CHAPTER 4

THE ANALYSIS OF NON-DEDUCTIVE REASONING

— in which the logical tool of a consequence relation is introduced through the work of Kraus, Lehmann & Magidor —

THE GOAL OF the investigations reported in this thesis is to commence a logical analysis of inductive reasoning. We will therefore need to have a good overview of what logic has to offer regarding the analysis of other forms of reasoning. This chapter is meant to provide such an overview.

Traditionally, logicians have concentrated on deductive or mathematical reasoning as their main object of study. It is certainly no exaggeration when I say that, in a not too distant past, most logicians were of the opinion that logic is **necessarily** deductive — and many of them probably still are[37]. This means on the one hand that the study of deductive reasoning has provided us with some rather advanced tools; but it also means that these tools, being tailor-made for analysing deduction, may be not immediately applicable to other forms of reasoning. I will try to present the logical theory of deductive reasoning currently in force with an eye for possible adaptations of that theory towards modelling such alternative forms of reasoning.

Conceivable alternatives include 'approximations' of deductive reasoning, such as plausible reasoning, as well as forms of reasoning such as induction, which are rather different from deductive reasoning. The former type of reasoning will be called *quasi-deductive*, while the latter will be called *a-deductive*. Jointly, they will be referred to as *non-deductive* reasoning.

The subject of plausible reasoning — or nonmonotonic reasoning, as it is usually called — started to attract attention from artificial intelligence researchers, and later logicians, some fifteen years ago. It is the first non-deductive form of reasoning that has been submitted to a relatively advanced logical analysis. Recently, this analysis has reached a stage in which it is sufficiently general so as to be applicable, at least in

[37] In the words of Dov Gabbay: 'some members of the traditional logic community are still very conservative in the sense that they have not even accepted non-monotonic reasoning systems as logics yet. They believe that all this excitement is transient, temporarily generated by computer science and that it will fizzle out sooner or later. They believe that we will soon be back to the old research problems, such as how many non-isomorphic models does a theory have in some inaccessible cardinal or what is the ordinal of yet another subsystem of analysis. I think this is fine for mathematical logic but not for the logic of human reasoning. There is no conflict here between the new and the old, just further evolution of the subject.'
(Gabbay, 1994, p.368, note 7)

principle, to other forms of non-deductive reasoning. The formal framework of consequence relations, developed by Kraus, Lehmann and Magidor (1990), will form the backbone of the logical analysis of induction attempted in this thesis, providing ample justification for the quite comprehensive elaboration of that framework in the present chapter.

§14. DEDUCTION: THE LOGIC OF SATISFACTION-PRESERVATION

Logic is the formal study of reasoning: the process of drawing conclusions from premisses. The unit of reasoning is the *argument*. In its most simple form, an argument consists of a set of *premisses*, and a *co*[redacted]*lusion*:

all human beings are mortal

Socrates is a human being

∴ Socrates is mort[redacted]

Th[redacted] the pr[redacted]ses from the co[redacted]lusion.

[redacted]his argument is *ded*[redacted]*ive*: its dist[redacted] is gu[redacted]anteed by the truth [redacted]s of a *truth-valued se*[redacted]r w[redacted]h a formula is true or false[38]. Su[redacted]et of *i*[redacted]*pretations*, rep[redacted]of aff[redacted]rs, and a *satisfact*[redacted]an int[redacted]pretation m satisfie[redacted] formula, and write m [redacted]φ. If all models of [redacted] of φ, we say that Σ *logically en*[redacted]*ls* φ or φ is a *logic*[redacted]*nsequence* [redacted]Σ, and write Σ [redacted]φ. If Σ is empty, φ is called a *tautology*; instead of ∅[redacted]φ, we write [redacted]. A formula [redacted]s a *contradiction* if ¬ψ is a tautology. I will refer to [redacted] the *standard* [redacted]mantics of deduction.

It is often said that d[redacted]uction is *truth-preserving*, in the sense that the conclusion of a deductive argument is true whenever the premisses are. Defining a formula to be *true* if it is satisfied by a distinguished *intended interpretation*, we see that the idea of truth-preservation is indeed captured by the standard deductive semantics. However, it seems more accurate to say that deduction is *satisfaction-preserving*: the conclusion of a deductive argument is satisfied by any interpretation satisfying the premisses[39]. Clearly, satisfaction-preservation implies truth-preservation, but the former generalises the latter by quantifying over all possible interpretations. Furthermore, deductive arguments are meaningful even if their premisses are false.

---

[38] A precise definition of this semantics is given in section 1.2.

[39] What I call ‘satisfaction’ some others may prefer to call ‘truth in a model’. However, the main point is that there is a distinction between *absolute* definitions of truth, and relative or *model-theoretic* definitions of truth (Haack, 1978, p.108). By using the more neutral term ‘satisfaction’ I hope to avoid confusion. What I want to stress here is that, whenever we reason about the ‘real world’ (which we often do), we would be satisfied with mere truth-preservation. Deduction offers us something much stronger!
See also (Etchemendy, 1990), especially Ch.2 and p.49.

§14. Deduction: the logic of satisfaction-prese[redacted]ation

Although logical entailment is a well-defined notion, t[redacted] question [redacted]hether a certain Σ logically entails a certain φ can be quite hard to answer. T[redacted]is is espec[redacted]lly true when the domain of interpretation is i[redacted]finite (as in the case of the [redacted]gers), s[redacted]ce in that case Σ may have an infinite numb[redacted] of models. *Proof theory* [redacted] method for calc[redacted]ting the re[redacted]n of logical entailment. [redacted]ic i[redacted] is to set a[redacted]rt a set of axioms, an[redacted]a set of in[redacted]ce rules that transform o[redacted]ore g[redacted]ven formul[redacted] into a new formula; a[redacted]oms and i[redacted]nce rules are jointly cal[redacted] proof procedure [redacted], we say t[redacted] is *provable* from Σ an[redacted] finite sequence o[redacted]ormulas φ[redacted]cces[redacted]e applicati[redacted]s inference rules [redacted] axioms, [redacted] combinations of [redacted]ese, whil[redacted] is the conclusion φ. Suc[redacted] exists, is called a[redacted]write [redacted] explicit referen[redacted] to particular proof p[redacted]cedure.

The question [redacted]hether [redacted]ually the same relation [redacted]es us to the *meta*[redacted] of deductive logi[redacted] We say th[redacted]proof procedure is *sound* i[redacted]henever [redacted] *complete* if Σ [redacted]whenever [redacted]φ. If a proof procedure can be [redacted]n to be [redacted]th [redacted] and complete, th[redacted]s called a *re*[redacted]*esentation theorem*. Representatio[redacted]theorems [redacted] obtained for a[redacted]iety of logical languages and proof theories, [redacted]cluding [redacted]rst-ord[redacted] pre[redacted]cate logic[redacted]urther metatheoretical notions include decidabilit[redacted]nd comp[redacted]tness. [redacted] log[redacted] is *decid*[redacted]*e* if there exists a[redacted] effective procedure for deci[redacted]ng wheth[redacted] Σ Pr[redacted]ositional l[redacted]ic is deci[redacted]ble (si[redacted] the number[redacted] models o[redacted] fo[redacted]nula, i.e. [redacted]uthval[redacted] ass[redacted]nments to[redacted]ropositio[redacted]toms[redacted]s always [redacted], whereas [redacted] tu[redacted]the[redacted]re proc[redacted] that produ[redacted]roof wh[redacted]ever one ex[redacted]s, but no p[redacted]cedure is[redacted] halt a[redacted]inate wit[redacted]e if no p[redacted]f exists. A [redacted]gic is *comp*[redacted] iff every[redacted]subset of[redacted] is (a set of [redacted]ula φ it [redacted]case that [redacted]and Σ [redacted]). First-or[redacted] to pro[redacted] several co[redacted]leteness r[redacted]lts in ch[redacted]

[redacted]learly, [redacted]a relation[redacted]equivalent to [redacted] enjoys a [redacted]alth of sp[redacted]

inclusio[redacted] if $\varphi \in \Sigma$, [redacted]
transitivity: if Σ [redacted] and Φ[redacted] then Σ [redacted]
monotonicity: if Σ [redacted]$\{\psi\}$ [redacted]all ψ;
deduction theorem:[redacted]ff Σ [redacted]
proof by refutation: Σ[redacted]ff Σ[redacted]$\{\neg\varphi\}$ [redacted]nconsistent.

I will refer to such properties a[redacted]*met*[redacted]*evel* pr[redacted]erties, since they can be used to reason *about* deduction as a form of r[redacted]soning. Clearly, some of these properties belong to the tool kit of every mathematician.

To wrap up the foregoing discussion: deductive reasoning can be analysed on three distinct levels. The semantic level is the most basic but the least practical, and formalises the idea that deduction is *truth-preserving*: the conclusion of a deductive argument is true whenever the premisses are. On the proof-theoretical level, one is concerned with the question how to actually *derive* the conclusion from the premisses. Finally, on the meta-theoretical level one can reason *about* deductive reasoning as such, establishing e.g. the completeness of a particular proof procedure, or the fact that deductive conclusions drawn on the basis of partial knowledge remain valid when the whole of available knowledge is taken into account.

Non-deductive logics

Clearly, a non-deductive logic is not truth-preserving. This means that every once in a while one draws a conclusion that turns out to be false. If a conclusion is refuted by further observations, we say that the conclusion is *defeated*; non-deductive logics are likewise called *defeasible*. Clearly, in a defeasible logic one must use all the information available to derive a conclusion, and one must always be prepared to give up earlier conclusions.

A better understanding of all these issues is obtained when non-deductive logics are analysed on the metalevel. Thus, whereas for deductive logic semantics and proof theory are the pillars upon which the metatheory is built, for non-deductive logics it seems to be just the other way around: we start by investigating which properties a certain non-deductive logic has or does not have, and only then we devise a semantics which precisely matches that specific set of properties. The process of devising a semantics matching a set of properties is called a *characterisation*.

The remainder of the present chapter will be devoted to the development of a metatheory of plausible reasoning, and a semantic characterisation of that metatheory, as put forward in a seminal paper by Sarit Kraus, Daniel Lehmann and Menachem Magidor, published in 1990 in the *Artificial Intelligence* journal. In that paper, the authors set out to "study general patterns of nonmonotonic reasoning and try to isolate properties that could help us map the field of nonmonotonic reasoning by reference to positive properties". Their approach will form a model for much of the foundational work on induction presented in this thesis.

§15. THE KLM FRAMEWORK FOR PLAUSIBLE REASONING

The need for a framework such as that developed by Kraus *et al*. (henceforth referred to as the *KLM framework*) is nowhere demonstrated more clearly than in the phrase 'nonmonotonic reasoning'. Usually, the kind of reasoning studied under that name is exemplified by the prototypical ornithological argument

typically, birds fly

Tweety is a bird

∴ Tweety flies

Clearly, such arguments are defeasible, and the corresponding kind of reasoning is nonmonotonic — yet so are virtually all non-deductive forms of reasoning. For instance, the inductive conclusion 'all swans are white' is defeated by the observation of a black swan, which demonstrates that induction is nonmonotonic as well. It is a regrettable fact that 'nonmonotonic reasoning' is used most frequently in the sense of 'plausible reasoning', whereas it is such a wider (and better-defined) term. To avoid confusion, I will use the term 'plausible reasoning' whenever I mean plausible reasoning.

Plausible consequence relations

Following Gabbay (1985), Kraus *et al.* focus the study of plausible reasoning on the level of consequence relations, where a consequence relation is a set of arguments.

> DEFINITION 4.1. Let L be a propositional language. A *consequence relation* $\mid\!\sim\ \subseteq L\times L$[40] is a set of pairs of formulas of L. Elements of a consequence relation are called *arguments*[41]; instead of $\langle \mathtt{p},\mathtt{q}\rangle \in\ \mid\!\sim$ we write $\mathtt{p} \mid\!\sim \mathtt{q}$. The left-hand formula is called the *premiss* of the argument (if it is a conjunction, each of the conjuncts is also called a premiss), and the right-hand formula is called its *conclusion*.

The intended interpretation of $\mathtt{p} \mid\!\sim \mathtt{q}$ is 'if $\mathtt{p}$, normally $\mathtt{q}$' or '$\mathtt{q}$ is a plausible consequence of $\mathtt{p}$'.[42]

It is the aim of Kraus *et al.* to develop sensible axiomatisations of the binary relation $\mid\!\sim$, each characterised by a suitable semantics. To this end, they employ a metalanguage containing a binary predicate $\mid\!\sim$ (written infix for convenience), variables ranging over formulas of L (denoted by Greek letters from the beginning of the alphabet), and constants referring to formulas of L (for convenience, these constants are simply the formulas themselves — `typewriter` font refers to formulas from L).

Kraus *et al.* choose a Gentzen-style notation of axiom schemata and inference rules to express structural properties of $\mid\!\sim$. An example of an axiom schema is

- **Reflexivity**: $\alpha \mid\!\sim \alpha$

Here, α is a variable in the metalanguage, ranging over formulas of L. A consequence relation satisfies such an axiom schema if it contains all instances of it (an instance is obtained by replacing the metavariable α with a formula of L). For example, if `bird` is a proposition of L then a consequence relation satisfying Reflexivity contains the argument `bird` $\mid\!\sim$ `bird` ('birds normally are birds').

An example of an inference rule is

- **Cautious Monotonicity**: $\dfrac{\alpha \mid\!\sim \beta\ ,\ \alpha \mid\!\sim \gamma}{\alpha\wedge\beta \mid\!\sim \gamma}$

The metalevel formulas above the line are the *antecedents* of the inference rule, and the formula below the line is its *consequent*. A consequence relation satisfies an inference rule if it is closed under that rule, that is, whenever it contains instances of all the antecedents

[40] A set of premisses Σ is treated as the conjunction of its elements; obviously, this requires Σ to be finite.

[41] Kraus *et al.* use the phrase 'conditional assertion'; I prefer a term which is also applicable to forms of reasoning other than plausible reasoning.

[42] It should be noted that the two intended interpretations are quite different: the first statement is reminiscent of material implication (if …, then …), while the second is analogous to logical entailment. Kraus *et al*. do not make this distinction, which is, I believe, a mistake — see §16.

(such that th[redacted]same va[redacted]ble is replaced by the same formula of *L* throughout), it a[redacted]o contains the [redacted]e consequent. For example, if a consequen[redacted]e relation cont[redacted]s $\mathtt{bird}$[redacted]$\mathtt{feathers}$ and $\mathtt{bird} \mathrel{|\!\sim} \mathtt{flies}$, then it satisfies Cauti[redacted] Monotonicity[redacted]nly if it [redacted]$\mathtt{ers} \mathrel{|\!\sim} \mathtt{flies}$ ('birds with feath[redacted]s normally fly'[redacted]

Deductiv[redacted]rgument[redacted]like 'penguins are birds' ca[redacted]lso be incorporated in the KL[redacted] framework. T[redacted]are n[redacted]on, but collected separately i[redacted] background t[redacted] *T*. If[redacted]enguins are birds' is known [redacted]om the background theory, this[redacted] written as *T* [redacted]$\mathtt{nguin}\rightarrow\mathtt{bird}$, as usual. In fact, [redacted] expli[redacted]t background the[redacted]y, the [redacted]LM framework utilises a res[redacted]cted set of models *U*, so that [redacted] means 'for all *m*[redacted]is technic[redacted] and for all practical purposes *U* can be thought [redacted]as being the set of models of *T*.

Deductive a[redacted]uments can be 'lifted' to the meta[redacted]vel, where they can be used in inference rules s[redacted]h as the following:

- **R[redacted]ht Weakening**: [redacted] $\dfrac{\alpha\rightarrow\beta\ ,\ \gamma \mathrel{|\!\sim} \alpha}{[redacted] \mathrel{|\!\sim} \beta}$

For instance, if [redacted]$\mathtt{flies}\rightarrow\neg\mathtt{penguin}$ ('no penguin flies')[redacted]nd the consequence relation contains $\mathtt{bird}$[redacted]$\mathtt{flies}$, then it satisfies Right Weake[redacted]ng only if it also contains $\mathtt{bird} \mathrel{|\!\sim} \neg\mathtt{penguin}$ ('birds are normally not penguins').

Rule systems for plausible reasoning

Axiom schemata and inference rules can be combined t[redacted] instance, by putting $\gamma=\alpha$ in Right Weakening and applyin[redacted]eflexivity, one obtains the following derived rule:

- **Entailment**: $\dfrac{[redacted]\alpha\rightarrow\beta}{\alpha \mathrel{|\!\sim} \beta}$

What this derivation tells us is that a consequence relation satisfying Reflexivity and Right Weakening contains at least all deductive arguments. Furthermore, note that Reflexivity can be obtained from Entailment by putting $\beta=\alpha$, which implies that in the presence of Right Weakening, Reflexivity and Entailment are equivalent, and only one of them needs to be included if both properties are required. Part of the KLM framework aims at finding elegant axiomatisations or *rule systems* for different sets of required properties.

Kraus *et al.* define five different rule systems, three of which are of immediate interest to us here. In order of increasing strength, these are the systems **C** (for cumulative reasoning), **P** (preferential reasoning) and **M** (monotonic reasoning). **P** is strictly stronger than **C** (and **M** is strictly stronger than the other two) in the following sense: every preferential consequence relation is cumulative, but some cumulative consequence relations are not preferential. Consequently, every preferential consequence relation satisfies all the properties of cumulative consequence relations, and the system **P** can be obtained from **C** by adding some additional rules.

We summarise the main definitions and results concerning these three rule systems below. The weakest rule system **C** contains, according to Kraus *et al.*, the minimal conditions under which a consequence relation can still [redacted]rm of plausible reasoning.

DEFINITION 4.2 (KLM 3.1[43]). A consequenc[redacted] *cumulative* iff it satisfies the following axiom sc[redacted]ma and inference rules:

- **Reflexivity**: [redacted]
- **Left Logical Equivalence**: $\dfrac{[redacted]\leftrightarrow\beta\ ,\ \alpha \mid\!\sim \gamma}{\beta \mid\!\sim \gamma}$
- **Right Weakening**: $\dfrac{[redacted]\rightarrow\beta\ ,\ \gamma \mid\!\sim \alpha}{\gamma \mid\!\sim \beta}$
- **Cut**: $\dfrac{\alpha\wedge\beta \mid\!\sim \gamma\ ,\ \alpha \mid\!\sim \beta}{\alpha \mid\!\sim \gamma}$
- **Cautious Monotonicity**: $\dfrac{\alpha \mid\!\sim \beta\ ,\ \alpha \mid\!\sim \gamma}{\alpha\wedge\beta \mid\!\sim \gamma}$

Together, these rules constitute the system **C**.

One of the derived rules of **C** is the following (KLM 3.3):

- **And**: $\dfrac{\alpha \mid\!\sim \beta\ ,\ \alpha \mid\!\sim \gamma}{\alpha \mid\!\sim \beta\wedge\gamma}$

The next rule system, **P**, takes the central position in the KLM framework.

DEFINITION 4.3 (KLM 5.1). A consequence relation $\mid\!\sim$ is said to be *preferential* iff it satisfies the rules of **C** and the following:

- **Or**: $\dfrac{\alpha \mid\!\sim \gamma\ ,\ \beta \mid\!\sim \gamma}{\alpha\vee\beta \mid\!\sim \gamma}$

Together, these rules constitute the system **P**.

One of the derived rules of **P** is the following (KLM 5.2):

- **S**: $\dfrac{\alpha\wedge\beta \mid\!\sim \gamma}{\alpha \mid\!\sim \beta\rightarrow\gamma}$

It can be shown (KLM 5.3) that an alternative axiomatisation of **P** is obtained by replacing Cut with And, resulting in the following set of properties: Reflexivity, Left Logical Equivalence, Right Weakening, And, Cautious Monotonicity, and Or.

Finally, the third rule system **M** is the strongest of the three. As we will see below it

[43] This refers to Definition 3.1 of (Kraus *et al.*, 1990).

models monotonic, deductive reasoning.

> DEFINITION 4.4 (KLM 7.1). A consequence [redacted]lation $\mathrel{|\!\sim}$ is said to be *monotonic* iff it satisfies the rules of **C** and the [redacted]lowing:
>
> - **Contraposition**: [redacted]
> $\neg\beta \mathrel{|\!\sim} \neg\alpha$
>
> Together, these rules constitute the system **M**.

An alternative axiomatisation (KLM 7.3) can be obtai[redacted]d from **P** by replacing Cautious Monotonicity with

- **Monotonicity**: [redacted]$\to\beta\ ,\ \beta \mathrel{|\!\sim} \gamma$
  $\alpha \mathrel{|\!\sim} \gamma$

Since Left Logical Equivalence is implied by Monotonicity, this results in the following set of properties: Reflexivity, Right Weakening, And, Monotonicity, and Or. Furthermore, we notice that Or could also be replaced with rule S (KLM 7.3).

Characterisation of plausible consequence relations

Each of the rule systems in the KLM framework is characterised by an appropriate semantics. Obviously, the weakest system **C** requires the most elaborate semantics, and each of the stronger systems simplifies this semantics in some respect. The basic idea of the semantics for **C** was already proposed by Shoham (1987), who introduced a partial ordering of preference between interpretations, stipulating that α plausibly entails β if every most preferred model of α satisfies β.

However, for cumulative reasoning the preference relation compares sets of models rather than single models. A further technicality is that the same set of models may appear at more than one place in the ordering. Therefore, preference is expressed between abstract states, each of which is labelled with a set of models (recall that U is a set of models, expressing an implicit background theory)[44].

> DEFINITION 4.5 (KLM 3.10). A *cumulative* structure[45] is a triple $\langle S,l,<\rangle$, where S is a set of *states*, l: $S\to 2^U$ is a function that labels every state with a nonempty set of models, and $<$ is a binary relation[46] on S.

---

[44] It should be noted that a state is fully determined by the set of models labelling it and its place in the preference ordering, and thus needs no further definition. In the words of Kraus *et al*.: 'We shall not define further the notion of a state, but suppose that every state is, in a [structure], labeled with a set of [models] (intuitively the set of all [models] the reasoner thinks are possible in this state)' (p.181).

[45] What I call a structure is called a model by Kraus *et al*., and what I call a model they call a world. I chose to change terminology because I prefer to use the term 'model' in its classical sense.

[46] $<$ is not necessarily a partial order, but it should satisfy a certain 'smoothness condition', which is for instance satisfied if $<$ does not have infinite descending chains.

Every cu[redacted]nsequence relation, as follows.

> D[redacted]INITION 4.6 (KLM 3.11, 3.13). Let $W = \langle S,l,<\rangle$ be a cumulative str[redacted]ture. A state $s \in S$ *satisfies* a formula $\alpha \in L$ iff for every model $m \in l(s)$, m [redacted]α; the set of states satisfying α is denoted by $[\alpha]$. The consequence rel[redacted]on defined by W is denoted by $\mathrel{|\!\sim}_W$ and is defined by: $\alpha \mathrel{|\!\sim}_W \beta$ iff every state minimal (wrt. $<$) in $[\alpha]$ satisfies β.

It is relatively easy to show that the consequence relation defined by a cumulative structure is in fact cumulative, that is, it satisfies the rules of **C** (KLM 3.16). Th[redacted] soundness result for cumulative consequence relations.

The corresponding completeness result requires that, for an arbitrary cumulativ[redacted] sequence relation $\mathrel{|\!\sim}$, there exists a cumulative structure W such that $\mathrel{|\!\sim}_W$ coincides [redacted] Briefly, the construction is as follows. Define an equivalence relation ~ by: $\alpha{\sim}\beta$ iff [redacted] and $\beta \mathrel{|\!\sim} \alpha$, and let S be the set of equivalence classes of formulas under ~. Further[redacted] each equivalence class is labelled with the set of normal models for the formulas [redacted] class, where a model $m \in U$ is a *normal* model for α iff it verifies all of its pla[redacted] consequences. Finally, s_1 is preferred over s_2 ($s_1 < s_2$) iff some formula in s_1 is a pla[redacted] consequence of some formula in s_2, and $s_1 \neq s_2$. The combination of soundnes[redacted] completeness gives us the following representation theorem.

> THEOREM 4.7 (KLM 3.25). *A consequence relation is a cumulative* [redacted] *consequence relation*[redacted]*f it is defined by some cumulative structure.*

The corresponding resu[redacted]s for **P** and **M** are obtained by constraining cumulative structures. A preferential str[redacted]ture is a cumulative structure where states are labelled with singleton sets of models, [redacted]r. We then essentially have a preference over models (ex[redacted]pt that the same model may label different states — see (Kraus *et al.*, 1990, p. 193) f[redacted] an example why this additional freedom is needed).

> DEFINITION 4.8 (KL[redacted] structure is a triple $W = \langle S,l,<\rangle$, where S is a set of st[redacted]es, l: $S \to U$ is a function that labels every state with a model, and $<$ is a[redacted]trict partial order[47] on S. A state $s \in S$ satisfies a formula $\alpha \in L$ iff $l(s)$ [redacted]α; the consequence relation defined by W is denoted by $\mathrel{|\!\sim}_W$ and is defined [redacted] in Definition 4.6.

For proving completeness the following preferential structure is built. Let S be the se[redacted] pairs $\langle m,\alpha\rangle$ of models m and formulas α, such that m is a normal model for α. Ev[redacted]y state $\langle m,\alpha\rangle$ in S is simply labelled with m. Finally, a state $\langle m,\alpha\rangle$ is preferred over [redacted]-other state $\langle n,\beta\rangle$ iff $\alpha \vee \beta \mathrel{|\!\sim} \alpha$ (expressing that α is not less ordinary than β[48]) and m [redacted]β

---

[47] I.e., $<$ is irreflexive and transitive. In addition, $<$ should satisfy the ‘smoothness condition’.

[48] In the words of Kraus *et al.*: ‘Indeed, if we would conclude that α is true on the basis that either α or β is true, this means that the former is not more out of the ordinary than the latter’ (p.195). Notice that by Or and Reflexivity $\beta \mathrel{|\!\sim} \alpha$ is a sufficient condition for $\alpha \vee \beta \mathrel{|\!\sim} \alpha$.

(this is just to make the ordering irreflexive). We then have the following represen[redacted] theorem.

> THEOREM 4.9 (KLM 5.18). *A consequence relation is a preferential* [redacted] *consequence relation iff it is defined by some preferential structure.*

A *monotonic*[49] structure is a preferential structure in which the preference relation < is empty. This removes the need for the intermediate level of states, since different states labelled by the same model can be considered identical. In p[redacted]ctice, a monotonic structure is thus defined by a subset *W* of *U*, and the consequence re[redacted]ion defined by a monotonic structure is equivalent to logical entailment over *W*.

> DEFINITION 4.10 (KLM 7.4). A *monotonic* struct[redacted] consequence relation $\mid\!\sim_W$ defined by a monotonic str[redacted]ture *W* is defined by: $\alpha \mid\!\sim_W \beta$ iff every model $m \in W$ that satisfies α, also s[redacted]isfies β.

In order to prove completeness, we need to construct a[redacted] given consequence relation $\mid\!\sim$, such that $\mid\!\sim$ and $\mid\!\sim_V$ coincide[redacted]o this end, *V* is defi[redacted] follows:

$$V = \{m \in U \mid \text{for all } \alpha,\beta \in L\text{: if } \alpha \mid\!\sim \beta\text{, then } m \text{ [redacted]} \alpha \rightarrow \beta\}$$

That is, *V* consists of those models that verify every plausible argument as if it [redacted] deductive argument. It is not difficult to show that the consequence relation define[redacted] corresponds to $\mid\!\sim$ if the latter is monotonic.

> THEOREM 4.11 (KLM 7.5). *A consequence relation is a monotonic* [redacted] *consequence relation iff it is defined by some monotonic structure.*

This concludes our discussion of the characterisation results for the systems **C**, **P** and **M**.

§16. DISCUSSION

In this chapter, I have tried to show how the standard analysis of deductive logic can be adapted in order to accommodate for a non-deductive form of reasoning. The KLM framework provides a systematic study of different forms of plausible reasoning, and will be a model for my subsequent investigations of different forms of induction. As I will argue below, this is possible because the KLM framework is so flexible that it in fact provides a methodology for analysing any kind of reasoning. I will also discuss criteria for comparing the strength of different rule systems.

The KLM approach as a methodology of descriptive logic

Although Kraus *et al.* were only concerned with plausible reasoning, their choice of considering plausible consequence as a metanotion on top of a classical propositional language (rather than devising a special-purpose language such as default logic (Reiter,

[49] Kraus *et al.* call such structures *simple preferential.*

1980) or circumscription (McCarthy, 1980)) turns their approach into a **methodology** for analysing arbitrary forms of reasoning on the metalevel. Indeed, the formal analysis in Part II of this thesis provides a case in point: although many of the ideas underlying that analysis were conceived independently of the KLM framework, it was the paper by Kraus *et al*. which prompted the formalisation of those ideas by means of conjectural consequence relations worked out in chapters 6 and 7.

I would like to add that both Kraus *et al*.'s analysis of plausible reasoning and my analysis of induction should be considered, in my view, as belonging to a distinguished branch of logic which I will call *descriptive logic*. The aim of descriptive logic is to provide a catalogue of different forms of reasoning, and to study the distinguishing qualities of each of those forms of reasoning, in their own right as well as in relation to each other. If such a catalogue deals only with reasoning forms of type *X*, I call it a *descriptive logical theory* of type *X*. For instance, the KLM framework establishes a descriptive logical theory of plausible reasoning; my primary aim in this thesis is to provide a descriptive logical theory of induction. In order to justify this new terminology, I have to explain why the systematic study of reasoning forms is useful, and why it requires a separate branch of logic.

The first point is easily dealt with. One only needs to consult the scientific literature in order to find answers to questions like: What is deduction? What is induction? Is any non-deductive logic inductive? What is abduction? and what is its relation to induction? What constitutes a logical system? Throughout the philosophical and logical literature, one will find either embarrassingly few answers, or an equally embarrassing variety of proposed answers, to these questions. In my view, this state of affairs puts a threat to the scientific credibility of logic, and reduces logic to 'the formal study of whatever logicians choose to study'.

Be that as it may, one might argue that the ontology of reasoning forms should be (and is) treated in the field known as *philosophy of logic*. To quote Haack:

> '...among the characteristically philosophical questions raised by the enterprise of logic are these: What does it mean to say that an argument is valid? that one statement follows from another? that a statement is logically true? Is validity to be explained as relative to some formal system? Or is there an extra-systematic idea that formal systems aim to represent? What has being valid got to do with being a good argument? How do formal logical systems help one to assess informal arguments? (…) Is there one correct formal logic? and what might 'correct' mean here? How does one recognise a valid argument or a logical truth? Which formal systems count as logics, and why? Certain themes recur: concern with the scope and aims of logic, the relations between formal logic and informal argument, and the relations between different formal systems.'
> (Haack, 1978, p.1)

It perhaps only requires a bit of good will to see that the systematic study of reasoning types would fit into Haack's description of the main questions studied by philosophy of logic. What bothers me, however, is that the issue of non-deductive or non-standard logics seems to occupy a rather peripheral position within that field. Furthermore, and perhaps more importantly, it seems to be a (regrettable) fact of life that issues raised in 'philosophy of *Y*' are usually taken to be of limited significance by practitioners of *Y*. I do, therefore, strongly advocate the view that the systematic study of logic forms an integral part of logic itself, rather than a meta-science studying the things logicians do. One can draw a parallel with the study of computer programming languages: although initially computer programming was identified with imperative programming, the advent of other programming paradigms such as functional and logic programming prompted computer scientists to reflect upon the essence of computer programming and programming paradigms, rather than banishing that subject to 'philosophy of programming languages' (although the term sounds good!).

To reiterate the point: descriptive logic, the systematic study of reasoning forms, is a branch of logic whose significance, especially for artificial intelligence researchers, can hardly be overestimated. The work by Kraus *et al*. represents an important contribution to descriptive logic, providing a descriptive logical theory of plausible reasoning. I will take advantage of the inherent flexibility of their framework, by using their methods to construct a descriptive logical theory of conjectural consequence relations, thus providing a constructive proof for the proposition that the KLM approach represents a methodology of descriptive logic.

The pragmatics of consequence relations

If one accepts this methodological view of the KLM framework, the distinction between the object level (a propositional language) and the metalevel (the language of plausible arguments) seems crucial. As soon as the metalevel consequence symbol $\mid\!\sim$ is interpreted as an object level connective, the methodological view seems to vanish.

Ironically, Kraus *et al*. are not very clear on this point, as is demonstrated by their account of the pragmatics of plausible consequence relations.

> 'The queries one wants to ask an automated knowledge base are formulas (of *L*) and query β should be interpreted as: *Is β expected to be true*? To answer such a query the knowledge base will apply some inference procedure to the information it has. We shall now propose a description of the different types of information a knowledge base has.
> The first type of information (…) is coded in the universe of reference *U* that describes both hard constraints (e.g., dogs are mammals) and points of definition (e.g., youngster is equivalent to not adult). Equivalently, such information will be given by a set of formulas defining *U* to be the set of all [models] that satisfy all the formulas of this set.
> The second type of information consists of a set of conditional assertions[50] describing the soft constraints (e.g., birds normally fly). This set describes what we know about the way the world generally behaves. This set of conditional assertions will be called the knowledge base, and denoted by **K**.

[50] Called 'plausible arguments' in this chapter.

> [redacted]e third type of information describes our information about the specific [redacted]uation at hand (e.g., it is a bird). This information will be represented by [redacted]formula. (…)
> [redacted]r inference procedure will work in the following way, to answer query β. [redacted]of reference *U* and a specific situation [redacted]scribed by α, it will try to deduce (…) the conditional assertion $\alpha \mathrel{|\!\sim} \beta$ [redacted]om the knowledge base **K**. This is a particularly elegant way of looking [redacted]ference process deduces conditional assertions [redacted]tions.’ (Kraus *et al.*, 1990, pp. 173–174)

For ins[redacted]ce, given the ‘hard’ rule

[redacted]`enguin`→`bird`

and the ‘soft’ rules

`bird` $\mathrel{|\!\sim}$ `flies`
`penguin` $\mathrel{|\!\sim}$ ¬`flies`

the conditional assertion `penguin` $\mathrel{|\!\sim}$ `flies`, which would establish a kind of contradiction, can be derived using rules from **M**, but not using only rules from **P**[51]; Kra[redacted]*et*[redacted] say that the assertion is monotonically, but not preferentially, entailed by the kno[redacted]edg[redacted]base.

[redacted]ow[redacted]s may be considered ‘a particularly elegant way of looking at the inference pro[redacted]ss’ [redacted]some — but it is not the kind of inference process I am concerned with in this thes[redacted]*l.* interpret a given plausible consequence relation as [redacted]**no**[redacted]**edge base**, while I interpret it as a description of the behaviour of a par[redacted]ular[redacted]**easoning agent**. Under the first interpretation the symbol $\mathrel{|\!\sim}$ represents the con[redacted]t then one wonders why this connective cannot be nest[redacted]ent plausible implications cannot be combined by mea[redacted]of[redacted]her connectives (as in $\alpha \mathrel{|\!\sim} \gamma \vee \beta \mathrel{|\!\sim} \gamma$)? Under the second interpretation these que[redacted]ons[redacted]imply make no sense: the expression $(\alpha \mathrel{|\!\sim} \beta) \mathrel{|\!\sim} \gamma$ is just as meaningless as (α [redacted]β) [redacted]γ.

[redacted]e p[redacted]tion taken in this thesis can be summarised as follows:

(*i*) an argument like $\alpha \mathrel{|\!\sim} \beta$ describes part of the behaviour of a particular reasoning agent;
(*ii*) a metalevel rule like Cautious Monotonicity is a rationality postulate for certain kinds of reasoning;
(*iii*) a consequence relation closed under the rules of a certain rule system is a complete description of the behaviour of an agent performing a certain kind of reasoning.

For instance, if a plausible reasoning agent would accept the conclusions `flies` and

[51] The proof proceeds via Transitivity, a derived rule in **M** but not in **P**. To see that `penguin` $\mathrel{|\!\sim}$ `flies` is not preferentially entailed, consider the preferential model consisting of three states $s<t<u$ with $l(s)$={`bird`,`flies`}, $l(t)$={`penguin`,`bird`}, and $l(u)$={`penguin`,`bird`,`flies`}.

`has_wings` from the premiss `bird`, we would consider the agent irrational if it wouldn't accept the conclusion `flies` from the premiss `bird`$\wedge$`has_wings`.

Comparing forms of reasoning

The different rule systems in the KLM framework are related by metalevel entailment: for instance, every rule of **P** can be derived from the rules of **M**, prompting Kraus *et al*. to call **M** stronger than **P**. Since consequence relations can be viewed as Herbrand interpretations of rule systems, an equivalent formulation of this observation is that every monotonic consequence relation is preferential. This can also be seen on the semantic level, by noting that every monotonic structure V establishes a preferential structure $W=\langle V,I_V,\emptyset\rangle$[52], such that the monotonic consequence relation $\mid\!\sim_V$ defined by V coincides with the preferential consequence relation $\mid\!\sim_W$ defined by W.

However, the connection between this formal relation of relative strength between rule systems and the intuitive relation of relative strength between forms of reasoning is not so clear-cut as one might conclude from the KLM framework. Intuition tells us that plausible reasoning is weaker or less restrictive than deductive reasoning, in the sense that plausible reasoning allows, in general, for more conclusions from given premisses than deductive reasoning does. Thus, the weakest possible form of reasoning would draw any conclusion from arbitrary premisses — let us call such a form of reasoning *flunky*. Flunky reasoning is axiomatised by the following rule:

- **Flunk**: $\alpha \mid\!\sim \beta$

The system **F** consists of the single rule of Flunk. Now, it is easy to see that every rule of **M** is a derived rule of **F**, so **F** is stronger than **M** according to the KLM criterion! Another way to see this is by noting that every flunky consequence relation (there is only one, *viz.* $L\times L$) is monotonic (i.e. it is defined by the empty monotonic structure).

I would argue that the criterion for comparing the strength of different forms of reasoning operates by relating semantic structures rather than rule systems. For instance, the claim that preferential reasoning is less restrictive than monotonic reasoning might be substantiated as follows. Let $W=\langle S,l,<\rangle$ be a preferential structure defining a preferential consequence relation $\mid\!\sim_W$, let $V\subseteq U$ be the set of background models labelling some state in S[53], and let $\mid\!\sim_V$ be the monotonic consequence relation defined by V. From Definitions 4.8 and 4.10 we see that $\mid\!\sim_V$ is indeed a subset of $\mid\!\sim_W$, and if $<$ is non-empty the inclusion is proper. That is, for a given formula α we can draw at least the same consequences using $\mid\!\sim_W$ as we can draw using $\mid\!\sim_V$; for instance, by virtue of the preference relation we may be able to conclude `bird` $\mid\!\sim_W$ `flies`, while `bird`$\not\mid\!\sim_V$ `flies`. Furthermore, $\mid\!\sim_V$ represents the most comprehensive monotonic consequence relation included in $\mid\!\sim_W$: any monotonic consequence relation $\mid\!\sim_X$ that is a superset of $\mid\!\sim_V$, and that includes an argument in $\mid\!\sim_W$ but not in $\mid\!\sim_V$ (such as `bird` $\mid\!\sim_W$ `flies`), would also include some argument not in $\mid\!\sim_W$ (such as $\neg$`flies` $\mid\!\sim_X$ $\neg$`bird`). $\mid\!\sim_V$ contains all the arguments from $\mid\!\sim_W$ that can be

[52] I_V: $V\rightarrow V$ denotes the identity function on V.
[53] Strictly speaking, V should be defined as $\{l(s) \mid s \text{ is minimal in } [\alpha] \text{ for some } \alpha\in L\}$, i.e. only including those models in U that are actually used to determine the preferential consequences of some α.

obtained by means of monotonic, i.e. non-preferential, means; it is called the monotonic *restriction* of $\mid\!\sim_W$. It is the existence of such a mapping between semantic structures which prompts us to call one form of reasoning *more restrictive* than another.

§17. SUMMARY AND CONCLUSIONS

In this chapter I have introduced the main formal tool that will be applied in this thesis: the general concept of a consequence relation, intended as a metalevel abstraction of an arbitrary reasoning form. I owe this tool to Kraus, Lehmann & Magidor, who use it to develop a descriptive logical theory of plausible reasoning. I have argued that the concept of a consequence relation establishes in fact a methodology of descriptive logic.

The rule systems developed by Kraus *et al.* will also play a role in the coming chapters, which is the reason why they have been presented here in some detail. For instance, my characterisation of explanatory reasoning will be based on a converse form of the rules of **M**. Furthermore, as has been mentioned in the previous chapter, confirmatory reasoning can be thought of as a form of preferential reasoning. This will be worked out more fully in chapter 7.

The question arises whether the methods for a logical analysis of reasoning outlined in this chapter are universal. Does every form of reasoning allow for a tripartite formalisation in terms of semantics, proof theory and metatheory? As we have seen in the previous chapter, Carnap's answer to this question is negative: his 'inductive logic' only provides a semantics in terms of a function assigning a degree of confirmation to arbitrary arguments, rendering the concept of a proof theory superfluous. On the other hand, the KLM framework does supply a tripartite formalisation of plausible reasoning. In this respect, it is illustrative to compare their approach to the standard formalisation of deductive reasoning: the only difference appears on the semantic level, in that the concept of satisfaction-preservation is replaced by the concept of what might be called *preferential* satisfaction-preservation. I will generalise this in the next chapter by introducing the concept of a *preservation semantics*, whose main virtue it is to define a certain semantic quality that the conclusion of an argument inherits from the premisses.

* * * *