here can be seen as extending in certain ways the account in Lascarides (1990). On the other hand, Lascarides and Asher (1991a, 1991b) extend the treatment of temporal relations explored here so as to permit the expression of discourse-structural, or rhetorical, relations, and thereby address the issue of non-local coherence. One way of expressing the difference between our treatment here and the fuller account which builds on it is to say that the binary relation between text and temporal structure characterised here is extended to a tertiary relation between text, temporal structure and discourse structure. The similarity lies in the fact that the reader's defeasible knowledge plays a central role in defining both relations.

A third proviso that should be registered is that cases of unmarked reversal are rare. By an unmarked reversal, we mean texts where the order in which events are described mismatches their order of occurrence, and this mismatch is unmarked by any syntactic indicators (such as the use of the perfect, or *because*). This may help explain the air of artificiality which surrounds the minimal pairs we have discussed. However, we argue below that combining the defeasible knowledge account with a simple generation model permits a natural explanation of the unnaturalness of the invented examples. In spite of this, worries may remain about the soundness of proceding from our minimal decontextualised cases.[4]The ubiquity of real context—linguistic and extra-linguistic—might make us feel that subjects confronted with our examples are like fish out of water, and their actions are inappropriate. Against this, we would argue that one of the major advantages of our account is that it provides a rather precise explanation of the differences in interpretation which arise as contexts are enriched, and made more natural. More specifically, our framework is rich enough to explain that the preferred reading in (10) is one where the pushing caused the falling, but in (11) the preferred reading is one where the falling caused the pushing.

(10) Max had a horrible accident yesterday. He fell. John pushed him.

(11) John and Max were at the edge of the cliff. Max felt a sharp blow to the back of his neck. He fell. John pushed him. Max rolled over the edge of the cliff.

This is because we exploit non-monotonic inference; and thus enlargement of the linguistic and extra-linguistic context can lead to changes in the temporal relations which are inferred.

# 3 Strategy

So we wish to tackle data in which temporal coherence and reliability are central problems. The standard approach to the semantics of the discourse constructs it from syntactic structure alone, and directly encodes temporal relations. As we have noted, the data creates puzzles for such approaches, if they are to be extended to deal with non-narrative texts. A very natural question, however, arises. If coherence and reliability aren't fixed by syntactic structure, where are they fixed?

The answer we wish to explore here is that these properties are fixed by a defeasible logic, which takes relatively simple semantic forms together with other knowledge sources, and infers the temporal/causal relations. A defeasible logic is one in which conclusions may follow from a set of premises, but not from a superset. Differing aspects of the approach have certainly been reconnoitred before. On the one hand, the idea that world knowledge can be used to guide discourse interpretation dates back to Aristotle, and has more recently been defended by Hobbs (1979, 1985) and Dahlgren (1988). On the other hand, the utility of defeasible logic for pragmatic interpretation has been explored by Appelt and Konolige (1988), Wainer and Maida (1990) and Crouch and Poznanski (1991).

### 3.1 The need for defeasible logic

In certain respects, the approach developed in the current paper refines that outlined by Hobbs and Dahlgren. In Hobbs' approach, the causal knowledge is encoded in a declarative but indefeasible knowledge representation. In Dahlgren's approach, the relevant knowledge is encapsulated in terms of episodic knowledge representation. One difference from Hobbs is that we place more emphasis on defeasibility; and one difference from Dahlgren is that we place more emphasis on the need to declaratively specify the relations between the representations. The main difference from both approaches, in fact, lies in the utilization of a constrained representation of defeasible knowledge, in which the underlying relation of logical consequence yields the interactions required.

In Hobbs' and Dahlgren's theories as they stand, it is not completely clear that the requisite notion of logical consequence could be defined, since there are no obvious relations between defeasible laws that ought to interact in certain specific ways. Why should this matter? The reason is that conflicts certainly arise among the knowledge sources their theories recruit, but there is no mechanical way of resolving such conflict. For example, Hobbs (1979) uses causal knowledge to choose the antecedent to the pronoun in text (12); that knowledge is still relevant to text (13), but further conflicting knowledge overrides it. As a result, a different antecedent to the pronoun is chosen.

(12) John can open Bill's safe. He knows the combination.

(13) John can open Bill's safe. He must change the combination.

Equally, Dahlgren (1988) represents linguistic knowledge concerning how textual order affects the temporal structure of the events described. This knowledge explains the difference between (1) and (14); but in the analysis of (2), it is overridden by conflicting world knowledge concerning the typical causal relation between falling and pushing.

(1) Max stood up. John greeted him.

(14) John greeted Max. He stood up.

(2) Max fell. John pushed him.

In both theories, the problem is that particular resolutions of conflict among knowledge sources appear arbitrary; they lack logical justification, given the way the knowledge is represented.

The current work can therefore be seen as building on that of Hobbs and Dahlgren by placing causal knowledge in a logical context where its defeasible implications can be precisely calculated. In this connection, the work which is closest in spirit to ours is that of Crouch and Poznanski (1991): they take care to ensure that the logical consequence relation has the inferential properties they require. They differ in encoding defeasible conditions directly into semantic representation, and in considering the retrieval of temporal relations in *intra*-sentential cases.

### 3.2 Choosing the right logic

Our specific strategy has evolved from Lascarides (1990). There, it was proposed that the relation between the descriptive order of events in text and their temporal order can be characterised in terms of two sets of defeasible rules. The first set represents preferences for *causal relations* among events and states (world knowledge) and the second set represents *linguistic rules* concerning Gricean-style pragmatic maxims (linguistic knowledge). This theory provides a simple account of the difference between discourses in which descriptive order matches temporal order, and those which don't. Sometimes, world knowledge (encoded as a defeasible preference for a causal direction) "overrode" the linguistic knowledge.

A declarative formal model was then developed, based on Levesque's (1990) autoepistemic logic for defeasible reasoning, which allowed different natural interpretations for texts with similar syntax. However, Levesque's logic as it stands does not support certain defeasible inferences that can potentially be exploited to define the temporal structure of text. Hence, even if it is accepted that we require a *logic* for defeasible reasoning, a further question remains: which is the *right* logic? One of our goals is therefore to determine an answer to this question; to choose between candidate logics.

### 3.3 Defeasible Patterns

We would argue that any defeasible logic will be adequate so long as it captures the patterns set out in Figure 1. The reason is that all these patterns are required on the model of textual interpretation we hold.

---

Insert Figure 1 about here

---

In this and subsequent figures, bold arrows correspond to indefeasible rules, and thin arrows to defeasible rules; circles correspond to propositions that feature in rules in the KB and double circles indicate which propositions will actually be inferred in the KB, given an adequate defeasible logic. The cross-hatching between some circles indicates that the relevant propositions are contradictory. In the Nixon Diamond pattern, the lack of double circles indicates that neither of the contradictory conclusions are drawn from the premises.

The patterns in Figure 1 are illustrated there in terms of more or less familiar examples from the knowledge representation literature. The reason we provide these graphical analogues is that they constitute a means of representing the main differences between the patterns, without requiring detailed specifications of the possible logics. For those familiar with defeasible logics, it is probably easier actually to follow the connexions between the formulae themselves. If so, then replace the thin arrows with the defeasible consequence relation which we represent as $>$; and replace thick arrows with the indefeasible connective $\rightarrow$. In summary, the difference between Defeasible Modus Ponens and the Penguin Principle is that in the latter pattern, a new premise is introduced, which can act as antecedent to a second defeasible rule. The consequents of the two rules conflict. However, the new premise indefeasibly implies the old one; because of this, we shall say that it is *more specific*, and we say that the rule which requires the more specific premise is *more specific* than the other rule. We want a defeasible logic to validate a pattern in which the more specific rule is preferred; this pattern turns out to be central to accounting for temporal reversals. The pattern of Resolvable Conflict is very similar to the Penguin Principle, except that the premises are no longer logically related, and one of the competing rules is indefeasible. This time, we want a logic to validate the conclusion of the indefeasible rule. The Nixon Diamond arises when we alter the Resolvable Conflict pattern to make both competing rules defeasible. We want a logic that draws neither of the two competing conclusions. Finally, the two patterns of Weakening and Dudley Doorite are indefeasible, and relatively straightforward.

There are logics of defeasible reasoning that validate these patterns of common sense entailment. Asher & Morreau (1991) present a logic called MASH, where the patterns of reasoning arise from the semantics of the defeasible rules themselves. Hierarchical Autoepistemic logic (HAEL) (Moore 1985, Appelt & Konolige 1988) captures the Penguin Principle by ordering the defeasible rules according to a taxonomic hierarchy, so that the rules with the more specific antecedents have priority over the rules with the less specific antecedents. There, the Penguin Principle is captured by a mechanism that is external to the semantics of defeasible rules. We do not aim to choose between these different approaches here. We will merely explore how the patterns of inference in Figure 1 contribute to local temporal coherence, reliability and precision. Our discussion below will indicate that a formalisation of the defeasible reasoning approach to temporal semantics is possible in either MASH or HAEL. A formalisation in MASH of some of the ideas presented here appears in Lascarides & Asher (1991).

## 4 Reversals: the basic story

Before indicating precisely why a defeasible logic needs to include all the patterns in Figure 1, and how they arise in each case, it is probably helpful to sketch the general form of the explanations we will be pursuing. Let us therefore turn to the first pair of examples: the orderly (1), and the "reversal", (2):

(1) Max stood up. John greeted him.

(2) Max fell. John pushed him.

It seems not unreasonable to suggest that in the second case, what makes the difference is that we have a piece of general knowledge which tells us that if we have a falling and pushing on our hands, it's more likely to happen one way than another. We would like to say that world knowledge (WK) includes some sort of law-like generalisation gained from perception and experience that relates falling and pushing in this way (in what follows, such generalisations will numbered so that their relation to the linguistic examples is reasonably obvious):

- **Causal Law 2**
  Connected events $e_1$ where $x$ falls and $e_2$ where $y$ pushes $x$ are normally such that $e_2$ causes $e_1$.

The generalisation is law-like in the familiar sense that it is not merely an accidentally true generalisation; it could therefore support subjunctive conditional claims. Note that there is no similar law for standing up and greeting. Our "law" is *defeasible*, in the familiar sense that additional information can prevent its conclusion from following. Such generalisations are standardly termed *ceteris paribus laws*; they hold true, other things being equal. If we have information which contradicts the explicit premises, the conclusion need not follow. Now, at first glance, it seems reasonable to say that something like this law is relevant when interpreting text (2). If we have decided not to encode the temporal/causal relation between the two sentences as part of their semantics, then Law 2 might allow us to infer the relation. But this cannot be the whole story; if the semantics is neutral for the sentences in (2), then it should also be neutral for (1). But there is no piece of WK which will help us infer the temporal/causal structure relating the eventualities mentioned in (1). Something else must do the job. Intuitively, it could be a piece of knowledge about language use like the following:

- **Narration**
  A text segment $\alpha.\beta$ (where $\alpha$ and $\beta$ are sentences) that is temporally coherent is normally such that the eventuality described in $\alpha$ immediately precedes the eventuality described in $\beta$.

The relation "$e_1$ immediately precedes $e_2$" in the above is interpreted as "$e_1$ and $e_2$ stand in a causal or part/whole relation and $e_1$ temporally precedes $e_2$". If $e_1$ and $e_2$ are events, then $e_1$ temporally precedes $e_2$ (written $e_1 \prec e_2$) if $e_1$'s culmination occurs before $e_2$'s culmination. So there are part/whole relations between the eventualities $e_1$ and $e_2$ that are compatible with $e_1 \prec e_2$.

Although we call the above rule *Narration*, the title is perhaps unfortunate, since it would not be used only in the narrative genre, if such a genre could be stably defined. Equally, it is not meant to be "the" default, casting all text in the image of narrative, unless there is information to the contrary. It's just a piece of linguistic knowledge (LK) that may help us retrieve temporal structure, so long as we're prepared to assume that the text is temporally coherent.

Plausible reasons for proposing this piece of knowledge are threefold. It will—for some text genres at least—correspond to the Dowtian protagonist's "order of discovery" (cf. Dowty 1986). In such genres, a narrator typically describes events in the order in which the protagonist views them: the temporal order predicted by Narration. But more generally, it will sometimes be the case that the *only* information available to an interpreter regarding temporal structure will be textual order. Take a case in which no information about the temporal structure of events is derivable from WK or clue words like *because* or adverbials. Then the descriptive order of the events provides the only vital clue which can be used to construct the model of what the text describes. In such cases, our rule will be the only thing available. Arguably, in the more general version of our approach, these cases will indeed be rare, since the interpreter will have information available about discourse structure (cf. Lascarides and Asher 1991b). In any case, the rule provides a means for an interpreter to exploit a co-operative text-generator, who has observed Grice's (1975) maxim of manner, and chosen one of several ways to "be orderly".

In essence, this story about the rule of Narration suggests that the theory should represent Grice's pragmatic maxims as defeasible rules. Such an approach to pragmatics has been suggested in, for instance, Joshi, Webber and Weischedel (1984), and Wainer and Maida (1990).

# 5 The Context of Interpretation

The general proposal, then, is that we have a set of defeasible rules available, some expressing WK, others expressing LK. A defeasible logic will then apply these rules to semantic structures, and generate temporal/causal structures for a text. In order to procede to the details of the proposal, we need firstly to put in place some terminology. This should allow a precise, yet general, representation for defeasible rules, and secondly a more exact specification of what an interpreter brings to a text.

## 5.1 Representation of default rules

We do not intend to provide here a full account of the truth conditions of the defeasible rules. We instead wish to emphasise that the ideas presented here are neutral between specific formalisations of defeasible reasoning—so long as the logic can support the patterns we have suggested. However, we will need to assume a certain syntactic structure to the rules and show how the reasoning patterns introduced earlier are syntactically structured, in order to calculate which defeasible inferences underly temporal semantics.

We will represent the default condition as $>$, so $\phi > \psi$ is read as "Other things being equal, if $\phi$ then $\psi$". $\phi > \psi$ is the way defeasible rules are represented in MASH. But for the purposes of this paper, $\phi > \psi$ can be considered syntactic sugar for $\phi \wedge \neg B \neg \psi \rightarrow \psi$ where $B$ is a modal belief operator, which is how default rules are represented in HAEL; or $\phi \wedge M\psi \rightarrow \psi$ where $M$ is Reiter's consistency operator; or $\phi \rightarrow normally(\psi)$ which is how default rules are represented in Update Semantics (Veltman 1989). The defeasible conditional *If Tweety is a bird then Tweety flies, if things are normal* will be represented as $bird(tweety) > fly(tweety)$.[5]We will represent all indefeasible laws as modal conditionals of the form $\Box(\phi \rightarrow \psi)$. *Penguins are birds* will thus be represented as $\Box(\forall x)(penguin(x) \rightarrow bird(x))$ for example. Such conditionals have a well-defined semantics in MASH, HAEL and in Reiter's logic.

Furthermore, let $\langle \alpha, \beta \rangle$ mean that the text $\alpha.\beta$ is temporally coherent; see assumption (ii) in the next subsection. $\langle\alpha,\beta\rangle$ will entail that the events described in $\alpha$ and $\beta$ are causally or temporally related somehow. Finally, let $me(\alpha)$ be a term that refers to the main eventuality described by $\alpha$.[6] For example, the main eventuality for *Max fell* is the event $e$ of Max falling.

### 5.2 Assumptions about the knowledge base

The assumptions we make about the premises or knowledge base (KB) when interpreting text are the following:

(i) The semantics of the sentences $\alpha$ and $\beta$, which form part of the KB, do *not* impose any conditions on the temporal order of the events they describe.

(ii) The reader assumes that text is locally temporally coherent, and so $\langle\alpha,\beta\rangle$ is part of the KB.

(iii) All defeasible WK, such as the above causal law, and LK, such as Narration, is part of the KB.

(iv) All indefeasible laws, such as *Penguins are birds* and *Causes Precede Effects* are part of the KB.

(v) The laws of logic are part of the KB.

On our account, then, this is all that an interpreter brings to a discourse. Obviously, any claims to psychological reality would have to be taken with a large pinch of salt, if only in view of assumption (v). In the next section, we will show how a model making these assumptions about an interpreter's KB can get the right structures for reversed and non-reversed texts.

## 6 The Penguin Principle and reversals

We will consider in turn the first three pairs of two-sentence texts introduced at the beginning of the paper. The first pair both involve sentences describing events; the second pair both involve describing an event and a state; the third pair both involve events and states, where more than one causal law is brought into play.

### 6.1 Events and causation

So: how do we account for the pair (1) and (2), and the differences between them?

(1) Max stood up. John greeted him.

(2) Max fell. John pushed him.

The story about (1) turns out to be quite simple. The only rule in our KB when interpreting (1) whose antecedent is satisfied is the following one ($iprec(me(\alpha), me(\beta))$ stands for "$me(\alpha)$ immediately precedes $me(\beta)$"):

- **Narration**
  $\langle \alpha, \beta \rangle > iprec(me(\alpha), me(\beta))$

In the light of this rule, the KB gives the pattern of Defeasible Modus Ponens, as shown in Figure 2.

---

Insert Figure 2 about here

---

Defeasible Modus Ponens yields the conclusions from the premises that the main event described in $\alpha$, i.e. Max standing up, precedes the main event described in $\beta$, i.e. John greeting him. This agrees with intuitions. However, as the basic story suggested, the account of (2) is more complex, and it turns out to be characteristic of all the cases we have labelled "reversals". (2) looks like (1), except for the lexical variations. This time, however, Narration is not the only relevant rule; we also have Causal Law 2.

To represent it formally, we introduce a predicate $r$ such that $r(e_1, e_2)$ means that the events $e_1$ and $e_2$ are temporally or causally *connected* somehow. For our purposes, $r(e_1, e_2)$ can be viewed as asserting that one of a constrained list of possible event connections holds:

(a) $e_1$ causes $e_2$ or $e_2$ causes $e_1$.

(b) $e_1$ is part of the preparatory phase of $e_2$ or $e_2$ is part of the preparatory phase of $e_1$.

(c) $e_1$ is part of the consequent phase of $e_2$ or $e_2$ is part of the consequent phase of $e_1$.

(d) $e_1$ and $e_2$ temporally overlap.

(e) $e_1$ immediately precedes $e_2$ or $e_2$ immediately precedes $e_1$.

So $r(e_1, e_2)$ holds just in case $e_1$ and $e_2$ are connected by one of the relations in the set $C$ introduced earlier. We now represent Causal Law 2 as follows:

- **Causal Law 2**
  $(r(e_1, e_2) \wedge fall(max, e_1) \wedge push(john, max, e_2)) > cause(e_2, e_1)$

This is a relatively simple rule: if the three premises are satisfied by the KB including the semantic structures for the sentences of the discourse, then *ceteris paribus*, we will conclude that the pushing caused the falling. The limitation of the rule to describing Max and John is purely exegetical; it is important to note that even a more general version of the rule should *not* be read as claiming that pushings normally cause fallings. There will be plenty of pushings around that don't cause fallings; and there may well be plenty of fallings that cause pushings. The point is that *if* a pushing and falling are connected, *then* although both causal directions may be permissable, the KB normally prefers one to the other.[7]

The second and third premise of the antecedent of Causal Law 2 are verified by the KB when interpreting (2), because we can assume that the semantics of the two sentences $\alpha$ and $\beta$ entail $fall(max, me(\alpha))$ and $push(john, max, me(\beta))$. This would follow directly from the intuitive definition of the function $me$ and the following DRT-type logical forms for the sentences in (2).[8]

$(\alpha)$ $\quad [e_1, t_1][t_1 \prec now, hold(e_1, t_1), fall(max, e_1)]$

$(\beta)$ $\quad [e_2, t_2][t_2 \prec now, hold(e_2, t_2), push(john, max, e_2)]$

In words, $\alpha$ introduces an event $e_1$ and time $t_1$ where $t_1$ is earlier than the time of speech $now$, and the event $e_1$ of Max falling holds at the time $t_1$. $\beta$ is the same save that the event is John pushing Max. These logical forms impose no relations between $e_1$ and $e_2$.

But the question remains: is the *first* premise of Causal Law 2 satisfied by the KB? As things stand, we have not said whether the KB contains the assumption $r(e_1, e_2)$. However, we can show that in both MASH and HAEL it does; and hence that the premises of the rule will be satisfied; and hence that the conclusion follows, in the absence of information to the contrary.

The demonstration procedes as follows. Our definition of temporal coherence provides the key: it yields the following indefeasible laws:

- **Definition of Coherence**
  $\Box((\langle\alpha, \beta\rangle \vee \langle\beta, \alpha\rangle) \leftrightarrow r(me(\alpha), me(\beta)))$
  $\Box(\langle\alpha, \beta\rangle \rightarrow r(me(\alpha), me(\beta)))$
  $\Box(\langle\beta, \alpha\rangle \rightarrow r(me(\alpha), me(\beta)))$

These indefeasible laws allow one to substitute the antecedent of Causal Law 2 to yield the following as part of the KB:[9]

- **New Causal Law 2**
  $\langle\alpha, \beta\rangle \wedge fall(max, me(\alpha)) \wedge push(john, max, me(\beta)) > cause(me(\beta), me(\alpha))$

The antecedent of New Causal Law 2 is satisfied by the KB in interpreting (2). So is Narration's. The antecedent of New Causal Law 2 entails that of Narration (by $\Box((\phi \wedge \psi) \rightarrow \psi)$). Now, the conclusions of these laws conflict in the light of the indefeasible law that causes precede effects:

- **Causes Precede Effects**
  $\Box(\forall e_1 e_2)(cause(e_1, e_2) \rightarrow \neg e_2 \prec e_1)$

Thus the defeasible rules in the KB don't quite form a Penguin Principle: the conclusions of the conflicting laws are not $p$ and $\neg p$. Nevertheless a "Complex" Penguin Principle is formed. In this case, the laws conflict in the context of Causes Precede Effects and so more specific laws win, just as in the Penguin example: see Figure 3.

Insert Figure 3 about here

If the logic supports the Complex Penguin Principle, then the conclusion gained is that the pushing caused the falling, as required. [10]

A more direct route to the Complex Penguin would be to restate Causal Law 2 directly as New Causal Law 2. This would avoid having to *infer* New Causal Law 2 in order to obtain the Penguin Principle pattern of reasoning required. In essence, instead of characterising causal laws as representing only WK, we would represent causal laws as a *mixture* of WK and LK. In words, New Causal Law 2 asserts that if the sentences are temporally coherent then there is a connection between the events; and given the kinds of events they are, the second event described caused the first, if things are normal. For the sake of simplicity, we will represent the causal laws directly as a mixture of WK and LK, as in New Causal Law 2.

It must be stressed that the reasoning pattern, and the temporal structure associated with (2), are dependent on the reader's KB. Suppose we add to that KB a further premise, stating that the falling preceded the pushing. Then the pattern of inference is different from a Complex Penguin; the consequent of New Causal Law 2 is not inferred, because it is inconsistent with the facts already held in the KB. Instead, only one remaining defeasible rule has its antecedent verified: Narration. It's therefore used to infer that the falling immediately preceded the pushing. So (2) is not always interpreted as a reversal. Rather, *in the absence of information to the contrary*, (2) is interpreted as a reversal. By contrast, (1) is interpreted as a narrative, in the absence of information to the contrary.

In summary, then, the natural interpretations of (1) and (2) can be formally distinguished; this even though their sentences are assigned similar logical forms. The basis for the distinction was a defeasible causal law whose antecedent was verified by (2) but not by (1). Because of this, the premises in interpreting (1) and (2) formed different patterns of defeasible inference; Defeasible Modus Ponens for (1) and the Penguin Principle for (2).

### 6.2 States and causation

Now, it turns out that the Penguin Principle also plays a crucial role in event-state pairs. Take (3), where we have the intuition that there is no causal link between the event and state described.

(3) Max opened the door. The room was pitch dark.

Suppose the logical forms of the sentences in (3) are respectively $\alpha$ and $\beta$. Then the appropriate KB for the analysis of (3) contains $\alpha$, $\beta$, $\langle\alpha, \beta\rangle$ and all the defeasible and indefeasible laws mentioned above. In addition we claim that the following defeasible piece of LK holds and is therefore contained in the KB:

- **States Overlap**
  $\langle\alpha,\beta\rangle \wedge state(me(\beta)) > overlap(me(\alpha), me(\beta))$

In words, when a temporally coherent segment of text $\alpha.\beta$ describes a state and we have no knowledge about how that state is related to other eventualities—gained from WK or syntactic markers like *because* and *therefore*—we assume that they temporally overlap.

This law can be seen as a manifestation of Grice's Maxim of Relevance—as suggested in Lascarides (1990). To construct a full picture of the situation described by NL text, the interpreter must infer the relative occurrences of the states and the culminations of the events, including where the states start and stop. Since culminations are punctual (holding at points of time), their relative order is inferred from rules like Narration. This models the fact that an author can typically describe eventualities in the order of perception. But states are extended, and the order of perception of states does not fully determine where the states start relative to the situation being described. So simply relating events and states in the order in which the protagonist views them does not in itself determine where the state starts. There are several linguistic mechanisms which can be used to indicate where a state starts relative to the other eventualities. The author could explicitly refer in the text to what caused the state, and so from the law that causes precede effects, the interpreter will know the relative place where the state starts. This is what's going on in (4).

(4) Max switched off the light. The room was pitch dark.

Alternatively, the author can use temporal adverbials to say where a state starts. But (3) shows that these two mechanisms are not enough: how do we determine where states start in texts that do *not* feature adverbials or causes? States Overlap is then a vital mechanism for determining where a state starts relative to other eventualities described in text. It basically says that if there is no "explicit" indication of where the state starts—via the mention of causes or the use of temporal adverbials—then the start of the state is assumed to be irrelevant. That is, the state started to hold *before* the situation that the text is concerned with occurs, resulting in temporal overlap. States Overlap can be viewed as a manifestation of the Maxim of Relevance, because it asserts that (unless there is indication in the text to the contrary) the point where a state starts is assumed to be irrelevant.

We assume that the logical form of the second sentence in (3), namely $\beta$, entails $state(me(\beta))$ by the classification of the predicate *dark* as stative. So the KB in the analysis of (3) verifies the antecedents to two defeasible laws: Narration and States Overlap. Furthermore, these laws conflict in the context of the following axiom that asserts that temporal overlap and precedence are mutually exclusive:

- **Overlap/Precedence are Exclusive**
  $\Box(\forall e_1 e_2)(overlap(e_1, e_2) \rightarrow (\neg(e_1 \prec e_2) \wedge \neg(e_2 \prec e_1)))$

So the KB provides another instance of the Complex Penguin Principle, as shown in Figure 4.

---

Insert Figure 4 about here

The temporal structure inferred for (3) is that the event and state overlap, as required.

We now turn to the other case, where there is some causal link between the event and the state. This time, the natural temporal interpretation is rather different: the event *precedes* the state.

(4) a. Max switched off the light.

b. The room was pitch dark.

Causal Law 4 below reflects the knowledge that the room being dark and switching off the light, if connected, are normally such that the event causes the state:[11]

- **Causal Law 4**
  $\langle\alpha, \beta\rangle \wedge \mathit{switchoff}(max, light, me(\alpha)) \wedge dark(room, me(\beta)) > cause(me(\alpha), me(\beta))$

Suppose the logical forms of (4a) and (4b) are respectively $\alpha$ and $\beta$. Then the KB when analysing (4) contains $\alpha$, $\beta$, $\langle\alpha, \beta\rangle$, Causal Law 4, States Overlap and Narration. The antecedents to these three defeasible laws are all verified by the premises. Causal Law 4 conflicts with States Overlap, which in turn conflicts with Narration. Causal Law 4 does not, however, conflict with Narration. Moreover, the antecedent of Causal Law 4 entails that of States Overlap (by the stative classification of the predicate *dark*), which entails that of Narration. Pictorially, this is represented as in Figure 5.

Insert Figure 5 about here

It is apparent that the pattern involved is not quite a Complex Penguin. So it isn't possible to tell at a glance whether a logic for defeasible reasoning that supports the Penguin Principle will also support the inference required from the above pattern, where the event in (4) causes the state. But intuitively, the same principle that is exemplified by the Penguin Principle should apply in the above. That is, the most specific defeasible law should take priority over less specific ones. Lascarides & Asher (1991b) show that conflict between defeasible rules is resolvable in MASH when the antecedent of one rule entails that of all the others; the consequent of the most specific rule holds, as required. In HAEL, since Causal Law 4 is the most specific defeasible law, it would have the highest priority in the hierarchy. So once again the logic will ensure that the law's consequent—that the event caused the state—is inferred. Whether or not Narration also "fires" in the above will depend on the particular details of the HAEL being used.

### 6.3 Most specific cause

We have illustrated that the Penguin Principle can be used to "filter" pragmatic maxims and causal laws so that only the appropriate event structures—the ones inferred from the more specific antecedents—are inferred. In essence, this means that the reader never ignores information that is derivable from the text and relevant to constructing the relations between events. For example, if the antecedents of Narration and States Overlap are both verified by the KB, then the Penguin Principle ensures that the following information is not ignored: a stative expression was used instead of an event expression.

Texts (5) and (6) demonstrate that the Penguin Principle can also filter defeasible causal laws.

(5) a. Max took an aspirin.

b. He was sick.

(6) a. Max took an overdose of aspirin.

b. He was sick.

Suppose the logical forms of (5a) and (5b) are $\alpha_1$ and $\beta$ and the logical forms of (6a) and (6b) are $\alpha_2$ and $\beta$. We assume that $\alpha_1$ features the atomic formula $take(max, aspirin, e_1)$, $\beta$ features $sick(max, e_2)$, and $\alpha_2$ features the conjunction $take(max, aspirin, e_1) \wedge overdose(e_1)$. Thus the entailment from $\alpha_2$ to $\alpha_1$ is cashed out in the traditional Davidsonian way (cf. Davidson 1967).

Let's first consider the natural interpretation of (5). Causal Law 5 captures the knowledge that sickness and taking aspirin, if connected, are normally such that the sickness caused one to take aspirin.

- **Causal Law 5**
  $\langle \alpha, \beta \rangle \wedge take(aspirin, max, me(\alpha)) \wedge sick(max, me(\beta)) > cause(me(\beta), me(\alpha))$

The KB in the analysis of (5) contains $\alpha_1$, $\beta$ and $\langle \alpha_1, \beta \rangle$. Thus the antecedents to Narration, States Overlap and Causal Law 5 are all satisfied (assuming that *sick* is classified as stative). The antecedent to Causal Law 5 entails that of States Overlap, which entails that of Narration. Moreover, Causal Law 5 conflicts with Narration in the context of the indefeasible law that Causes Precede Effects. The premises thus form a pattern of defeasible reasoning that is very like that for text (4), and the same remarks about the expressiveness required from a default logic apply again. See Figure 6.

---

Insert Figure 6 about here

---

Now consider text (6). Again, the appropriate KB verifies the antecedents of Causal Law 5, States Overlap and Narration. The antecedent to Causal Law 6 is also verified:

- **Causal Law 6**
$\langle\alpha,\beta\rangle \wedge take(max, aspirin, me(\alpha)) \wedge overdose(me(\alpha)) \wedge sick(max, me(\beta))$
$\quad > cause(me(\alpha), me(\beta))$

Causal Law 6 reflects the defeasible knowledge that taking an overdose and being sick, if connected, are normally such that the overdose caused the sickness. The antecedent to Causal Law 6 entails that of Causal Law 5. Moreover, Causal Law 6 conflicts with Causal Law 5 in the context that Causes Precede Effects, and also conflicts with States Overlap. It does not, however, conflict with Narration. Thus the premises in the KB form the pattern in Figure 7.

---

Insert Figure 7 about here

---

Again, the general principle—specific defeasible laws override less specific ones—will yield the conclusion that we require, that the event of taking an overdose in (6) caused the sickness. So if the logic for defeasible reasoning captures this principle, then the account of temporal interpretation presented here will capture an intuition: a reader never ignores information that is salient in text when calculating the relations between the events described. In (6) this means that the reader doesn't ignore the crucial information that the taking of aspirin was an overdose.

# 7 The Nixon Diamond and Incoherence

## 7.1 Local incoherence

We now turn to the awkwardness of text (7), and compare it with the acceptable (8).

(7) ?Max won the race. He was home with the cup.

(8) Max won the race. He arrived home with the cup.

The interpretation of (8) is similar to that of (1); an instance of Defeasible Modus Ponens where the relevant defeasible law is Narration. The awkwardness of (7) can be explained as follows. The defeasible law below captures the WK that if Max wins the race and if Max is at home, then these events don't temporally overlap, if things are normal. In other words, it is unusual for the finish line of the race to be at the winner's house:

- **Law 7**
$win(max, race, e_1) \wedge athome(max, e_2) > \neg overlap(e_1, e_2)$

Note that Law 7 does not require the antecedent to assert that the event and state are connected. For the intuition Law 7 captures is that the event and state don't normally temporally overlap, regardless of whether they are connected or not.

The appropriate KB in the analysis of (7) satisfies Law 7, States Overlap and Narration. Moreover, the antecedent of Law 7 does *not* entail that of States Overlap or Narration, since it does not contain in the antecedent the formula $\langle \alpha, \beta \rangle$. But as we have mentioned, the antecedent of States Overlap entails that of Narration. Law 7 conflicts with States Overlap. So the premises in the KB form the pattern in Figure 8.

---

Insert Figure 8 about here

---

By the general principle that more specific overrides less specific, the rule of Narration is deemed irrelevant here; States Overlap is more specific. So the example is like the Nixon Diamond: there is irresolvable conflict between States Overlap and Law 7. Hence as in the Nixon Diamond, we *fail* to conclude from the KB that $iprec(me(\alpha), me(\beta))$, $overlap(me(\alpha), me(\beta))$ or that $\neg overlap(me(\alpha), me(\beta))$. So thanks to the Nixon Diamond, the KB supports $\langle \alpha, \beta \rangle$ but *fails* to support any relation between $me(\alpha)$ and $me(\beta)$. We assume an axiom of consistency on KBs which forms a new assumption (vi) on the structure of KBs given in section 3:

(vi) Any KB that verifies $\langle \alpha, \beta \rangle$ and fails to verify $rel(me(\alpha), me(\beta))$ for any relation $rel$ in the set of event connections (listed in section 4) is inconsistent.

We reject all inconsistent KBs. And if a text cannot be processed so as to yield consistent KBs, then the text is incoherent, i.e. the assumption that $\langle \alpha, \beta \rangle$ is part of the KB must be dropped. Thus (7) is incoherent.

## 7.2 Incoherence and discourse popping

We have argued that a text is locally temporally incoherent just in case the logic for defeasible reasoning refuses to draw a conclusion about temporal structure, once it has run into an irresolvable conflict of the Nixon Diamond variety. Lascarides & Asher (1991a,1991b) exploit this approach to local temporal incoherence in their account of discourse popping. In a text such as (15),

(15)
a. Guy experienced a lovely evening last night.
b. He had a great meal.
c. He ate salmon.
d. He devoured lots of cheese.
e. He won a dancing competition.

one attempts to attach the sentence being processed to the text that has been processed so far. In our current terms, the method they use is to check whether the sentence currently being processed forms a (locally) temporally coherent text segment with the previous sentence. So upon processing (15e), the KB contains the assumption that (15d,e) forms a coherent text segment. With the relevant defeasible WK and LK the KB runs into an irresolvable conflict of the Nixon Diamond variety just as we did with text (7). From this, the conclusion drawn is that (15d,e) does *not* form a coherent text segment. Thus (15e) must attach to one of the remaining sentences in (15d) which are open to clause attachment, which in the above case is defined as (15a) and (15b). Thus the Nixon Diamond is shown to provide an account of local temporal incoherence, and local temporal incoherence in turn provides the key to discourse *popping*.

# 8 Dudley Doorite, Weakening and Temporal Ambiguity

We have shown that defeasible WK and LK can be used to infer temporal structure. A corollary of the approach is that interpretation is always carried out relative to the reader's KB which, among other things, contains facts and laws that define the context in which the text is uttered. So interpretation in this account is reader-specific and context-specific.

This has significant implications for the question of textual ambiguity. The assumptions we've made on the KB entail that as long as a text is coherent, it is also temporally *precise*. That is, whenever the reader infers that one of a proper subset of the relations in the set $C$ hold between the eventualities described, then there is always sufficient information to choose *which* among these plausible alternative temporal structures holds. This is in contrast to the Nixon Diamond of irresolvable conflict, where no proper subset of relations in $C$ could be inferred in the first place.

To illustrate this point, consider the following; in the realm of thermostats, a bimetallic strip's changing shape and a temperature fall are causally connected, but there is an ambiguity over the direction of the causal relation. Even so, text (9) on our account is not interpreted as ambiguous.

(9) The bimetallic strip changed shape. The temperature fell.

Because of the *order* in which the events are described the reader can choose between the two alternative causal relations (that the bimetallic strip's change of shape caused the temperature drop, or *vice versa*). In fact, as Herb Clark has suggested (personal communication), the reader infers that the descriptive order in (9) matches temporal order, unless there are contrary facts already known to the reader.

This interpretation of (9) is explained as follows: the appropriate KB contains these laws:

- **Change = Bending or Straightening**
  $change(strip, e) \leftrightarrow (bend(strip, e) \vee straighten(strip, e))$
- **Causal Law 9a**
  $\langle \alpha, \beta \rangle \wedge bend(strip, me(\alpha)) \wedge fall(temperature, me(\beta)) > cause(me(\alpha), me(\beta))$
- **Causal Law 9b**
  $\langle \alpha, \beta \rangle \wedge straighten(strip, me(\alpha)) \wedge fall(temperature, me(\beta)) > cause(me(\beta), me(\alpha))$

The antecedents of Law 9a and 9b are *not* satisfied by the KB in the analysis of (9). However, Weakening the Consequent and Dudley Doorite can be used to infer the event structure for (9): Weakening yields Causal Laws 9a$'$ and 9b$'$ from 9a and 9b:

- **Causal Law 9a$'$**
  $\langle \alpha, \beta \rangle \wedge bend(strip, me(\alpha)) \wedge fall(temperature, me(\beta))$
  $\qquad > cause(me(\alpha), me(\beta)) \vee cause(me(\beta), me(\alpha))$
- **Causal Law 9b$'$**
  $\langle \alpha, \beta \rangle \wedge straighten(strip, me(\alpha)) \wedge fall(temperature, me(\beta))$
  $\qquad > cause(me(\alpha), me(\beta)) \vee cause(me(\beta), me(\alpha))$

Laws 9a$'$ and 9b$'$ then form the premises for Dudley Doorite, and using the law that a strip changing shape is either a bending or a straightening, we obtain Law 9c:

- **Causal Law 9c**
  $\langle \alpha, \beta \rangle \wedge change(strip, me(\alpha)) \wedge fall(temperature, me(\beta))$
  $\qquad > cause(me(\alpha), me(\beta)) \vee cause(me(\beta), me(\alpha))$

The KB satisfies the antecedents to Law 9c and Narration and the former antecedent entails the latter. But the consequents of Narration and Law 9c *don't* conflict. As a result, Defeasible Modus Ponens on Causal Law 9c and on Narration yield the following:

(16) $\quad cause(me(\alpha), me(\beta)) \vee cause(me(\beta), me(\alpha))$

(17) $\quad iprec(me(\alpha), me(\beta))$

(16) and (17) together entail $cause(me(\alpha), me(\beta))$; i.e. the strip changing shape caused the temperature fall. So the temporal structure of (9) inferred is not ambiguous even if the reader has no information prior to interpreting (9) about the direction of the causal relation.

A similar story holds for texts that describe an event and then a state. Suppose one knows from WK that the event and state are causally connected but one doesn't know the direction of the causal relation. The consequent of States Overlap is compatible with the state causing the event, but not with the event causing the state. As a result, a similar pattern of reasoning (that is: Defeasible Modus Ponens on States Overlap and a Causal Law that follows from Dudley Doorite and Weakening) yields the inference that the state caused the event.

Taken together with the assumptions in section 4 about the contents of the reader's KB, this indicates that a KB will always be such that if the reader infers a plausible set of alternative temporal structures for a text, she can always infer *which* of these alternatives holds. Crucially, information about textual order is used to choose among these alternatives, via rules like Narration and States Overlap. On this theory, there is no ambiguity at the level of interpretation relative to the reader's KB. Ambiguity does arise, however, at another level. As we've mentioned, a text like (2) can have different interpretations with respect to different KBs.

(2) Max fell. John pushed him.

We showed that (2) is interpreted as a reversal in the absence of information to the contrary, but as a narrative if the KB *does* contain (additional) information to the contrary. This is a consequence of using non-monotonic inference: as context is enriched, the temporal structure inferred relative to the context may change. So a text is ambiguous in that it receives different interpretations with respect to different contexts. In that sense, all the texts we've mentioned are ambiguous.

## 9 Defeasible Knowledge and Semantic Structure

We have deliberately left open the question of how the inference regime presented here affects the semantic structure of text. This is because at least three options regarding the semantic structure of text are available, given that we have imposed conditions so far only on the semantic structure of *sentences*.

The first choice is that the semantic structure of text is *neutral* about the temporal order of the events it describes, and the defeasible inferences presented above provide the means of explaining why the reader infers more than textual semantic structure alone would warrant. Under this proposal, the logical form of the text is built from the syntactic structure of its sentences, and the representation of the text $\alpha.\beta$ would be something like (18):

(18) $\alpha \wedge \beta \wedge \langle \alpha, \beta \rangle$

(18) imposes no conditions on the relations between $me(\alpha)$ and $me(\beta)$, given the properties of $\alpha$ and $\beta$ we have already mentioned. These relations would be worked out using the reader's knowledge base and defeasible inference. One can view this option as keeping semantic structure as neutral as possible, and letting pragmatics, at a "higher" level of interpretation, do a lot of the work, but not at the sacrifice of formality. The temporal structure derived from (18) is dependent on the KB being considered, and is therefore reader-specific and context-specific.

The second option also views the semantic structure of text as being built from the syntactic structure of its sentences. But this time, each text has several logical forms, one corresponding to each of the possible event connections given the syntactic structure of the sentences concerned. Under this option, the inference regime presented here can be viewed as a way of choosing among these logical forms, one of which corresponds to the reader's preferred interpretation, given what she knows. In this way, the current work can be taken to provide for Hobbs (1985:XXX) the required formal mechanism for deriving the "best" interpretation from the set of plausible interpretations. Once again, the choice of which among a large set of possible interpretations is preferred is reader-specific and context-specific.

A third option is adopted in Lascarides & Asher (1991a,1991b): the logical form of text is constructed dynamically. This is formalised as follows: we can view the logical form of text as a pair of sets $\{\phi_1, \phi_2\}$, where $\phi_1$ is the set of logical forms of sentences and $\phi_2$ a set of relations between the eventualities introduced in $\phi_1$. When incorporating the second sentence $\beta$ of a text $\alpha.\beta$ into $\{\{\alpha\}, \emptyset\}$—the logical form of the text so far—$\beta$ is added to the first set, and defeasible reasoning as presented in this paper is used to calculate the relation to be added to the second set. Thus the logical form of (1) would be $\{\{\alpha, \beta\}, \{iprec(me(\alpha), me(\beta))\}\}$

and the logical form of (2) would be $\{\{\alpha, \beta\}, \{cause(me(\alpha), me(\beta))\}\}$; they are different in spite of their similar syntax. This view of dynamically constructing the logical form of text is explored extensively in Asher (forthcoming), in which $\phi_2$ contains discourse relations rather than temporal relations.

This third option is different from the first two in at least one crucial respect. The relationship between logical form and syntax is not uniform, as it was with the first two options. Under those options, the defeasible logic is a pragmatic mechanism which either augments the semantic structures passed to it, or chooses between them. But under the third option, the semantic structure of text itself is dependent on the reader's knowledge, making the semantic structure of text reader-specific. This contrasts with the way logical forms are constructed in Kamp & Rohrer (1983), Partee (1984) and Hinrichs (1986), where the logical form for the *whole* text is built from syntax, before any interpretation occurs. Our discussion here has been neutral; all three options are possible ways of applying the defeasible strategy outlined in this paper. Our proposal thus opens up a choice as to where to place the barrier between semantics and pragmatics. But wherever it goes, formal precision isn't sacrificed.

## 10 Defeasible Knowledge in Generation

We have examined at length the role defeasible knowledge—about language and the world—may play in interpretation. Such knowledge can also play a role in generation. Recall the second question posed at the beginning of this paper. Given the causal and temporal relations between events in a knowledge base, what are the ways they can be described in text? In fact, there are many ways of describing complex temporal structures, and Joshi, Webber and Weischedel (1984) exploit default reasoning to choose between candidate descriptions. A speaker $S$ can generate a set of possible utterances, and then by using her knowledge of her hearer $H$, eliminate utterances where $S$ has reason to believe $H$ will draw a conclusion $S$ knows to be false. So if $S$ believes that $H$ lacks some defeasible knowledge, $S$ will eliminate utterances whose coherence, reliability or precision relied on the use of it.

In the larger context of discourse structure, Oberlander & Lascarides (1991) discuss various ramifications of this approach to interactive defaults. Here, let us just note the following consequences. Clashing defaults lead to problems about coherence; *missing* defaults lead to problems about reliability and hence precision.

Take the old case where John's pushing Max caused the latter to fall. Suppose $S$ has a KB which will allow her to generate (2). This text is coherent, precise and reliable for $S$ because the causal law (about the usual causal relation between pushings and fallings) is more specific than the linguistic rule (about Narration). Suppose $S$ knew $H$ to lack the appropriate causal information. (2) will trigger a different inference pattern in $H$; one in which Narration wins after all. $S$ must block this by changing the utterance. There are two basic options. If clause order is kept fixed, then $S$ could shift tense into the pluperfect, as in (19). Alternately $S$ can insert a *clue word*, such as *because*, into the surface form, and thereby replace (2) with (20). Such a tactic relies on mutual knowledge of a specific linguistic rule about *because*.

(2) Max fell. John pushed him.

(19) Max fell. John had pushed him.

(20) Max fell because John pushed him.

On the other hand, if clause order is not taken to be fixed, then $S$ can simply *reorder* the sentences in (2), and let Narration do the rest. However, there will be problems with cases involving state-descriptions. In the absence of appropriate causal laws, temporal overlap will be predicted in cases where it should not. A solution would be to replace the state expression with an event expression.

Where clause reordering is permitted, the lack of causal information will thus have (possibly surprising) ramifications. Here, we end up restricting $S$ to using event expressions only. However, rather than accept these restrictions, an obvious move is to introduce further clue words, and appropriate linguistic rules for reasoning about them. This means exploiting LK to overcome the gaps in WK. If $S$ has reason to believe $H$ lacks relevant WK, but believes $H$ to possess appropriate LK, then $S$ will shift in the ways specified towards exploiting the latter in her utterances. This may help explain the proviso made when we first introduced the data about reversals. Recall the observation that texts which describe events in reverse to temporal order, without marking the reverse, may be quite rare. It's easy enough to interpret such texts, when we have the appropriate WK. But if a considerate speaker or writer has reason to believe that some or all of her audience lacks that WK, then she will either avoid such descriptive reversals entirely, or mark them with the type of clues we have discussed.

## 11 Conclusion

Defeasible reasoning is a useful inference regime. It can be applied to NLP via causal laws and pragmatic maxims; such knowledge will be useful so long as inference is formally characterised. The logic can help characterise temporal coherence and reliability, and explain at least two aspects of connected temporal discourse. These are reversals (or the lack of them), explained via the Penguin Principle; and incoherence, explained via the Nixon Diamond. As part of a more general theory of discourse structure, such patterns of defeasible inference can therefore constitute elements in an account of interpretation, and they should also feature in a theory of text generation based on the strategy of interactive defaults.

## Footnotes

1. Later, we shall cash out the part/whole relation in terms of Moens & Steedman's (1988) event ontology, where an event is associated with three parts: a preparatory phase, a culmination and a consequent phase. So "$e_1$ is part of $e_2$" means that "$e_1$ is part of the preparatory phase or consequent phase of $e_2$".

2. We assume that an event $e_1$ precedes an event $e_2$ if $e_1$'s *culmination* occurs before $e_2$'s. So there are part/whole relations between $e_1$ and $e_2$ that are compatible with $e_1$ temporally preceding $e_2$.

3. Note that our use of these terms therefore differs from that in Hobbs (1985), where the local coherence of an utterance is determined by its relation to the whole surrounding discourse, and its global coherence by the utterance's relations to discourse, conversational plans, and world knowledge.

4. Herb Clark has pursued this point in conversation.

5. For the sake of simplicity we will not be concerned with the way quantifiers are represented in the defeasible rules; these have no bearing on the defeasible rules for temporal interpretation since they will be propositional.

6. $me(\alpha)$ is formally defined in Lascarides & Asher (1991b) in a way that agrees with intuitions.

7. This is not the place to enter into the metaphysics of causality; we are only trying to represent a language user's causal preferences. However, at least some philosophical accounts suggest that $A$s cause $B$s when $A$s probabilise $B$s. In particular, when an $A$ does not occur, the probability of $B$ is lower than when an $A$ does occur. These are by no means the only views in the literature, but there is much to commend them. Suffice it for now that our representation of a language user's beliefs about causal relations is compatible with such accounts.

8. These are the forms adopted in Lascarides (1990) and Lascarides & Asher (1991a,1991b). Note that the logical forms of (1's sentences would be the same, save that $fall$ and *push* are replaced by *standup* and $greet$. We are not here concerned with committing ourselves to a specific logical form of sentences like *Max fell*. The examples are simply meant to illustrate that logical forms with the required properties are available.

9. Similarly, there is also a rule $\langle\beta, \alpha\rangle \wedge fall(max, me(\beta)) \wedge push(john, max, me(\alpha)) > cause(me(\beta), me(\alpha))$ indicating that *regardless* of the order in which the pushing and falling are described, the defeasible preference is that the pushing caused the falling.

10. Lascarides & Asher (1991b) show that MASH validates the Complex Penguin Principle. HAEL also does so, since the default rules will be ordered in the hierarchy according to the specificity of their antecedents.

11. For the sake of simplicity, we ignore the problem of inferring that the light is in the room.

## References


**Appelt, D. and Konolige, K. [1988]** A practical nonmonotonic theory for reasoning about speech acts. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, State University of New York at Buffalo, Buffalo, N.Y., 7–10 June, 1988, pp170–178.

**Aristotle [1984]** *The Complete Works of Aristotle: the revised Oxford translation.* Barnes, J. (ed.). Princeton: Princeton University Press.

**Asher, N. [forthcoming]** *Reference to Abstract Objects in English: A Philosophical Semantics for Natural Language Metaphysics.*

**Asher, N. & Morreau, M. [1991]** Common Sense Entailment: A Modal Theory of Nonmonotonic Reasoning. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence*, Sydney Australia, August, 1991.

**Bach, E. [1986]** The algebra of events. *Linguistics and Philosophy*, **9**, 5–16.

**Caenepeel, M. [1991]** Event Structure versus Discourse Structure. In *Extended Abstracts for the* DANDI *Workshop on Discourse Coherence*, Edinburgh, April, 1991. Available from Human Communication Research Centre, University of Edinburgh.

**Crouch, R. and Poznanski, V. [1991]** Factual and Counterfactual *Before*. Unpublished ms, available from the Computer Laboratory, University of Cambridge.

**Dahlgren, K. [1988]** *Naive Semantics for Natural Language Understanding.* Norwell, Mass.: Kluwer.

**Davidson, D. [1967]** The logical form of action sentences. In Rescher, N. (ed.) *The Logic of Decision and Action.* Pittsburgh: University of Pittsburgh Press.

**Dowty, D. [1986]** The effects of aspectual class on the temporal structure of discourse: semantics or pragmatics? *Linguistics and Philosophy*, **9**, 37–61.

**Grice, H. P. [1975]** Logic and Conversation. In Cole, P. and Morgan, J. L. (eds.) *Syntax and Semantics*, Volume 3: *Speech Acts*, pp41–58. New York: Academic Press.

**Grosz, B. J. and Sidner, C. L. [1986]** Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, **12**, 175–204.

**Hobbs, J. R. [1979]** Coherence and Coreference. *Cognitive Science*, **3**, 67–90.

**Hobbs, J. R. [1985]** On the Coherence and Structure of Discourse. Research report No. CSLI–85–37, Centre for the Study of Language and Information, Stanford, Ca., October, 1985.

**Hinrichs, E. [1986]** Temporal anaphora in discourses of English. *Linguistics and Philosophy*, **9**, 63–82.

**Joshi, A., Webber, B. and Weischedel, R. [1984]** Default reasoning in interaction. In *Proceedings of the Non-Monotonic Reasoning Workshop*, AAAI, New York, October, 1984, pp144–150.

**Kamp, H. [1981]** A theory of truth and semantic representation. In Groenendijk, J. A. G., Janssen, T. M. V. and Stokhof, M. B. J. (eds.) *Formal Methods in the Study of Language*, Volume 136, pp277–322. Amsterdam: Mathematical Centre Tracts.

**Kamp, H. and Rohrer, C. [1983]** Tense in texts. In Bauerle, R., Schwarze, C. and von Stechow, A. (eds.) *Meaning, Use and Interpretation of Language*, pp250–269. Berlin: de Gruyter.

**Lascarides, A. [1990]** Knowledge, Causality and Temporal Representation. Research Report No. HCRC/RP–8, Human Communication Research Centre, University of Edinburgh, Edinburgh, June, 1990. Revised version submitted to *Linguistics and Philosophy*.

**Lascarides, A. and Asher, N. [1991a]** Discourse Relations and Defeasible Knowledge. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, Berkeley, Ca., 18–21 June, 1991, pp55–62.

**Lascarides, A. and Asher, N. [1991b]** Discourse Relations and Common Sense Entailment. DYANA deliverable 2.5b, available from Centre for Cognitive Science, University of Edinburgh. Submitted to *Journal of Logic, Language and Information*.

**Levesque, H. J. [1990]** All I Know: A Study in Autoepistemic Logic. *Artificial Intelligence*, **42**, 263–309.

**Mann, W. C. and Thompson, S. A. [1987]** Antithesis: A Study in Clause Combining and Discourse Structure. Technical Report No. RS–87–171, USC Information Sciences Institute, Marina Del Rey, Ca., April, 1987.

**Moens, M. [1987]** Tense, Aspect and Temporal Reference. PhD Thesis, Centre for Cognitive Science, University of Edinburgh.

**Moens, M. and Steedman, M. J. [1988]** Temporal ontology and temporal reference. *Computational Linguistics*, **14**, 15–28.

**Moore, R. C. [1985]** Possible-world Semantics for Autoepistemic Logic. Report No. CSLI–85–41, Center for the Study of Language and Information, December, 1985.

**Oberlander, J. & Lascarides, A. [1991]** Generating Discourse with Defeasible Knowledge. To be presented at American Association for Artificial Intelligence Fall Symposium on Discourse Structure in Natural Language Understanding and Generation, Asilomar, Ca., November, 1991.

**Partee, B. H. [1984]** Nominal and temporal anaphora. *Linguistics and Philosophy*, **7**, 243–286.

**Reiter, R. [1980]** A logic for default reasoning. *Artificial Intelligence*, **13**, 81–132.

**Scha, R. and Polanyi, L. [1988]** An augmented context free grammar. In *Proceedings of the 12th International Conference on Computational Linguistics and the 24th Annual Meeting of the Association for Computational Linguistics*, Budapest, Hungary, 22–27 August, 1988, pp573–577.

**Veltman, F. [1989]** Defaults in Update Semantics. ITLI-Prepublication LP–89–09, University of Amsterdam, 1989.

**Wainer, J. and Maida, A. [1990]** Good and bad news in formalizing generalized implicatures. In *Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society*, Berkeley, Ca., 16–19 February, 1990, pp530–540.

**Webber, B. L. [1988]** Tense as discourse anaphor. *Computational Linguistics*, **14**, 61–73.

**Webber, B. [1991]** Structure and Ostension in the Interpretation of Discourse Deixis. **Language and Cognitive Processes**, **6**, 107–135.

Figure 1: A Pictorial Representation of Defeasible Inference

## Figures

Figure 2: Narration

Figure 3: Explanation

## Figure Legends

1. A Pictorial Representation of Defeasible Inference
2. Narration
3. Explanation
4. States Without Causation
5. States With Causation
6. Aspirin
7. Overdose
8. Incoherence

Figure 4: States Without Causation

Figure 5: States With Causation

Figure 6: Aspirin

Figure 7: Overdose

Figure 8: Incoherence