Chapter 14

A WBAnalysis

The incident that forms our running example occurred on 5 September 1995 at Brussels Airport (BRU). As reported in our primary source of information in Section 13.1.1, a DC10, Northwest Airlines Flight 52, on its way to Frankfurt (FRA) from Detroit, landed instead at BRU, much to everyone's surprise – except for the passengers and cabin crew, who watched the route on the cabin video, and air traffic control (ATC), who thought it was supposed to be going there.

The history included in the source article is simple, and we shall extend it to a rigorous analysis of the incident which will demonstrate what has to be determined to conclude where the errors occurred. (We shall see that there is indeterminacy in the reports we work from, which is also to be expected).

14.1 The Ontology

Our aim is to distinguish and represent every significant factor contributing to the final incident and their causal and explanatory interrelations. We distinguish four types of factors [LL98]:

1. States are individuated by the collection of state predicates which are true in that state. So states might be interpreted as types, because a given collection of state predicates might be true at two or more different moments. That is, the state occurs more than once in time. Furthermore states may have a duration, i.e. the values of the state predicates considered do not change over a period of time. Notice that one state might include another, if the collection of state predicates true in the first includes the collection true in the second. Along with states, we have therefore occurrences or instances of states - particular occasions at which the system or world is in that state.

Notation: $\langle \mathbf{A} \rangle$

2. Events are particulars, representing specific changes in state. However, events usually belong to event types, and the description of an event could be given by describing its type, coupled with an indication of when it occurred. Actions are types of event – events that fall under some common description

Notation: [A]

3. Processes (as described in [Lew86]) are defined as state/event mixes of bounded duration which describe undifferentiated actions. When everything is running normally, we might want to use a process to characterise it, rather than decompose it unnecessarily into events and states.

Notation: $\{A\}$

4. Non-events are states corresponding to the non-occurrence of awaited events (see below). Since events (more properly, event types) are characterised by descriptions, a state predicate corresponding to that non-event can be formed by negating the description.

Notation: (A)

We call all these objects "nodes". This ontology is justified further in [Lad96a, The Formal Semantics] and [GLL97b], and actually originates with Mill [Mil73a], although we arrived at it in other ways. Put briefly: events induce change of state; a change of state can be described by pointing out the relevant relation between the before-state and the after-state; a binary relation between states is a set of pairs of states according to the standard mathematical notion; an individual event is one of these pairs, but since the events have to be described, an event description will single out potentially many individual events, namely those that satisfy the description. These are the actions. The ontology we have outlined corresponds roughly to the USAF's act, omission, condition, circumstance (events or processes, non-events, state predicates respectively).

The distinction between which processes to analyse further into components (so: sequences of events and states) and which processes to leave 'undifferentiated' depends upon the purpose of the analysis. A procedure which takes place over an interval of time, but which is executed correctly, is likely to be of less interest to an accident investigator than a process executed in a faulty manner. One is likely to want to analyse the latter further - to decompose it into component state-event sequences. For example, an appropriate pilot reaction to an aircraft change-of-state can be considered to be an undifferentiated event or process; whereas a faulty pilot reaction can be considered in terms of the PARDIA classification - the chain of perception-attention-reasoning-decision-intention-action (Chapter 18) which can be used to indicate at which precise stage some given error was made.

Much depends on the level of analysis chosen. A procedure or the behavior of a system can be presented simply in terms of the major features and how they change (for example, representing a communication channel as a lossless FIFO buffer), or in great detail in terms of its implementation (for example, a sliding-window protocol to ensure reliable communication – that is, to simulate a lossless FIFO buffer – over an unreliable medium). The level of detail chosen will depend on the goal of the analysis. Again, a subsystem or procedure which operates successfully may be described often at the high level, but a faulty system may require much more detailed specification of its implementation and design in order to pinpoint a fault.

14.2 First, Determining a Temporal Succession

We commence by extracting step by step the information about what happened and why from the source text. We need a *history*, a statement of temporal succession on which we shall base our inquiry. (The source text may not use temporal order - histories make free use of verbal aspect and other devices to establish the temporal order of nodes.) Such histories are a preferred way to relate the story of an accident without attributing causality (see Section 13.2 for examples of such précis). There's no algorithm here – it's just look and see. WBA has methods for refining the initial 'guess', as we shall see.

The Aircraft (AC) crosses Shannon ATC (SATC) region and London ATC (LATC) region, finally reaching Brussels ATC (BATC) and landing at BRU runway (RWY) 25. All these pieces of information are mentioned in [dWL95]. Thus at the top level of our investigation we may define four nodes:

- [1] AC lands at BRU RWY 25
- $\langle 2 \rangle$ AC in BATC area
- $\langle 3 \rangle$ AC in LATC area
- $\langle 4 \rangle$ AC in SATC area

The aircraft passes through these nodes in a determinate temporal order:

$$\langle 4 \rangle \hookrightarrow \langle 3 \rangle \hookrightarrow \langle 2 \rangle \hookrightarrow [1]$$
 (14.1)

A word about time and temporal extent. We assume a linear temporal succession, composed of points. Nodes may have (usually will have!) temporal extent or duration. For those which occur 'at an instant' (which do not have temporal extent), an assertion $A \hookrightarrow B$, for two instantaneous states, events, or non-events A and B (processes always have temporal extent) means that A and B both occurred and that B succeeds or is contemporaneous with A. If A and B have temporal extent (thus could be processes also), an assertion $A \hookrightarrow B$ means that there is at least one temporal part of A which precedes all temporal parts of B.

The connective \hookrightarrow plays only a minor inferential role in a WBA, so we feel no need to be more precise about the notions of 'temporal part' and 'precedence'.

Nodes without temporal extent will usually occur during description of the behavior of some physical device which requires some physics to explain. For example, an aircraft passing a particular spatial point in cruise; the speed of an accelerating aircraft reaching (and passing) a particular value (an event ensured by, say, the Mean Value Theorem of Calculus); a climbing aircraft passing through a definite altitude. Examples of nodes with temporal extent are specific actions of agents, such as pilots: flipping a switch, changing the dynamic state of the aircraft from cruise to descent; and of course also many state predicates. The temporal extent of a node may or may not be consequential: when there are no concurrent or intervening significant events, a node with temporal extent may often be treated as though it had none; as if it were *atomic*, temporally non-decomposed.

For an axiomatic explanation of one version of the \hookrightarrow relation, see [Lam86]. We, however, shall only use \hookrightarrow to state top-level theorems to be proved; it suffers an immediate attempted reduction to \Longrightarrow^* in WBA. Where the reduction cannot be made, one may conclude that the history afforded no clues to the causal history, and the analysis must remain incomplete. The connection between causality and temporal succession (also considering temporally-extended nodes) may be made through Mill's observation that:

Whether the cause and its effect be necessarily successive or not, the beginning of a phenomenon is what implies a cause, and causation is the law of the succession of phenomena[...]

Whether the effect coincides in point of time with, or immediately follows, the hindmost of its conditions, is immaterial. At all events it does not precede it; and when we are in doubt, between two co-existent phenomena, which is cause and which is effect, we rightly deem the question solved if we can ascertain which of them preceded the other. [Mil73a, Chap. V, Book III, Volume 7]

(We are not considering the viability of backwards causation in this work.)

14.3 Rules for Causality

By introducing new facts and perhaps assumptions not mentioned particularly in the sources, we are now going to improve (14.1) successively. We first try to replace " \leftrightarrow " by causal chains between its arguments. According to Mill's explanation of causality, temporal succession is (at least) a hint towards causality (sometimes taken as more than a hint – see Hume, Mill [Mil73a], Moffett *et. al.* [MHCM96] and Johnson [TJ96] as criticised in [Lad96b]). We formulate this insight as the following axioms:

Axiom 1
$$\vdash$$
 $(A \Longrightarrow^* B) \Rightarrow (A \hookrightarrow B)$

An n-fold succession of causal-factor relations between chained factors implies a temporal succession between (at least) the first and the last factor of this chain.

The relation \Rightarrow^* of causality is intended to denote the *transitive closure* of \Rightarrow , 'causal factor of', the primary causal operator advocated by Lewis [Lew73a] and defined by him in terms of the counterfactual conditional from [Lew73b]. Although one cannot define completely the transitive closure of a relation from that relation in first-order logic, one can nevertheless axiomatise it soundly. One simply uses the 'standard' recursive definition (corresponding to Axioms 2, 3 and 4 below), and since proving is a finitary activity, the things one can prove about the transitive closure are exactly those which follow from a finite number, say n, of proof steps, which are exactly those that follow from an n-fold succession of causal-factor relations. Rather than use the recursive definition, we can axiomatise the relation \Rightarrow^* for practical purposes by asserting that it is an extension of \Rightarrow and that it is transitive (Axioms 2 and 5).

Axiom 2
$$\vdash$$
 $(A \Rightarrow B) \Rightarrow (A \Rightarrow *B)$

Axiom 3
$$\vdash$$
 $(A \Longrightarrow^* B) \land (B \Longrightarrow C) \Rightarrow (A \Longrightarrow^* C)$

Axiom 4
$$\vdash$$
 $(A \Rightarrow B) \land (B \Rightarrow^* C) \Rightarrow (A \Rightarrow^* C)$

Axiom 5
$$\vdash$$
 $(A \Longrightarrow^* B) \land (B \Longrightarrow^* C) \Rightarrow (A \Longrightarrow^* C)$

The relation \Rightarrow * is, according to Lewis, true causality. Using *modus ponens*, Axiom 1 leads to a derived inference rule we can use in analysis:

$$\frac{A \Rightarrow^* B}{A \hookrightarrow B} \tag{14.2}$$

Our investigation method attempts to reverse this order, like proof search, by working from the conclusion as true, and searching out a causal hypothesis involving \Rightarrow^* from which it may be derived. Not all temporal succession will be causal, so which instances of \hookrightarrow to choose to explain as causal is up to the investigator, as we saw with Wood and Sweginnis's example in Section 12.1.

The analysis proceeds as follows. We want to demonstrate (14.1) and must use Inference Rule 14.2, that is, we must try to establish a causal chain between each pair of temporally-successive nodes in 14.1. We may fail, because there are temporally successive nodes that are simply not causally related in any sense

meaningful to us (people will be born after we die: unless they are our descendants, we would not necessarily want to assert a causal connection). If we fail to establish a causal connection in a particular analysis, then that could be taken as reason to focus on different temporally-successive nodes to start our analysis.

Similarly to Axiom 1 and Rule 14.2, we can formulate derived inference rules from *modus ponens* and Axioms 2 and 5:

$$\frac{A \Rightarrow B}{A \Rightarrow^* B} \tag{14.3}$$

$$\begin{array}{ccc}
A \Longrightarrow^* B \\
B \Longrightarrow^* C \\
\hline
A \Longrightarrow^* C
\end{array} \tag{14.4}$$

14.4 Proving Causal Dependency

210

Suppose we have replaced the relation \hookrightarrow with \Longrightarrow^* ; and have replaced \Longrightarrow^* by a chain (that is, a conjunction) of assertions of the \Longrightarrow relation *simpliciter*. We must now try to prove the \Longrightarrow clauses, and for this we need a new inference rule. Given that we are trying to establish the relation $A \Longrightarrow B$ for some A and B, how do we go about doing it? Remembering the Lewis definition of \Longrightarrow in terms of counterfactuals, his general formal definition is

$$A \Longrightarrow B \triangleq (A \square \rightarrow B) \land (\neg A \square \rightarrow \neg B) \tag{14.5}$$

We formulate it rather as an inference rule thus [Lew86]:

We won't need the rule in quite this form, because we consider only cases of $A \Rightarrow B$ when A and B are true. When A is true, the assertion $A \Box \rightarrow B$ turns out to be logically equivalent to $A \wedge B$, as we shall see shortly.

14.4.1 The Mathematical Semantics

The semantics of $\square \rightarrow$ is a possible-world semantics [Lew73b]. Lewis bases this semantics on the Kripke semantics for modal logic, with an additional relation of nearness:

World X is at least as near as world Y to world W

Lewis's syntactic inference rules for counterfactual dependency (the relation $\Box \rightarrow$) are technical, and do not really exhibit the insight into the notion that we need for judging the truth of basic counterfactual assertions that arise in the WBA. We think it best to evaluate the truth of assertions involving " $\Box \rightarrow$ " by using semantic arguments with Lewis's "nearest possible world" semantics [Lew73b]. We therefore need to explain the mathematical basis for the semantics.

We shall adapt this general Rule 14.6 to our specific application, in which we are explaining a history, by reducing it in this case to Rule 14.20. Some semantic argument about the mathematical structure of the nearness relation is required.

Fix W for the moment. This relation of nearness relative to W yields a binary relation $\leq W$, namely

$$X \leq_W Y \tag{14.7}$$

The relation $\leq W$ is supposed to be an ordinal measure according to Lewis. An ordinal measure is one in which any two things can be compared as smaller or larger, or the same as each other [KLST71, 1.1.1, 1.3.1]. That means mathematically that $\leq W$ should be axiomatised as a *total preorder* (called a weak order in [KLST71]). A preorder is a binary relation \leq that satisfies the following properties:

$$x \leq x \quad (reflexivity)$$
 (14.8)

$$(x \leq y) \land (y \leq z) \Rightarrow (x \leq z) \quad (transitivity)$$
 (14.9)

A preorder is *total* if and only if any two objects in the preorder are comparable:

$$(x \leq y) \lor (y \leq x) \quad (totality)$$
 (14.10)

Given a preorder \leq , one can define a relation \simeq as:

$$x \simeq y \Leftrightarrow (x \preceq y) \land (y \preceq x) \tag{14.11}$$

and the relation \simeq is then an *equivalence relation*, namely a relation that satisfies:

$$x \simeq x \quad (reflexivity)$$
 (14.12)

$$(x \simeq y) \Rightarrow (y \simeq x) \quad (symmetry)$$
 (14.13)

$$(x \simeq y) \land (y \simeq z) \Rightarrow (x \simeq z) \quad (transitivity)$$
 (14.14)

It is well-known that an equivalence relation partitions everything into a collection of mutually disjoint sets called equivalence classes, such that every object in the domain of the equivalence relation belongs to exactly one of the equivalence classes: x and y belong to the same equivalence class just in case $x \simeq y$. It is also well known that the equivalence relation \simeq defines a *linear order* or *total order* \preceq equiv on the equivalence classes: let [x] be the equivalence class of x. Then

$$[x] \leq_{equiv} [x] \quad (reflexivity)$$
 (14.15)

$$([x] \leq_{equiv}[y]) \land ([y] \leq_{equiv}[x]) \Rightarrow ([x] = [y]) \quad (antisymmetry)$$
 (14.16)

$$([x] \leq_{equiv}[y]) \land ([y] \leq_{equiv}[z]) \Rightarrow ([x] \leq_{equiv}[z]) \quad (transitivity)$$
 (14.17)

$$([x] \leq_{equiv}[y]) \vee ([y] \leq_{equiv}[x]) \quad (totality) \tag{14.18}$$

That means that any two worlds can be compared in terms of their similarity to world W; either the one or the other is more similar, or they are both equally similar.

14.4.2 Causes from Counterfactuals

Fix the 'real world'. Then for other possible (not actual) worlds (such as in our case that the flight flew from LATC to Maasstricht ATC and thence to FRA), there is a relation of nearness to the actual world, with the above properties. The Lewis semantics for $\Box \rightarrow$ is that

 $A \longrightarrow B$ in a world W if and only if B is true in all the nearest worlds to W in which A is true.

How does one then define the concept 'nearest'? A world X is nearest to a world W if and only if, for all worlds Y, world X is at least as near as world Y to world W. Suppose A is true in world W. Then the set of nearest worlds to W in which A is true consists of precisely W itself. Then $A \square \to B$ is true in W just in case B is also true. Thus we have the axiom:

Axiom 6
$$\vdash$$
 $((A \Box \rightarrow B) \land A) \Rightarrow B$

from which it clearly also follows that

$$\frac{A}{(A \square \to B) \equiv (A \land B)} \tag{14.19}$$

Thus the Lewis rule for counterfactuals in the form in which we use it, to explain the causal-factor relation between facts A and B rather than fictions, reduces to the following:

$$\begin{array}{c}
A \wedge B \\
\neg A \Box \rightarrow \neg B \\
\hline
A \Longrightarrow B
\end{array} \tag{14.20}$$

14.5 Finding Causal Candidates

To begin with, we blend together the information contained in Formula 14.1 with more explicit symbols for the ontological information to produce the first Why..Because Graph, or WB-graph, as shown in Figure 14.1.

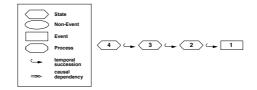


Figure 14.1: Beginning stage of the WB-graph

We wish to replace the temporal " \hookrightarrow " by a series of causal, rather than purely temporal, dependencies. An intuitive way to find intermediate causal factors to replace the temporal " \hookrightarrow " by such a causal series is by stating "Why..Because"-questions:

Why
$$B$$
? – Because C
Why C ? – Because A

An arbitrary node will generally be causally affected not by single but by multiple factors.

We start our analysis by questioning:

Several potentially relevant facts can be detected in the article. We don't yet have any formal criterion for relevance – this is at this 'discovery' stage of the investigation an intuitive judgement. However, the proof will later require demonstration of a sort. We identify the following two causally-relevant facts:

- [11] Crew (CRW) realizes they are landing at the wrong airport
- [12] CRW opts to continue landing

We can test the relations [11] \Rightarrow [1] and [12] \Rightarrow [1] using inference rule (14.20) and intuitive reasoning to establish the \Box relation, and find that the relations hold. Thus they give a partial explanation for the immediate causes of [1]. Including these new nodes into the existing graph results in figure 14.2.

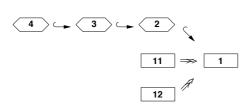


Figure 14.2: View of the WB-graph after first step

The node [12] is a crew decision, and because

- the crew is part of the system, and
- crew decisions are the output of a reasoning process (broadly meant) which the crew undertakes

one can also ask if this outcome is correctly derived from the 'input' plus the reasoning. Further analysis of [12] – which at this point means scanning the text [dWL95] for further information – results in a new node:

(121) CRW has safety reasons for continuing landing

Looking now at [11], finding a causal factor is not that easy, since we do not find any hints in [dWL95] about the reasons why they finally realize the problem. Use of a secondary source [Lad95a] provides additional information:

[111] CRW gets visual contact to BRU airport {112} CRW notices that the BRU airport layout is different from FRA's

Process [111] itself is caused by

[1111] AC breaks out under clouds, $\langle 1112 \rangle$ AC near BRU

and $\langle 2 \rangle$, which is already present in (14.1). We can gain this information directly from the text. There are, however, some things not said. An expert would say that there were also

(1113) CRW procedures

contributing to [111]. We shall see more on including expert knowledge below. We have succeeded now in filling out a causal chain from $\langle 2 \rangle$ to [1] (figure 14.3).

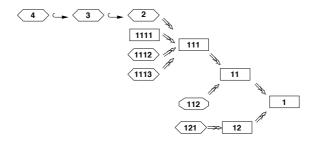


Figure 14.3: Establishing the first causal chain

State $\langle 1112 \rangle$ seems to be worthy of further analysis, because according to its flightplan the aircraft was not supposed to be there under 'normal' circumstances.

It raises the question how and why they got there, and harks forward to use of the criteria of *contrastive explanation* (Chapter 15.2).

Both primary and secondary sources, as well as the intuitive meaning of 'realize', provide information that

(11121) CRW did not realize that they were on wrong course, UNTIL:[11]

(The last part of the phrase explicitly bounds the temporal extent of the state predicate by citing an event beyond which it is not asserted to hold. Clearly, for every event-denoting verb such as 'realizes', the state predicate that something is not realized holds precisely up to this event. This could be formulated as an axiom, but this would go deeper into the logical structure of individual assertions than we think is appropriate here.)

This is an interesting point. How can we know that something *did not* happen, since the states and events we used so far and which are likely to appear in a simple history described what *did* happen?

14.6 Non-Events and Deontics

How on earth can one possibly determine that a non-event – the absence of something – is causally important? Formally, a non-event is a state. Something we await – according to our knowledge about the situation the system is in as well as the obligations following from the procedures which govern the system – does not occur. That kind of argumentation uses deontic reasoning: the procedures ought to be followed. Therefore the event ought to have happened. But it did not (– and this is explicitly remarked). To capture this formally, we can introduce a new inference rule which says exactly this. An event that ought to have happened but didn't, and whose absence is causally factorial, is called a counteracting cause by Mill, who understood the necessity for dealing with non-events [Mac74, Chapter 3].

Before giving the inference rule, we need to analyse the situation a little further. We base the rule on the following principle:

We assert the existence of a non-event, given procedures Proc in a situation S if, given S and always Proc, that the procedures are continually followed, the event must necessarily occur, either then or later; but in fact it doesn't.

This involves two modalities, in the technical sense of the term in modal logic: necessity and tense. Because tense logics and logics of necessity are often considered separately, the same notation is used for both: we need to distinguish notation. We shall use \succ for the Lewis-Langford [LL32] relation of $strict impli-cation: A \succ B$ if B necessarily follows as a matter of logic from A; and \Box , \diamondsuit for the always and eventually operators of simple linear-time tense logic.

There is a third modality to introduce: obligation. The deontic axiom says that procedures *ought* to be followed:

Axiom 7 \vdash O(Procedures)

Suppose an event is a necessary consequence of following procedures in the given situation. Since the procedures ought to be followed, the event ought to occur. It should occur. How do we say necessary consequence? Since we're talking about procedures or systems with behavior, I shall use the tense logic TLA (which includes axioms for simple linear temporal logic, as well as axioms to support the ontology of events and processes, namely TLA actions, and state predicates) and the following axiom:

$$\frac{(\vdash_{TLA}A \Rightarrow B)}{A \succ B} \tag{14.21}$$

This can be used as a proof rule in hierarchical proofs in the following way: to prove $A \succ B$, the proof proceeds according to the proof of $A \Rightarrow B$ using the proof rules of TLA as given in, say [Lad97, Lam94c]. Given this rule for \succ , we may now formulate the *Deontic Rule* which says that when the occurrence of an event is a necessary consequence of procedures, the event ought to happen:

$$\frac{(Hypotheses \land \Box Procedures) \succ \diamondsuit Event}{(Hypotheses \land O(Procedures)) \Rightarrow O(\diamondsuit Event)}$$
(14.22)

It follows trivially as a derived rule from Rule 14.22 and Axiom 7 that

$$\frac{Hypotheses}{(Hypotheses \land \Box Procedures) \succ \Diamond Event}}{O(\Diamond Event)}$$
(14.23)

The event may not in fact occur, even though it should have, because it is perfectly possible that the procedures weren't followed and thus allowed the event not to occur. In our analysis, we need to remark and reason with these events that should have occurred but didn't. We call them non-events. What kinds of objects are they? Well, non-events persist: the system state does not change in the relevant way because the event that causes this change does not occur, so non-events describe *states* whose occurrence is inferred from our knowledge of procedures, and of the current situation.

However, it is difficult to formulate this final step, the existence of non-events, as a formal inference rule, because it really tells us explicitly to remark a particular fact, and there is no way of expressing such meta-facts in the syntax. We have a meta-rule:

Axiom 8 MetaAxiom: Explicitly add to the history those states $\langle \neg E \rangle$ in which E is an event, O(E) is derivable, and E does not occur.

How do we apply this to the example? We have to identify procedures and situation, and an event which should have but didn't occur. As far as the Crew is concerned we learn that

 $\{111211\}$ CRW addresses BATC controller as "Frankfurt" several times,

(111212) ILS has different frequency for FRA.

and that

[111213] CRW asks for the Bruno VOR's frequency.

All these three occurrences should have alerted the crew to the fact that there was a problem. By use of the deontic reasoning we can learn that - contrary to international procedures -

(111214) Brussels did not question the addressing error although it happened more than once

The relevant procedure which (111214) contravenes is something like:

CommProc: If an ATC is addressed by an aircraft using an identifier other than its correct identifier, this identifier should be explicitly queried.

It should be intuitively clear that $\langle \neg 111214 \rangle$ follows from CommProc. The procedure Commproc is a statement of what is standard operating procedure for ATC (also for aircrew). Similarly, one can consider the procedure NavProc:

NavProc: If a navigation aid (navaid) identifier is different from what is expected, or if a navaid is used in clearance instructions that does not appear on the current navigation chart, the navaids and frequency differences should be explicitly queried of ATC by the aircrew.

(Both CommProc and NavProc will be formulated more precisely in TLA when we proceed with the formal proof.) Again, this or something like it is standard operating procedure for aircrew, and both (111212) and [111213] contravene it. The Metaaxiom explains why they explicitly figure in our WBA; intuitive use of the meta-axiom may explain why they were explicitly remarked in the journal history [dWL95, Lad95a, Lad95b].

Adding all this new information leads to figure 14.4

We have gotten so far, but our explanation is by no means complete, yet. We must eliminate the two remaining \hookrightarrow ; and we must somehow show that our partial explanation suffices for this part. We address the second issue next.

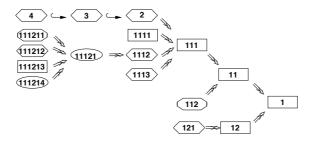


Figure 14.4: Introducing non-events