Chapter 20

The Logic EL

In order formally to verify that a causal explanation is indeed correct and relatively sufficient, we need a formal reasoning system, a formal logic, in which to do so. We describe the logic EL. Since our explanations are explanations of happenings-through-time, the basic concept of a world with behavior will be that of tense logic; accordingly we use the tense logic TLA and the specification language TLA+, and the worlds will be TLA models, TLA *behaviors*. In order to discuss the counterfactual aspects of causality, *what would have happened if...*, we shall need to discuss alternative behaviors to the actual world; accordingly we use the possible-worlds structure of modal logics. This modality, of causally-alternative worlds, will be a different modality to that of tense. We shall also require an alethic modality, to speak of logical necessities, and a deontic modality to speak of behavioral norms. It turns out that both the alethic and deontic modalities may be defined in terms of the counterfactual modality. We introduce the formal rules of EL and illustrate the semantics.

EL uses the TLA rules, which were defined in [Lam94c] and given translation into natural-deduction form in [Lad97]. The sections explaining the classical, temporal and invariance rules modified versions of those in [Lad97]. Figure 20.1 contains a table of all the EL connectives.

20.1 Classical Rules

One may simply assume some set of rules or procedures for classical (propositional and predicate) logic, as does Lamport in the TLA definition [Lam94c].

For completeness, we give a Prawitz-type natural deduction system [Pra65] with the classical rules, presented in the traditional manner: proofs 'move' downwards towards their desired conclusion.

Propositional Logic Operators

$\wedge$	logical notation for *and*
$\vee$	logical notation for *or*
$\neg$	logical notation for *not*
$\forall$	universal quantifier ("*for all*")
$\exists$	existential quantifier ("*exists*")
$\Rightarrow$	*implication*
$\Leftrightarrow, \equiv$	*equivalence*
$\triangleq$	*definition*

Inferences

$\begin{array}{l} A \\ B \\ \hline C \end{array}$	From the antecedents A and B, the consequent C can be inferred.
$\begin{array}{l} A \\ B \\ \hline\hline C \end{array}$	A and also $\begin{array}{l} C \\ \hline A \\ B \end{array}$ $\begin{array}{l} B \\ \hline C \end{array}$

Modal Operators

$\Box$	temporal modality of *always*
$\Diamond$	temporal modality of *eventually*
$\rightsquigarrow$	*leads to* operator (whenever the first argument is true, the second will eventually become true)
$\Box[A]_f$	future validity of A including *stuttering* steps
$\hookrightarrow$	*temporal succession*
$O(\ldots)$	deontic modality of *obligation*
$\Box\!\!\rightarrow$	(Lewis) *counterfactual* dependency
$\succ$	Lewis-Langford relation of *strict implication*
$\preceq_W$	*total preorder* (according to Lewis)
$\Rrightarrow$	*causal explanation*
$\Rrightarrow^*$	*causality*
$\Box\!\!\Rightarrow$	*causal sufficiency*

Figure 20.1: The Logical Connectives of EL occurring in this book

20.1.1 Propositional Rules

$$\text{(\&-intro)}\ \frac{A \quad B}{A \wedge B} \qquad \text{(\&-elim)}\ \frac{A \wedge B}{A} \quad \frac{A \wedge B}{B}$$

$$\text{($\vee$-intro)}\ \frac{A}{A \vee B} \quad \frac{B}{A \vee B} \qquad \text{($\vee$-elim)}\ \frac{A \vee B \qquad \begin{array}{c}[A]\\ C\end{array} \qquad \begin{array}{c}[B]\\ C\end{array}}{C}$$

Note that these rules strictly speaking build proof *trees*, not simple sequences of formulas, since proof branches converge in certain rules. Branches converge as we move 'downwards' in the proof towards the desired conclusion. The root of the tree is the desired theorem, and the leaves are the initial assumptions.

Define $(\neg A)$ to be the formula $(A \Rightarrow \bot)$

$$\text{($\Rightarrow$-intro)}\ \frac{\begin{array}{c}[A]\\ B\end{array}}{A \Rightarrow B} \qquad \text{($\Rightarrow$-elim)}\ \frac{A \quad A \Rightarrow B}{B}$$

$$(\bot_{\mathsf{I}})\ \frac{\bot}{A} \qquad (\bot_{\mathsf{C}})\ \frac{\begin{array}{c}[\neg A]\\ \bot\end{array}}{A}$$

A restriction on both $\bot$ rules is that A is different from $\bot$. A restriction on $\bot_{\mathsf{C}}$ is that A is not of the form $B \Rightarrow \bot$ (but this is only for convenience: Exercise:- Show that if A has the form $\neg B$ in $\bot_{\mathsf{C}}$ that there's another way of deriving the conclusion). Note that $\bot_{\mathsf{C}}$ subsumes $\bot_{\mathsf{I}}$.

The use of $\bot$ is logically important: if $\bot_{\mathsf{C}}$ is used, the theorems are those of classical propositional logic; if $\bot_{\mathsf{I}}$ is used, the theorems are those of intuitionistic propositional logic. The *formal* difference between the two logics is thus reduced just to that of how negation is treated.

20.1.2 Quantifiers

we may include rules for quantifiers as follows:

$$\text{($\forall$−intro)}\ \frac{A}{\forall x : A[a/x]} \qquad \text{($\forall$−elim)}\ \frac{\forall x : A}{A[x/t]}$$

in which a does not occur in the assumption set of premise A, nor within the scope of a quantifier binding x in A; resp. no occurence of x in A occurs within the scope of a quantifier binding a variable of t (i.e. the standard logic stuff usually embodied in the definition of *<term> is free for <variable> in <formula>*).

The rules for the existential quantifier are

$$(\exists\text{–intro}) \quad \frac{A[x/t]}{\exists x : A} \qquad (\exists\text{–elim}) \quad \frac{\exists x : A \qquad \overset{[A[a/x]\,]}{B}}{B}$$

with appropriate restrictions here too.

20.2 Modal Rules

Following the approach of Prawitz [Pra65], we first define the notion of *essentially tense-modal* formula:

- $\Box A$ is essentially tense-modal;
- $\bot$ is essentially tense-modal;
- the essentially tense-modal formulas are closed under $\wedge$, $\vee$, $\exists$

The rules are

$$(\Box\text{–intro}) \quad \frac{A}{\Box A} \qquad (\Box\text{–elim}) \quad \frac{\Box A}{A}$$

in which the assumption set of the premise of ($\Box$–intro) must include only essentially-modal formulas.

This set of rules defines a natural deduction system whose theorems are those of S4 (as shown in [Pra65]).

Axioms are proof rules with conclusions but without hypotheses, and rather than write an axiom in proof rule notation as $\frac{}{A}$ we write $\vdash A$. The modal notions are more commonly formulated as Hilbert-style systems, with restricted proof rules (usually just *modus ponens* and *necessitation*) and many axioms.

So, alternatively, one may add the rule of *necessitation*:

$$\frac{F}{\Box F}$$

with the constraint that $\mathit{assumptions}(F) = \emptyset$,
and various of the axioms

(K) $\quad \vdash \Box(A \wedge B) \equiv (\Box A \wedge \Box B)$
(T) $\quad \vdash \Box B \Rightarrow B$
$(S4)$ $\quad \vdash \Box B \Rightarrow \Box\Box B$
$(S5)$ $\quad \vdash \Diamond_A B \Rightarrow \Box_A \Diamond_A B$

as one wishes, to obtain the system one needs [HC96]. The basic propositional modal logic contains the axiom K along with the rule of necessitation, and is itself

known as K. A model for a modal logic extending K is a collection of classical (propositional or predicate-logical) models, along with a binary relation $\mathcal{W}$ on these models, known as *accessibility*. If $X\mathcal{W}Y$, we say that world Y is accessible from world X. We speak of formulas being satisfied by worlds in a model M, thus $X \models A$ in M means that formula A is true at world X in model M. The inductive rules for truth at a world in a model for the classical connectives and quantifiers are identical to those of classical logic; for example

$$X \models (A \wedge B) \triangleq X \models A \wedge X \models B$$

$$X \models \neg A \triangleq \neg(X \models A)$$

suffice for propositional logic, since $\{\neg, \wedge\}$ form a sufficient set of connectives for classical propositional logic. The interpretation of the operators $\Box, \Diamond$ uses the accessibility relation essentially:

$$X \models \Box A \triangleq \forall Y(X\mathcal{W}Y \Rightarrow Y \models A)$$

The rule for $\Diamond$ may be derived from the rule for $\Box$, since $\Diamond \triangleq \neg\Box\neg$, and is

$$X \models \Diamond A \triangleq \exists Y(X\mathcal{W}Y \wedge Y \models A)$$

In the Kripke (possible worlds) semantics for modal logic, the accessibility relation in any model for (T) is reflexive and, conversely, any Kripke model with a reflexive accessibility relation is a model for (T). The logic S4 is axiomatised by the axioms $(K) + (T) + (S4)$. The accessibility relation in any model for S4 is reflexive and transitive and, conversely, any Kripke model with a reflexive, transitive accessibility relation is a model for S4. The logic S5 is axiomatised by the axioms $(K)+(T)+(S4)+(S5)$. The accessibility relation in any model for S5 is reflexive, transitive and symmetric, that is, an equivalence relation; conversely, any Kripke model with a reflexive, transitive and symmetric accessibility relation is a model for S5.

20.3 Temporal Rules

We need tense-logical rules sufficient for '*simple linear-time temporal logic*' (STL), as some computer scientists call it. It is equivalent to S4.3. There are also some special TLA rules for handling the '*prime*' operator. While we could extend the Prawitz rules with axioms for S4.3, the existing TLA system is formulated axiomatically by Lamport [Lam94c]. We include the Lamport axioms and rules here. We refer readers interested in a natural deduction formulation to that presented in [Lad97].

STL1: $\dfrac{F}{\Box F}$
in which $assumptions(F) = \emptyset$
STL2: $\vdash \Box F \Rightarrow F$
STL3: $\vdash \Box\Box F \equiv \Box F$
STL4: $\dfrac{F \Rightarrow G}{\Box F \Rightarrow \Box G}$
STL5: $\vdash \Box(F \wedge G) \equiv (\Box F) \wedge (\Box G)$
STL6: $\vdash (\Diamond\Box F) \wedge (\Diamond\Box G) \equiv \Diamond\Box(F \wedge G)$
STL7: $\vdash \Box\Diamond\Box F \equiv \Diamond\Box F$

To these STL axioms we add the **Basic TLA Rules** which specifically axiomatise '*prime*':

TLA1. $\dfrac{P \wedge (f' = f) \Rightarrow P'}{\Box P \equiv P \wedge \Box[P \Rightarrow P']_f}$ TLA2. $\dfrac{P \wedge [\mathcal{A}]_f \Rightarrow Q \wedge [\mathcal{B}]_g}{\Box P \wedge \Box[\mathcal{A}]_f \Rightarrow \Box Q \wedge \Box[\mathcal{B}]_g}$

where the notation

$$[\mathcal{A}]_f \triangleq \mathcal{A} \vee (f' = f)$$

20.4 Behaviors and the Rules They Engender

In order to allow proof by induction over temporal time-steps, we need to add an induction rule:

INV1: $\dfrac{I \wedge [\mathcal{N}]_f \Rightarrow I'}{I \wedge \Box[\mathcal{N}]_f \Rightarrow \Box I}$

in which $\mathcal{N}$ is an action formula, thus defining a relation between a given state and the next. We suppose also that f is the state function which is a sequence whose elements are the variables of $\mathcal{N}$ (TLA includes ZF, so sequences can be formed). The rule says intuitively: if it is provable that I is preserved under state-relation $\mathcal{N}$, then it follows if $\mathcal{N}$ always holds and I holds, I will always hold. We also add the technical rule:

INV2: $\vdash \Box I \Rightarrow (\Box[\mathcal{N}]_f \equiv \Box[\mathcal{N} \wedge I \wedge I']_f)$

These two rules are known as the 'Invariance Rules'. We hope that it is clear that these rules are sound. They suffice for many purposes, for example to prove *safety properties* of systems.

20.5 Strict Implication

We observe that the only place that strict implication is used in EL is to restrict the subproof to TLA logic. We do not anticipate use of combinatorial properties

of $\succ$ in reasoning about failures in EL. We nevertheless define them, in case they should be needed.

Strict implication $\succ$ is supposed to represent alethic modal implication, that is, (logical or philosophical) *necessity*. Rather than being axiomatised directly as a binary logical operator, $B \succ C$ is usually defined as

$$B \succ C \quad \triangleq \quad \Box_A(B \Rightarrow C) \tag{20.1}$$

and $\Box_A$ is axiomatised. We take this course and take $\Box_A$ to be axiomatised by the S5 logic, as is commonplace for the alethic modality intended to capture logical necessity.

Alternatively, we could use a Prawitz-like approach, defining the notion of *essentially alethically-modal* formula, and then giving rules for $\Box_A$ isomorphic to those above for $\Box$, which yields the logic S4, and adding the $(S5)$ axiom.

20.6 The Deontic Modalities

For the deontic part, we may use *Standard Deontic Logic* (SDL) [MWD78]. Along with *modus ponens* and classical logic, SDL consists of the axioms/rules:

(K_O) $\quad \vdash O(A \Rightarrow B) \Rightarrow (O(A) \Rightarrow O(B))$

(D_O) $\quad \vdash \neg O(\bot)$

(N_O) $\quad \dfrac{A}{O(A)}$

(P) $\quad \vdash P(A) \equiv \neg O(\neg A)$

(F) $\quad \vdash F(A) \equiv O(\neg A)$

However, a clean way of implementing SDL inside an S5 modal logic was discovered by Anderson [And58] (see also [MWD78]). Given an alethic S5-operator $\Box_A$, define a propositional constant V (standing intuitively for '*Violation*'), and further define

$$O(B) \equiv \Box_A(\neg B \Rightarrow V) \tag{20.2}$$

Since we have an alethic S5 operator for the definition of strict implication, we may use the Anderson definition in order to define the deontic modality.

20.7 Lewis Semantics for Counterfactuals

The Lewis possible-world semantics for counterfactuals [Lew73b] is based 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

which we denote $AsNear(X, Y, W)$ The Lewis semantics for $\Box\!\!\rightarrow$ is that

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

We use semantic arguments about nearest possible worlds to evaluate formulas of the form $A \boxrightarrow B$. We repeat here for ease of reference the mathematical structure concerning nearness amongst possible worlds which is necessary to make such arguments. The relation of nearness yields a collection of binary relations $\preceq_W$, one for each world W, by means of the definition

$$(X \preceq_W Y) \triangleq AsNear(X, Y, W) \tag{20.3}$$

The relation $\preceq_W$ shall be a *total preorder*, satisfying the following properties:

$$x \preceq x \quad (reflexivity) \tag{20.4}$$

$$(x \preceq y) \wedge (y \preceq z) \Rightarrow (x \preceq z) \quad (transitivity) \tag{20.5}$$

$$(x \preceq y) \vee (y \preceq x) \quad (totality) \tag{20.6}$$

The relation $\simeq$ is defined from $\preceq$ as:

$$x \simeq y \triangleq (x \preceq y) \wedge (y \preceq x) \tag{20.7}$$

It follows that the relation $\simeq$ is an *equivalence relation*, namely one that satisfies:

$$x \simeq x \quad (reflexivity) \tag{20.8}$$

$$(x \simeq y) \Rightarrow (y \simeq x) \quad (symmetry) \tag{20.9}$$

$$(x \simeq y) \wedge (y \simeq z) \Rightarrow (x \simeq z) \quad (transitivity) \tag{20.10}$$

The equivalence classes are defined by the condition: x and y belong to the same equivalence class just in case $x \simeq y$. It follows from 20.4 20.5 20.6 20.7 20.8 20.9 and 20.10 that there is a *linear order* or *total order* $\preceq_{equiv}$ on the equivalence classes: let $[x]$ be the equivalence class of x. Then

$$[x] \preceq_{equiv} [x] \quad (reflexivity) \tag{20.11}$$

$$([x] \preceq_{equiv} [y]) \wedge ([y] \preceq_{equiv} [x]) \Rightarrow ([x] = [y]) \quad (antisymmetry) \tag{20.12}$$

$$([x] \preceq_{equiv} [y]) \wedge ([y] \preceq_{equiv} [z]) \Rightarrow ([x] \preceq_{equiv} [z]) \quad (transitivity) \tag{20.13}$$

$$([x] \preceq_{equiv} [y]) \vee ([y] \preceq_{equiv} [x]) \quad (totality) \tag{20.14}$$

Finally, an additional requirement on $\preceq_W$ is that W is more similar to itself than any other world which is similar to it:

$$X \preceq_W X \Rightarrow W \preceq_W X \tag{20.15}$$

This entails that the relation $\preceq_W$ yields an *ordinal measure* or *ordinal scale* [KLST71] on worlds: 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.

A mathematical way of phrasing the semantics of $\boxrightarrow$ [Lew73b, p16] is:

- $A \boxrightarrow B$ is true at world W if and only if either
  1. A is true at no world accessible from W at any distance; or
  2. A is true at some world accessible from W, and $A \Rightarrow B$ is true at all worlds in $\{X | X \preceq_W A\}$.

20.8 Rules for Counterfactual Conditionals

It follows from the semantics for $\boxrightarrow$ that all the inference rules for logic still hold if all the formulas are prefaced by '$A \boxrightarrow$': that is, for example

$$\text{(\&-intro)}\quad \frac{S \boxrightarrow A \quad S \boxrightarrow B}{S \boxrightarrow A \wedge B}$$

$$\text{(\&-elim)}\quad \frac{S \boxrightarrow A \wedge B}{S \boxrightarrow A} \qquad \frac{S \boxrightarrow A \wedge B}{S \boxrightarrow B}$$

$$\text{($\vee$-intro)}\quad \frac{S \boxrightarrow A}{S \boxrightarrow A \vee B} \qquad \frac{S \boxrightarrow B}{S \boxrightarrow A \vee B}$$

$$\text{($\vee$-elim)}\quad \frac{S \boxrightarrow A \wedge B \qquad [S \boxrightarrow A] \quad [S \boxrightarrow B]}{S \boxrightarrow C \qquad S \boxrightarrow C}$$

$$\text{($\Rightarrow$-intro)}\quad \frac{[S \boxrightarrow A]}{S \boxrightarrow B}$$

$$\text{($\Rightarrow$-elim)}\quad \frac{S \boxrightarrow A \quad S \boxrightarrow A \Rightarrow B}{S \boxrightarrow B}$$

These should be derived rules from a reasonable axiomatisation of the counterfactual conditional. Lewis axiomatises his preferred logic of counterfactuals, VC, as follows [Lew73b, p132]. There are two rules:

$$\text{Modus Ponens} \tag{20.16}$$

$$\frac{(A_1 \wedge \ldots \wedge A_n) \Rightarrow B}{((C \boxrightarrow A_1) \wedge \ldots \wedge (C \boxrightarrow A_n)) \;\Rightarrow\; (C \boxrightarrow B)} \tag{20.17}$$

in which the second Rule 20.17 is actually a rule schema, one rule for each value of $n > 1$. The axioms are:

Axiom 11 *Truth-functional tautologies*

Axiom 12 *Definitions of the non-primitive related operators (which we don't use here)*

Axiom 13 $A \mathrel{\Box\!\!\rightarrow} A$

Axiom 14 $(\neg A \mathrel{\Box\!\!\rightarrow} A) \Rightarrow (B \mathrel{\Box\!\!\rightarrow} A)$

Axiom 15 $(A \mathrel{\Box\!\!\rightarrow} \neg B) \vee (((A \vee B) \mathrel{\Box\!\!\rightarrow} C) \equiv (A \mathrel{\Box\!\!\rightarrow} (B \Rightarrow C)))$

Axiom 16 $(A \mathrel{\Box\!\!\rightarrow} B) \Rightarrow (A \Rightarrow B)$

Axiom 17 $(A \wedge B) \Rightarrow (A \mathrel{\Box\!\!\rightarrow} B)$

20.9 Defining the Other Modalities

From a counterfactual satisfying *VC* (or the Lewis semantics in general), one may define a modal operator

$$\Box_C A \triangleq (\neg A) \mathrel{\Box\!\!\rightarrow} A \tag{20.18}$$

called by Lewis the *outer modality*. If the outer modality satisfies in addition the rule (*S*5), then the outer modality forms the S5 modal logic. The (*S*5) axiom for the outer modality corresponds to the following axiom for $\mathrel{\Box\!\!\rightarrow}$:

Axiom 18 $\neg(A \mathrel{\Box\!\!\rightarrow} \neg A) \Rightarrow ((A \mathrel{\Box\!\!\rightarrow} \neg A) \mathrel{\Box\!\!\rightarrow} \neg(A \mathrel{\Box\!\!\rightarrow} \neg A))$

Lewis calls the logic *VC* with the additional Axiom 18 *VCU* [Lew73b, p121].

This means that if we take a *possible world* to be a model of TLA (a *behavior*, a sequence of classical models ordered like the non-negative integers), in which each pertinent state variable is a assigned a definite set-valued value), and the set of possible worlds to satisfy the Kripke semantics (that is, that there is a binary relation of *accessibility* between possible worlds), and furthermore the worlds accessible from any given world W are totally-preordered by *nearness* to W, then we may axiomatise $\mathrel{\Box\!\!\rightarrow}$ as above (including Axiom 18), define the S5 alethic modality $\Box_C$ from $\mathrel{\Box\!\!\rightarrow}$ using Definition 20.18, (that is, $\Box_A \triangleq \Box_C$), define $\succ$ from $\Box_C$ using Definition 20.1, and define the deontic modality O from $\Box_C$ using Definition 20.2.

We may define the operator $\Rightarrow\!\!\!\!\succ$ from $\mathrel{\Box\!\!\rightarrow}$ as in Definition 14.5:

Axiom 19 $A \Rightarrow\!\!\!\!\succ B \triangleq (A \wedge B \wedge (\neg A \mathrel{\Box\!\!\rightarrow} \neg B))$

The transitive closure $\Rightarrow\!\!\!\!\succ^*$ is not definable in a first-order way from the relation of which it is the transitive closure, $\Rightarrow\!\!\!\!\succ$, as is well-known. Thus the axioms giving the recursive definition of $\Rightarrow\!\!\!\!\succ^*$ remain in EL. (This is not to claim that $\Rightarrow\!\!\!\!\succ^*$ is not definable in *VCU*-EL, but rather ignorance as to whether it is.)

20.10 Causal Sufficiency

We gave two inference rules for causal sufficiency in Section 15.1.3. These were Rule 15.5:

$$\frac{\begin{array}{l} C \\ B \\ \neg C \ \Box\!\!\rightarrow \ \neg B \\ \neg B \ \Box\!\!\rightarrow \ \neg C \end{array}}{C \ \Box\!\!\Rightarrow \ B}$$

and Rule 15.6

$$\frac{\begin{array}{r} A \ \Box\!\!\rightarrow \ C \\ B \ \Box\!\!\rightarrow \ C \end{array}}{(A \vee B) \ \Box\!\!\rightarrow \ C}$$

which suffice to demonstrate the CCT. We resisted inverting Rule 15.5 because we considered deriving the logically minimal causal conditions to be an impractical constraint, as we argued in Section 15.1.4. We preferred to add the Rule 15.7

$$\frac{\begin{array}{l} \mathit{Hypotheses} \\ \mathit{Procedures} \\ (\mathit{Hypotheses} \wedge \Box \mathit{Procedures}) \succ \Diamond \mathit{Event} \\ \Diamond \mathit{Event} \end{array}}{(\mathit{Hypotheses} \wedge \mathit{Procedures}) \ \Box\!\!\Rightarrow \ \Diamond \mathit{Event}}$$

which also necessitated adding the Rule 15.9:

$$\frac{\begin{array}{l} X \\ C \\ X \succ C \\ (C \wedge A) \ \Box\!\!\Rightarrow B \end{array}}{(X \wedge A) \ \Box\!\!\Rightarrow B}$$

We do not rule out that appropriate properties for $\Box\!\!\Rightarrow$ will still be found wanting, and that therefore other reasonable inference rules for $\Box\!\!\Rightarrow$ will need to be added to EL.

20.11 The Well-Formed Formulas of EL

We define the *well-formed formulas* of EL to be those formulas of *VCU*, with the extra defined primitives as in Section 20.9, plus $\Rightarrow\!\!\!\!\succ^*$, and with formulas of TLA substituted for the propositional primitives. That is, the primitive formulas are those of TLA, and the well-formed formulas of EL are the formulas constructed from these by the syntactic rules of *VCU* with the addition of the defined operator

symbols and the symbol $\Rightarrow\!\!\!\!\succ^*$. This ensures that TLA-tense-operators may only occur *strictly within the scope of* the other modalities.

This syntax definition makes semantic sense. *Possible worlds* in EL are those ways the world might have been. We have construed 'the world' as a temporal succession of states and events, namely those states and events that happened during the incident, in addition to a lot of others. So 'the world' is a model for TLA, a TLA *behavior*. Alternative worlds are alternative behaviors. Thus a *possible world* in the Kripke sense for the counterfactual logic *VC* (and thus for *VCU*) consists of a TLA behavior, which is a model for TLA formulas. The *VCU* logic makes assertions concerning propositions true or false in these possible worlds. Thus the *VCU* logic is syntactically built from 'propositional primitives' which describe propositions true or false in these possible worlds, i.e. TLA formulas.

Hence the definition of well-formed formulas for EL follows directly from the requirement that a *possible world* for the $\Box\!\!\rightarrow$ logic consists of a TLA behavior.

20.12 Soundness and Completeness Observations

Let us denote by *VCU*-EL the fragment of EL axiomatised by *VCU*, with the defined operators (and therefore without $\Rightarrow\!\!\!\!\succ^*$), on top of the TLA primitives, plus the TLA axioms and rules.

The binary relation of *accessibility*, needed for the Kripke semantics on possible worlds, is definable from the ternary relation *nearness* as follows. Let accessibility by $accessible(W, X)$. Then

$$accessible(W, X) \triangleq X \preceq_W X \tag{20.19}$$

so all one needs for the semantics is a ternary relation

$$AsNear(W, X, Y) \triangleq X \preceq_W Y$$

whose projections on the first argument W, namely $\{\langle X, Y\rangle | X \preceq_W Y\}$, are total preorders with least element W (W must be least, according to Axiom 20.15), that satisfy the *Uniformity Condition*

Axiom 20 *For any* W *and* Z, $\{Y | accessible(W, Y)\} = \{Y | accesible(Z, Y)\}$

The Uniformity Condition is the semantic counterpart of Axiom 18: *VC*-models that satisfy the Uniformity Condition are *VCU*-models, and *VCU*-models satisfy the Uniformity Condition [Lew73b, pp120-121]. Axiom 18 is said to be the *characteristic axiom* for the Uniformity Condition.

If one defines all the modal operators from primitive $\Box\!\!\rightarrow$, then the resulting logic must be a conservative extension of *VCU*, and thus its soundness follows from the soundness of *VCU*, as shown in [Lew73b, pp122,124,133]. Its completeness follows from the fact that Axiom 18 is the characteristic axiom for the Uniformity Condition.

Axiom 1 $\vdash (A \mathrel{+\!\!\!\succ}^* B) \Rightarrow (A \leadsto B)$

Axiom 2 $\vdash (A \mathrel{+\!\!\!\succ} B) \Rightarrow (A \mathrel{+\!\!\!\succ}^* B)$

Axiom 5 $\vdash (A \mathrel{+\!\!\!\succ}^* B) \wedge (B \mathrel{+\!\!\!\succ}^* C) \Rightarrow (A \mathrel{+\!\!\!\succ}^* C)$

Axiom 6 $\vdash ((A \boxdot\!\!\!\rightarrow B) \wedge A) \Rightarrow B$

Axiom 7 $\vdash O(\mathit{Procedures})$

Figure 20.2: Special Axioms of EL

However, since the propositional primitives of *VCU*-EL are TLA formulas, this argues for the soundness and completeness of *VCU*-EL only relative to TLA.

TLA is complete for proof of formulas that occur during hierachical verifications in TLA (namely, of the form $\mathit{TLA.Spec}_1 \Rightarrow \mathit{TLA.Spec}_2$), and it is sound, assuming that ZF set theory is sound [Lam94c].

Thus follows the relative soundness and completeness of *VCU*-EL from that of *VCU* and TLA. EL itself is not complete, because the intended interpretation of the relation $\mathrel{+\!\!\!\succ}^*$ is as the transitive closure of $\mathrel{+\!\!\!\succ}$, whereas many other relations, such as the universal relation on nodes and their Boolean combinations, also satisfy Axioms 2 and 5. That this relation, which exists in every model of *VCU*-EL, satisfies Axioms 2 and 5 shows that EL is sound.

EL itself contains some extra rules which are not necessarily derived rules of *VCU*-EL, although many are. So the soundness and completeness of EL does not necessarily follow from that of $\mathit{VCU} - \mathit{EL}$.

20.13 Special EL Rules

Although we could define the relation $\leadsto$ using the tense-logical rules, we don't need to and haven't. The EL rules as expressed here clearly do not form a non-redundant set, since some of them follow from others, as noted in the text. Furthermore, many of them are derived rules if non-temporal modalities of EL are defined from $\boxdot\!\!\!\rightarrow$, as in Section 20.8. They are listed in Figures 20.3 and 20.4.

Of these,

- Axiom 1, Axiom 2, Axiom 5, Rule 14.2, Rule 14.3, and Rule 14.4 are not derived rules of *VCU*-EL, because the operator $\mathrel{+\!\!\!\succ}^*$ does not appear in *VCU*-EL; neither could it be reasonably defined in EL as far as we can see;
- Rule 14.6, Axiom 6, Rule 14.19, Rule 14.20, Rule 14.21 and Rule 15.5 are derived rules of *VCU*-EL;

Rule 14.2 $\dfrac{A \Rightarrow\!\!\!\!\succ^* B}{A \hookrightarrow B}$

Rule 14.3 $\dfrac{A \Rightarrow\!\!\!\!\succ B}{A \Rightarrow\!\!\!\!\succ^* B}$

Rule 14.4 $\dfrac{A \Rightarrow\!\!\!\!\succ^* B \quad B \Rightarrow\!\!\!\!\succ^* C}{A \Rightarrow\!\!\!\!\succ^* C}$

Rule 14.6 $\dfrac{A \mathrel{\square\!\!\rightarrow} B \quad \neg A \mathrel{\square\!\!\rightarrow} \neg B}{A \Rightarrow\!\!\!\!\succ B}$

Rule 14.19 $\dfrac{A}{(A \mathrel{\square\!\!\rightarrow} B) \equiv (A \wedge B)}$

Rule 14.20 $\dfrac{A \wedge B \quad \neg A \mathrel{\square\!\!\rightarrow} \neg B}{A \Rightarrow\!\!\!\!\succ B}$

Figure 20.3: Special Rules of EL: Part 1

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

Rule 14.22
$$\frac{(Hypotheses \wedge \Box Procedures) \succ \Diamond Event}{(Hypotheses \wedge O(Procedures)) \Rightarrow O(\Diamond Event)}$$

Rule 14.23
$$\frac{Hypotheses \quad (Hypotheses \wedge \Box Procedures) \succ \Diamond Event}{O(\Diamond Event)}$$

Rule 15.5
$$\frac{C \quad B \quad \neg C \ \Box\!\!\rightarrow \neg B \quad \neg B \ \Box\!\!\rightarrow \neg C}{C \ \Box\!\!\Rightarrow \ B}$$

Rule 15.7
$$\frac{Hypotheses \quad Procedures \quad (Hypotheses \wedge \Box Procedures) \succ \Diamond Event \quad \Diamond Event}{(Hypotheses \wedge Procedures) \ \Box\!\!\Rightarrow \ \Diamond Event}$$

Rule 15.9
$$\frac{X \quad C \quad X \succ C \quad (C \wedge A) \ \Box\!\!\Rightarrow B}{(X \wedge A) \ \Box\!\!\Rightarrow B}$$

Figure 20.4: Special Rules of EL: Part 2

Rule 19.1 Conflict Resolution

$$\frac{\begin{array}{l} \mathit{Hypotheses} \\ \mathit{Hypotheses} \wedge \Box \mathit{Procedures} \succ \left(\begin{array}{ll} \wedge & \mathit{Reason}(X, A) \\ \wedge & \mathit{Reason}(X, \neg A) \end{array} \right) \\ \left(\begin{array}{ll} \wedge & \mathit{Hypotheses} \\ \wedge & \mathit{Reason}(X, A) \\ \wedge & \mathit{Reason}(X, \neg A) \\ \wedge & \Box \mathit{Phases.Decision} \end{array} \right) \succ O(\mathit{Decide}(X, A) \vee \mathit{Decide}(X, \neg A)) \\ \mathit{Decide}(X, A) \end{array}}{\left(\begin{array}{ll} \wedge & \mathit{Hypotheses} \\ \wedge & \Box \mathit{Procedures} \\ \wedge & \Box \mathit{Phases.Decision} \\ \wedge & \mathit{Decide}(X, A) \end{array} \right) \Box\!\!\Rightarrow \mathit{Decide}(X, A)}$$

Rule 19.6 Inconsistency

$$\frac{\begin{array}{l} \mathit{Hypotheses} \\ \mathit{Hypotheses} \wedge \Box \mathit{Procedures} \succ \bot \\ \mathit{Hypotheses} \succ \mathit{Phases.Mode.InMode} \end{array}}{O \left(\begin{array}{r} \vee \mathit{Decide}(X, \mathit{Phases.Mode.ExitMode}) \\ \vee \mathit{Decide}(X, \mathit{Phases.Mode.RemainInMode}) \end{array} \right)}$$

The module *Phases* has the specific form described in Chapter 19, and *Phases.Decision* consists of specific behavioral conflict-resolution axioms.

Figure 20.5: Behavioral Rules of EL: Part 1

Rule 19.9 Conflict-Exit

$$\begin{array}{l} \mathit{Hypotheses} \\ \mathit{Hypotheses} \wedge \Box \mathit{Procedures} \succ \bot \\ \mathit{Hypotheses} \succ \mathit{Phases.Mode.InMode} \\ \mathit{Decide}(X, X, \mathit{Phases.Mode.ExitMode}) \end{array}$$

$$\left(\begin{array}{ll} \wedge & \mathit{Hypotheses} \\ \wedge & \Box \mathit{Procedures} \\ \wedge & \Box \mathit{Phases.Decision} \\ \wedge & \mathit{Decide}(X, \mathit{Phases.Mode.ExitMode}) \end{array} \right) \Box\Rightarrow$$

$$\Box\Rightarrow \mathit{Decide}(X, X, \mathit{Phases.Mode.ExitMode})$$

Rule 19.10 Conflict-Remain

$$\begin{array}{l} \mathit{Hypotheses} \\ \mathit{Hypotheses} \wedge \Box \mathit{Procedures} \succ \bot \\ \mathit{Hypotheses} \succ \mathit{Phases.Mode.InMode} \\ \mathit{Decide}(X, X, \mathit{Phases.Mode.RemainInMode}) \end{array}$$

$$\left(\begin{array}{ll} \wedge & \mathit{Hypotheses} \\ \wedge & \Box \mathit{Procedures} \\ \wedge & \Box \mathit{Phases.Decision} \\ \wedge & \mathit{Decide}(X, \mathit{Phases.Mode.RemainInMode}) \end{array} \right) \Box\Rightarrow$$

$$\Box\Rightarrow \mathit{Decide}(X, \mathit{Phases.Mode.RemainInMode})$$

The module *Phases* has the specific form described in Chapter 19, and *Phases.Decision* consists of specific behavioral conflict-resolution axioms.

Figure 20.6: Behavioral Rules of EL: Part 2

- Rule 14.22 and Rule 14.23 have syntactically restricted components ('*Procedures*', '*Event*') devised by the intent and judgement of the analyst, and therefore cannot be derived rules of *VCU*-EL. Rule 14.23 is, however, obviously derivable from Rule 14.22. Rule 14.22 itself is not derived, when *Procedures* and *Event* are propositional variables, and $O(P)$ is defined as $\Box(\neg P \Rightarrow V)$ - here is a simple argument. The consequent of the rule is equivalent to

$$\Box((\mathit{Hypotheses} \wedge \Box(P \vee V)) \Rightarrow (\Diamond E \vee V))$$

Suppose we consider an instance in which $\mathit{Hypotheses} \equiv \top$, i.e.,

$$\Box(\Box(P \vee V) \Rightarrow (\Diamond E \vee V))$$

then this is equivalent to

$$\Box\Box(P \vee V) \Rightarrow \Box(\Diamond E \vee V))$$

which in turn is equivalent in S5 to

$$\Box(P \vee V) \Rightarrow \Box(\Diamond E \vee V))$$

The antecedent of the rule is equivalent to

$$\Box P \Rightarrow \Box\Diamond E$$

Suppose now we take $P \equiv \neg V$: then the antecedent is

$$\Box\neg V \Rightarrow \Box\Diamond E$$

and the consequent is

$$\Box(\Diamond E \vee V))$$

A set of possible worlds in which there is a violation in some world, but in which $\Diamond E$ doesn't hold in this world, satisfies the antecedent but not the consequent. Therefore when P and E are not suitably restricted, this rule is unsound. So in particular it is not derivable from *VCU*-EL.

- For similar considerations as for Rule 14.23, Rule 15.7 is not a derived rule of *VCU*-EL, and neither is Rule 15.9.

This completes the explanation of EL. We repeat the caveat that we do not necessarily consider the set of special axioms and rules in Figures 20.2, 20.3 and 20.4, along with the axioms and rules of VCU-EL, to suffice for our practical goals, because we do not necessarily have all the appropriate rules governing sufficiency, $\Box\!\!\Rightarrow$. Since we declined to define $\Box\!\!\Rightarrow$ as minimal sufficient causal explanation, it is a matter for consideration, judgement and reasoned argument

which rules for $\Box\!\!\Rightarrow$ are appropriate. We do not claim to have given, or to have found, all such appropriate considerations, therefore we leave open the possibility of adding more rules for $\Box\!\!\Rightarrow$ to VCU-EL.

A particular extension may come from considering formulae such as

$$O(Decide(X, A) \vee Decide(X, \neg A)) \wedge Decide(X, A) \Box\!\!\Rightarrow Decide(X, A) \quad (20.20)$$

which involve a mixture of PARDIA primitives with VCU-EL primitives. This formula says that if a decision between A and $\neg A$ is procedurally required, and A is decided, then these two facts alone explain why A was decided: the human had to choose one or the other, and did so (rightly or wrongly). Which such formulas could be added as axioms depends upon the reasoning given for their general truth. We suggest that adding them would turn PARDIA from a classification system into a model, because rather than simply classifying which cognitive states a human might be in, they make assertions about how the presence of certain cognitive and deontic states constitute an explanation of certain others. This is more than simple classification of which states occurred.

Because of our reluctance to propose PARDIA as a model, we are also reluctant to add formulas such as (20.20) as axioms. This means that we must establish the truth of individual instances explicitly in a formal proof as required, as we have done below in the formal proof of completeness and relative sufficiency of [11]. However, we do not suggest that these formulas are false, or otherwise unworthy to become axioms. Those who think that PARDIA could be developed into a worthwhile and adequate model for human processes in incident analysis will see reason to add such formulas to EL. We don't object to doing so.

20.14 Axioms and Processes for WBA

Certain special axioms must be included for each WBA analysis, namely that (human) procedures and (machine) specifications are deontic axioms. These are properly regarded not as axioms of EL itself, since various procedures or machine specifications could be contradictory if care is not taken with scoping the variables used. In any particular application of EL to perform a WBA, the following axioms must be added, for each set of procedures or system specifications:

$$\vdash O(Procedures)$$

$$\vdash O(Spec)$$

Similarly, there are two 'meta-axioms' that are more properly processes to be followed by the analyst in constructing a WBA, and are not themselves part of the EL logic. These are

> *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.*

For any B which has a causal factor (that is, for which there is an A such that $A \Rightarrow\!\!\!\!\!\!\sim\ B$ is established), the set $\{A|A \Rightarrow\!\!\!\!\!\!\sim\ B\}$ must be a satisfactory set of causal factors for B.

The difference between constructing proofs in EL and proofs in, say, TLA should be noted. A step in a TLA proof is always a theorem of TLA. The TLA rules are rules of proof – if the hypotheses are TLA theorems then the consequents are. Rules of proof are theoremhood-preserving rules. However, a step in an EL proof of correctness and sufficiency of an explanation is not necessarily a theorem of EL. For example, the truths of the matter about the incident to be explaining occur as steps in an El proof of an explanation. These are mostly not theorems of any logic! The EL rules are rules of inference – truth-preserving rules: if the premisses are true, the conclusions are guaranteed to be true.

To understand the difference, one can consider the TLA rule of *temporal generalisation* (TG): $A/\Box A$. If A is a theorem of the logic, then so is $\Box A$. This is certainly valid – if one can prove something in TLA, then it is a temporal-logical validity and thus holds everywhere; at all states in all models. Therefore so does $\Box A$.

Now consider the same rule used for the modality $\Box_A$. Could we use $A/\Box A$ (*Necessitation*) as a rule of inference in EL? Suppose A is any statement that happens to be true. If we were to use *Necessitation* as a rule of inference, we would be able to infer $\Box A$. That is, any truth is a necessary truth. But that means that all alternative worlds have exactly the same truths as this one, and that would mean that any sentence of the form $(A \Rightarrow B)$ is true at all worlds including the real world, or none (also including the real world) which in turn would reduce the meaning of $A \,\Box\!\!\rightarrow B$ simply to $A \Rightarrow B$, which would destroy our enterprise.

EL is a natural deduction system. One may make assumptions, and *discharge* some of those assumptions according to certain inference rules. Or leave them in. Every step in an EL proof is taken to be true; either true because assumed so, or true because inferred from other statements. At the end of a proof, it is convenient to list all assumptions. The constructor of the proof bears the responsibility of ensuring that the assumptions are indeed all true.

20.15 VCU-EL Semantics Illustrated

We illustrate the semantics of the logic *VCU-EL* by means of the example in Figure 20.7.

We see in Figure 20.7 ten 'possible worlds'. One, in the center, is the '`Real World`', and the other possible worlds are called `World 1` through `World 9`. Each possible world consists of a temporal-logic model, a *behavior*, which consists of an unending discrete linear sequence (the first three are indicated) of models for

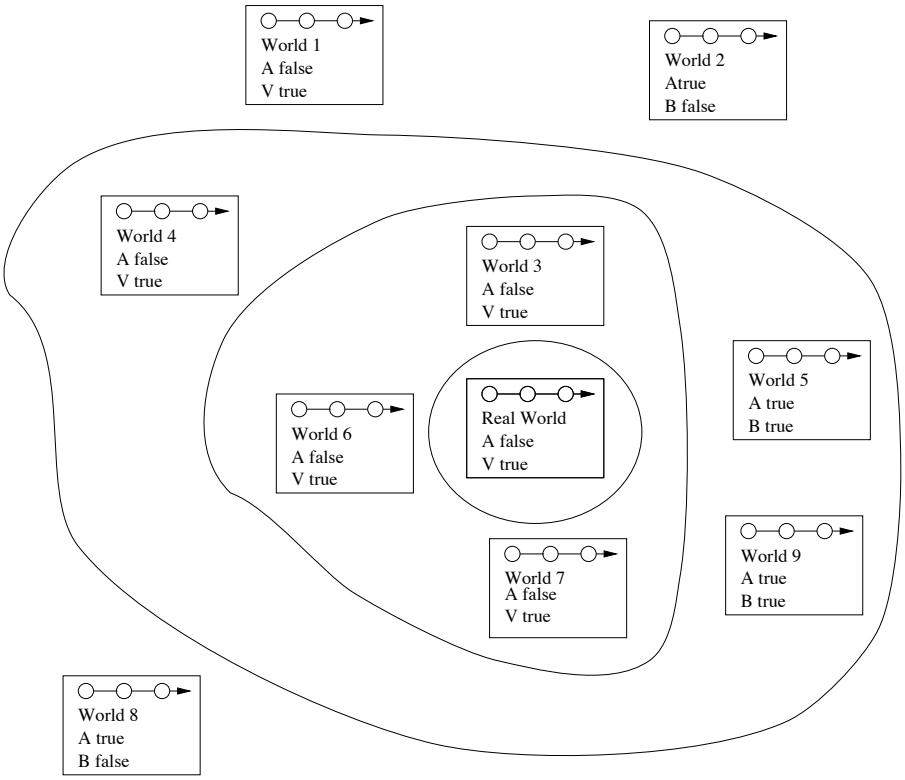

Figure 20.7: Illustration of the Lewis Semantics

normal first-order logic in which all the variables, constants and state predicates are interpreted. A state in the behavior, represented by a tiny circle, is one of these first-order models. The sequence represents the succession of states in the behavior, and an action holds between two adjacent states just in case the action formula is true of the pair of two states, with 'normal' variables interpreted in the first state and primed variables in its successor. We won't be illustrating the TLA semantics by means of Figure 20.7.

The relation $\preceq$ `Real World` is illustrated by the 'nearness rings', indicating the equivalence classes of possible worlds under the equivalence relation generated by $\preceq$ `Real World`. There are four 'nearness' rings, that is, the nearest world to the `Real World` is the `Real World` itself. The next, equal, nearest are `World 3`, `World 6` and `World 7`. Further away than these are (equally) `World 4`, `World 5` and `World 9`. Finally, equal furthest away are `World 1`, `World 2` and `World 8`.

Let us now determine the truth of $A \mathbin{\Box\!\!\rightarrow} B$ at `Real World` in this model of VCU-EL. The semantics says that we must find the nearest world in which A is true, and evaluate $A \Rightarrow B$ on all worlds at that distance or nearer.

In the `Real World`, A is false, as it is in `World 3`, `World 6` and `World 7`. In the next ring, however, containing `World 4`, `World 5` and `World 9`, there is a

world – in fact two, `World 5` and `World 9` – in which A is true. So this nearness ring bounds the set of worlds we have to look at to determine the truth of $A \boxdot\!\!\rightarrow B$ at `Real World`: namely we need to evaluate $A \Rightarrow B$ in `Real World`, `World 3`, `World 6`, `World 7`, `World 4`, `World 5` and `World 9`. In fact, in all of these worlds, $A \Rightarrow B$ is true, because either A is false, or B is true. Hence $A \boxdot\!\!\rightarrow B$ is true at `Real World`.

Notice that there are two worlds, `World 2` and `World 8`, in which $A \Rightarrow B$ is false, because A is true and B is false. However, because these worlds are outside the nearness ring in which the nearest worlds with A true lie, they play no role in assessing the truth of $A \boxdot\!\!\rightarrow B$.

However, they do play a role in assessing the truth of $A \succ B \triangleq \Box_A(A \Rightarrow B)$. For this to be true, $A \Rightarrow B$ must be true at *all* possible worlds, and since $A \Rightarrow B$ is false at `World 2` and `World 8`, $\Box_A(A \Rightarrow B)$ is false.

In all the worlds in which A is false, namely `Real World`, `World 1`, `World 3`, `World 4`, `World 6`, and `World 7`, V is true. Hence $\neg A \Rightarrow V$ is true at all worlds, i.e., $\Box_A(\neg A \Rightarrow V)$ is true, and since $O(A) \triangleq \Box_A(\neg \Rightarrow V)$, this means that $O(A)$ is true.

Notice that we have spoken of sentences $\Box_A(...)$ and $O(...)$ as being true or false, rather than true-at-`Real World`, false-at-`Real World`. That is because evaluation of the operator involves looking at *all* worlds to see whether something is true in each. So if a sentence $\Box_A(...)$ and $O(...)$ is true at any world, it is true at all worlds. So we just might as well say 'true' or 'false'. However, $A \boxdot\!\!\rightarrow B$ might very well be true at the `Real World`, as here, and false at some other world, say `World 1`, because the nearness rings of `World 1` do nto necessarily bear any relation to the nearness rings of `Real World`.

20.16 Extensions and Modifications

The EL logic has stayed moderately stable during its use to perform the proof of the example. It is, however, quite possible that the rules will need to be modified somewhat in light of other needs. We do not rule out this possibility. For example, one may consider it worthwhile to drop the $\Diamond$ in the consequent of hypothesis and conclusion to Rule 14.23 (also Rule 14.22).

20.16.1 Giving Priority to Causal Factors

During his consideration of the counterfactual analysis of causal statements, John Mackie has argued in [Mac74, Ch. 2] for ways of assigning greater importance to some sorts of causal factors rather than others. Suppose a factor is normally present in similar circumstances to those of the accident circumstance, but in which there was no accident, then even though it might have been a necessary factor fulfilling the counterfactual semantics, we do not count it as such. For

example, we do not usually count it a causal factor of an aircraft landing accident that the aircraft took off some hours before; or that the aircraft was built at all. Nevertheless, these events both satisfy the counterfactual semantics for being a factor. Mackie uses the term *causal field* for the collection of such features.

We have not pursued any attempt to distinguish amongst causal factors except for distinguishing those which themselves are not considered to have any factors from those which are explained by other factors. This comes down implicitly to considering the causal field in Mackie's sense, and we have provided no guidelines so far as to how to do so. This would be an obvious extension of WBA.

20.16.2 Closed World Assumptions

Accident reports use a closed-world assumption (CWA) , namely that either all the significant events and states are known, or those that are not known are known to be not known. Both the CWA and other non-monotonic reasoning can be expressed in the ontology introduced above. WBA does not necessarily suffer from this weakness (should it be considered to be a weakness), since one starts a WBA from certain facts in temporal order, and completes an explanation to a desired degree of granularity. Either one has all the facts needed for that explanation, or one becomes aware of ignorance and uses the PAD approach to signal the indeterminacy. However, implicitly a CWA is still made. One may determine that the reason an aircraft crashed was an inflight breakup initiated by a weak structural member, but one does not entertain the possibility that there were little green men sitting on the wing pulling it apart. That is a closed world assumption – that one does not and can not consider all the possible events that could have caused a given event. One identifies only those which one has reason to suspect were there. One can consider this a form of Occam's Razor if one likes, but it is a CWA of some sort.

Intuitively, it seems to make little sense to worry about this sort of non-monotonicity. This kind of assumption seems to be endemic to any form of causal explanation of historical events.

20.16.3 Other Non-Monotonicity

In principle, the 'world' consists only of states or events obtained *directly* from instruments like cockpit voice recorder (CVR) and digital flight data recorder (DFDR, 'black box'); photographs; on-site investigation of wreckage; states, events or processes derivable by temporal, causal and deontic reasoning from these. Formally, for every 'new' node (representing new knowledge of one of these states, events or processes) we introduce in our analysis, we have to check whether former reasoning is still valid (there are thus two cases: simple incompleteness and non-monotonicity - see below). Whenever we make an *assumption* about a cause for a state/event/process, we limit the explanatory power of the system to

explanations which fulfil this assumption. To keep this limitation within bounds (we prefer to base analysis on formal argumentation rather than speculation), it would make sense formally to clone the 'existing' world before we introduce the new information, as in the method of semantic tableaux. We would need to control the potential exponential growth of the number of worlds to consider. Alternatively, we can be content with justifying 'reasonable' assumptions and ignore alternatives, but we may have to be prepared to revise these in light of further discovery (non-monotonicity). Examples:–

Cali (incompleteness, monotonic reasoning):
DFDR recordings show that the machine turned left for 90 seconds. This could not be explained, until an undamaged FMC was discovered and its non-volatile memory decoded. In this case, the WB-method would yield an incomplete, but causally correct graph, which contains all information discovered, but not including grounds for the left turn. The additional information gleaned from the FMC several months after the accident can be introduced to 'complete' the graph. Such 'completions' result in addidtional subgraphs, but do not change the rest of the graph.

Lauda Air, Thailand (assumption, non-monotonic):
Evidence from CVR that reverse thrust (RT) was 'deployed'; but there's an interlock.
Conclusion: upset cannot be directly explained. Subsequently found a failure mode of the interlock, which in principle could allow RT to actuate in flight. Report contains no probable cause, but considers this to be a likely scenario.

Mont Ste. Odile, Strasbourg (assumption, non-monotonic):
Autopilot modes not available on DFDR; flight path shows rapid descent starting exactly at FAF. Descent rate in fpm is almost identical with required flight path angle in degrees; also the autopilot descent mode would have been engaged at FAF, where divergent behavior started. Autopilot mode control is unlabelled toggle; mode annunciation is via small letters, rate/angle larger figures. Again, this 'likely cause' is presumed.

Summary: All accident reports make a CWA: the relevant facts are those we know plus those we know we don't know. Assumptions about 'likely happenings' introduce either an extra (formal) modal dimension or non-monotonicity.

20.16.4 Casual Defeasibility

It turns out that in practical use in accident reports, the notion of cause is defeasible by properties of social institutions. For example, an instrument approach is designed to be used in conditions of low or absent visibility by all pilots; a pilot is obligated to decline the approach, or to break it off, if it cannot be flown 'safely' (within hisher abilities). Suppose an aircraft crashes into terrain on approach while under full control (so-called CFIT accidents, for example the

August 1997 Korean Air accident in Guam). This happens mostly, if not entirely, in conditions of low or absent visibility (night, cloudy weather). According to the counterfactual semantics, this absent visibility is a cause (*had* the pilot been able to see where (s)he was, (s)he would have avoided the terrain). But according to accident investigators, it is *not* a cause, because the approach procedure was designed to be safe in precisely these conditions. So certain institutional facts defeat the obvious physical causality. Is this always so? No – had the approach procedure not been designed to be safe (say, there was none, but the pilot was trying anyway), the low visibility remains a causal factor – to which is added the pilot misjudgement.

We do not know yet how to handle this non-monotonicity in WBA.

20.16.5 Summary

The notion of causal field and assigning weights to causal factors, considered in Section 20.16.1, will need to be addressed explicitly in the further formal development of WBA. The form of CWA considered in Section 20.16.2 is not one that we imagine we shall ever worry about in WBA. We illustrated how to use PADs in Chapter 17 to handle the kinds of indeterminacy considered in Section 20.16.3; whether this method always suffices for these kinds of indeterminacy will be determined by experience. However, the defeasibility recounted in Section 20.16.4 is significant, and WBA cannot yet handle it. We consider this to be a weakness, but have no remedy to hand at present. We are loathe to build in any of the current approaches to defeasible reasoning because of their logical complexity and because of the problems with formal defeasible reasoning that have not yet been solved. It does appear, though, that at some time some version will have to be included.