A Non-Classical Modal Logic For Belief

Michael P. McPartlin

Ph.D
University of Edinburgh
1991

Abstract of Thesis

The standard model of knowledge and belief attributes to agents the ability to reason perfectly in classical logic. This is known as the problem of logical omniscience and, in accordance with the requirements of their contexts of use, has led to the development of a number of alternative epistemic logics. Some of these alternatives can, like the standard model, be regarded as presenting for discussion and analysis in a base language a system of reasoning, or consequence relation: the relation under which beliefs are closed. Adopting this perspective with regard to a useful four-valued logic, the resulting extension of the standard model is described and many technical points of comparison with the original model are given.

Table of Contents

Chapter 1

Introduction

The classical approach to modelling epistemic notions by means of Kripke structures may be seen to assume a particular philosophical characterisation of how mental states are primarily to be understood, namely in the rôle they play in explaining rational action. Although the following interpretation is in no way forced upon this approach, it seems a natural one to make. To assume that an agent is rational is in part to attribute to him the ability to recognise with some degree of clarity the various possible outcomes of the alternative actions he is free to perform at any time: these possible states of the world are those that are conceivable or possible for the agent. Belief, for example, may then be taken to be a particular relation between these possible states of the world, relative to the agent in question, which picks out a subset of the set of all possible states and in which what is believed in fact holds true. In reasoning about actions that may be performed, only outcomes where his present beliefs are true are considered possible, and conversely, whatever is true in that set of possible worlds defines precisely the beliefs of the agent. The problem with this, of course, is that there are many things that will be true in those worlds appropriately related to the actual world which we would be reluctant to say that the rational agent must believe. For example, he would acknowledge that there may be sentences now true the truth of which remain unaffected by his actions or by his learning them, but which he does not know. This is known as the problem of logical omniscience, for this means of characterising epistemic states would appear to be forced to attribute to any agent the belief in and knowledge of all logically valid propositions, for these are true at all possible states of the world. Moreover, all classical logical equivalents of a given belief are also believed, and even worse, the agent's set of beliefs is closed under classical logical consequence. It appears that within this paradigm we must accept that this model of epistemic concepts in fact explains some ideal of rationality, perhaps that in terms of approximation to which we describe and understand the representational mental states of agents; or else we modify the model, perhaps with the aim of changing the ideal or, at the limit, perhaps with the aim of explaining the defining or minimal conditions of rational belief, the conditions which must hold for any ascription of rational belief to sentient beings to make sense. Most practitioners in the field can be seen as attempting the latter under some guise, though mainly from a different perspective.

The last point is important to note, given the diversity of the possible uses of epistemic logic, so it is helpful to labour this point in order to avoid seeing rivalries between models where they do not in fact exist. Given a theory, we may wish to discuss new concepts or perhaps we may wish to avoid a consequence of the theory which is undesirable for our present purposes. For example, the context with which we are concerned may dictate that the terms of the original theory behave differently here. We have the choice either of extending or revising the theory in places while retaining for the most part the methodological assumptions it embodies, or of rejecting the old theory in a more radical sense and developing a new theory which accommodates the phenomena in the particular context we are interested in. Just why the consequence is undesirable depends on what we understand the theory to be for and affects the criteria for an effective revision: how we propose to use the theory contributes to determining whether the choice we make of revising the theory in a particular way or of developing a new theory is the correct one. This debate does not take place at the level of models.

Furthermore, to develop just such a new theory is not to deny the validity of the original one and its applicability to phenomena and contexts other than those to which ours is designed to apply: all that is denied is its adequacy to accommodate the new concepts with which our own model is concerned, which is only a rival to the original theory or its modifications to the extent that they may provide a conflicting interpretation of the terms of our theory as they behave in our context - or perhaps if they are proposed as a general unifying theory the terms of which apply across a number of contexts. Otherwise there is no common data with regard to which there could be conflicting interpretations. There are of course more general levels at which they could conflict: we could regard them as theories which model the same type of things through some abstraction from the particular phenomena represented (e.g., they are both epistemic theories) and then regard the two theories as providing different answers to the question of the nature of representation (e.g., of informational content). At this level epistemic theories are always comparable, but this is not usually the level at which the evaluation of a theory takes place. This is not of course to imply that different areas of application should happily co-exist in ignorance of each other: on the contrary, the similarities of the techniques used and the problems faced would tend to suggest that there are benefits to be had in methodological and technical comparisons which may illuminate the assumptions behind the theories and the choice of models. Indeed, developing a theory to avoid an unwanted consequence may reveal commonalities in these areas. We now look at some such contexts for epistemic logic.

The methodological inadequacies of an ad hoc practical solution to an epistemic problem can throw into focus the questioning of an aspect of the original paradigm. Thus we may be interested in modelling some related epistemic concepts, and then decide that this requires revision of the paradigmatic model of epistemic concepts. Because of some such concerns we may in a philosophical mood regard the problem of modelling logical reasoning in some area of research as in part a problem of explaining how reasoning can be said to increase one's knowledge or beliefs given that the information contained in the end state of the reasoning process is in some sense already contained in the premises. An initial attempt to remedy the situation may simply reject the view that logical relations hold between the objects of belief, and so take a belief state to be simply a set of sentences each of which satisfies a belief predicate, perhaps closed under an incomplete set of deduction rules. This by itself is no solution to the problem, for although we can now quantify the beliefs in a number of belief states and so compare their size we have lost the ability to model reasoning, as well as the potential to explain how non-syntactic contextual factors contribute to determining the content of the attitudes. In most applications of epistemic logic these would deem the theory inadequate, throwing into focus the paradigm's model of reasoning and our philosophical interpretation of it.

For example Vardi[1], principally within the field of the study of distributed systems in computer science, proposes a semantic analogue of this approach which takes a belief state to be a set of propositions which satisfies certain closure conditions, specified according to the reasoning abilities to be modelled. The interest of workers in computer science in epistemic logics is natural when multiple agents are accommodated in the formal model, for then it seems natural to attempt to characterise the state of knowledge which results after an agent receives information, in particular given certain assumptions on the nature of the information and the medium of communication. In this context the logic models the information one part of a communication system has as well as the information the whole system has, for example under the assumption

---

[1]Fagin and Vardi 1985, Vardi 1985, Fagin and Vardi 1986 and Vardi 1986. A fuller description of the logic below is to be found in chapter 2.

that communications do not always arrive, or are not always transmitted on time. In Vardi's proposal, however, there are these differences from much of the other work in the field which uses something like Kripke structures: although possible worlds are still fundamental notions, there is an assignment to agents of collections of sets of these which is functional rather than relational, with the consequence that epistemic notions are no longer seen as founded on the philosophical picture outlined above. Instead propositions rather than states of the world are fundamental to this epistemic theory, thus diverging in its philosophical presuppositions from the interpretation of Kripke structures outlined above, although there is no need to regard them as rivals at the level of their primary intended interpretations. The result, however, is a technically flexible and semantically straightforward schema, which he demonstrates may serve as a companion to Kripke structures in modelling various epistemic states. It can model more refined notions of reasoning than the traditional approach, lacking the assumptions of closure under consequence and knowledge of all validities[2].

Even reasoning moves such as adjunction and $\wedge$-elimination are not valid in all epistemic structures. The carrier set of the aforementioned Kripke structure is a set of what he calls knowledge structures, the conditions on the construction of which are dictated by the intended purpose of the model. Very roughly, the construction of a world proceeds by distinguishing levels of construction according to the depth of embedding of operators in formulas that are assigned to agents at that level.

Level 0 is an assignment of classical truth values to atomic formulas; level 1 is an assignment to each agent of a collection of sets of 0-ary worlds, the set of propositions believed by the agents; level 2 assigns the agents sets of sets of 1-ary worlds, what the agents believe they (or others) believe, and so on. Conditions on the construction capture reasoning; e.g. if we require that the collection of

---

[2]Vardi 1986, in particular, discusses modal systems weaker than K.

sets of worlds assigned form a filter in $\mathcal{P}(\mathcal{P}(W))$, W being the set of worlds at the previous level of construction, this determines the modal logic K.

Restrictions on the construction of the modal world straightforwardly allow for the representation of logics weaker than K, but no weaker than E[3]: in other words, logical equivalents cannot be distinguished, but there is no equivalent of closure under consequence or the rule of necessitation, and most other rules of inference can be allowed to fail. The joint knowledge and the common knowledge of a number of agents can also be modelled in this way, which is of great importance in his chosen application of analysing distributed systems[4].

Relations on a Kripke structure which is a set of epistemic worlds can be used to compare states of knowledge, or to circumscribe the state of knowledge of an individual: i.e. to express concepts such as "knowing more than" or "all a knows is α"[5]. Thus knowledge worlds (belief worlds) and Kripke structures complement each other in that knowledge worlds model agents' states of knowledge whereas Kripke structures can be used to model collections of knowledge states: taking the relation for each agent on the structure to be set inclusion, for example, allows new epistemic concepts such as "knowing more than" to be modelled. a knows at least as much in knowledge structure f as he does in knowledge structure g iff for each level of assignment in f and g, what a is assigned in f is set-theoretically contained in what a is assigned in g. So a has in g at least all the possibilities he has in f, and possibly more, and f is a possible state for a in g.

---

[3]The Kripke semantics for systems below K is very different and no longer so straightforward. Though see Blok and Kohler 1983 for some positive results. The names of the logics may be found in Chellas 1980.

[4]As in Fagin and Vardi 1986. There is also however some philosophical interest in this question e.g. Humberstone 1985.

[5]See Vardi 1985.

If the epistemic axioms are reasonably strong then, as Vardi proves, there is the an equivalence between determining the truth of epistemic formulas by looking at the 'internal' semantics of the knowledge world and looking at the knowledge structures that are possible for the agent at that world: a knows that α at g iff α is true at all states f possible for a at g. Other relations may also be definable between knowledge worlds, for example to model knowledge acquisition through communication[6].

The main concerns of this research are thus essentially practical, the choice of the particular model of the attitudes being motivated by the specific technical task at hand, and the motivation for the whole method being its flexibility in this regard; though we have noted that this is not without philosophical consequence for its theory of content. And the divorce effected between the objects of agents' beliefs and how they reason with them - between the descriptive and the normative - serves to highlight for us the question of the relative status of these two aspects of an epistemic logic. His main motivation, however, is brought out by his argument in its favour and against Kripke structures. Conceptually, he argues, a possible world is a primitive notion which means different things in different uses of possible world semantics: in dynamic logic it may be a programme state, whereas in temporal logic it may be a point in time, and then Kripke structures are tailored according to the intuitive understanding of what the possible worlds are supposed to be. Finally axioms are found to describe the structure. In epistemic logics, Vardi maintains, things are different: first the axioms are selected, and then the structures are tailored to fit the axioms. And since in this application it is not clear what a possible world is, he asks how we can construct the appropriate Kripke structure without understanding its basic constituents. In accordance with this analysis, he advocates explicitly building complex structures to correspond to single worlds in Kripke structures,

[6] Fagin and Vardi 1986.

according to the specific application in mind - he describes the 'internal semantics' of each Kripke world. This is possible because in this particular context of use of epistemic logic it makes sense to ascribe belief without ascribing powers of reasoning - other than, as should be noted well, the technically residual ability given by closure under logical equivalence in the classical base logic. So the abandonment of inbuilt normative assumptions - not of the particular norm of classical consequence, but of normativity itself - inevitably means that a relational Kripke-type structure cannot be used to model this form of belief, given that these structures carry such assumptions.

We now briefly mention an example of a perhaps more philosophically motivated alternative to standard techniques of modelling the attitudes in a particular application of epistemic logic in a context which demands a more radical rejection of these. It does not directly address the problem of logical omniscience but it illustrates that more abstract objections to the paradigmatic model are possible. This is Rosenschein's attempt to remain neutral with regard to the nature of representation in artificial intelligence and robotics, in a way in which he views other workers as having singularly failed to do[7]. He thus disputes the philosophical presuppositions of other epistemic theories rather than their 'local' explanatory adequacy with respect to the task at hand. Clearly inspired by work in situation theory he asks what it really means for a machine to know that α, or to satisfy the axioms of our model-theoretic semantics, and rejects the classical AI approach based on interpreted symbolic structures. Classically, he argues, the state of the machine is seen as encoding symbolic data objects for which the designer provides some particular interpretation mapping its parts to parts of the world. These data structures are knowledge representation structures solely in virtue of the assignment of content provided by the intended interpretation function the designer has in mind. Instead, he argues, we need to know how the

---

[7]Rosenschein 1986 and Rosenschein 1987.

actual world corresponds to our abstract model to understand what it means for computers to know and act. And so, rejecting the idea that the ascription of knowledge depends upon the attitude of a designer, Rosenschein's work is based on the assumption that knowledge is an objective property of the way the machine is embedded in the world, and so is characterisable in terms of objective correlations that hold between machine states and world states. The results of his work on 'situated automata' show that the systematic assignment of propositional content to computational states does not require the assumption that these are prestructured as interpreted linguistic entities.

Returning the problem of logical omniscience, another computational area which has given rise to epistemic logics with principled limitations on the permitted inferences is the search for a representation language to sanction semantically the generation of knowledge from knowledge in a knowledge-base query system. Because of the data-base environment where the main aim is generating rather than discussing knowledge, an explicit modal operator is often taken to superfluous on this approach, and so there is no immediate demand for the nesting of epistemic concepts. The rejection of closure under consequence is quite differently motivated here from in some other applications of epistemic logics, and so a successful solution may differ from other solutions that have been proposed since there are different criteria that it must meet. What is required here is a means of avoiding all forms of classical logical omniscience to allow for fast though perhaps limited reasoning, but without simply making arbitrary restrictions on the syntactic deduction rules; the semantics of models doing the latter tend simply to reflect the inference process as it proceeds, and so to include syntactic entities in their structure, rather than dictating what correct inferences should be on the basis of truth and falsity. In this setting there are reasons for avoiding omniscience over and above that of plausibility: if closure under modus ponens is allowed in a knowledge representation utility then the computational demands of a query would in many cases be practically intractable, and closure under classical logical equivalence is far too coarse-grained to be informative in the typically domain-specific applications of the logic. In addition, however there is also a requirement to avoid the inference from any sentence to a tautology, as well as "ex falso quodlibet", the inference from a contradiction to anything. For instance, to allow EFQ would allow the semantics to sanction as correct any response of the knowledge-base to a query if at the time it is storing inconsistent information. The alternative method of detecting and removing contradictions before a response, or perhaps after a data-base update, is much harder computationally than using an error-tolerant logic, which limits the impact of those errors to sentences that are naturally related to the erroneous information.

These considerations strongly suggest the use of a form of relevance logic, and a number of authors have advocated the use of a form of tautological entailment in epistemic logic, motivating this choice from its advantages within the field of knowledge representation in artificial intelligence[8]. The methods by which the above constraints are achieved and which distinguish this approach from the classical models of epistemic logic are as follows. In classical possible worlds semantics there is in effect only one contradictory proposition because there is only one function which takes each world to false, and similarly, there is only one logically valid proposition. So all contradictions are true in the same worlds - namely in none - and all tautologies are true in the same worlds - in all of them, since the worlds are complete. In relevance logic, however, different tautologies are taken to describe different situations and so it is possible for one to be true without another being so; the same of course goes for different contradictions being false at different (unrealisable) situations. This suggests four-valued se-

[8]See in particular Levesque 1984, Lakemeyer 1987, Patel-Schneider 1985 and Lakemeyer 1986, which are examined in chapter 3. Thistlewaite, Meyer and McRobbie 1986 contains references to uses of relevance logic in theorem-proving and data-base management, where it is used to partition large data-bases into deductively relevant parts.

tups, or partial situations[9], where sentences may receive the values $\emptyset$, $\{1\}$, $\{0\}$ or $\{1,0\}$, and so may, in one common presentation of this logic, be viewed not as a function but rather as a relation of sentences, situations and the set $\{1,0\}$[10]. Then given that a truth-functional sentence entails another if the consequent is true whenever the antecedent is, and the antecedent is false whenever the consequent is, 'ex falso quodlibet' does not hold because there are valuations in which the contradictory antecedent receives the value 1 whereas the consequent does not: for example, if α has the value $\{1,0\}$ in a given situation, then $\alpha \wedge \neg\alpha$ will not, as it turns out, entail any sentence with the value $\{0\}$ there. Moreover modus ponens no longer holds once we stipulate that there is no intrinsic relation between the truth value of a sentence and that of its negation, so agents are not logically omniscient. The existence of efficient and decidable algorithms for determining whether tautological entailments hold makes these variations of relevance logic suitable for knowledge representation systems due to its speed and its coherent semantics. Levesque[11] presents a propositional logic of this form with non-embeddable operators for both active or explicit belief, and for tacit or implicit belief, the latter taken to be the classical logical consequences of explicit belief which are determined by any complete and consistent extensions of the

---

[9]Partial possible worlds are not a new phenomenon and have also been motivated in other ways, for example the model sets of Hintikka 1962, and, as examples of more recent examples, Humberstone 1981, and van Benthem 1985 and 1986.

[10]This is not of course the only means of developing the semantics of relevance logic within a possible worlds framework; indeed, something along the lines of Fine 1986 could meet the required conditions with greater conceptual parsimony and without adopting partial situations in this sense. For the use of restrictions on realizability functions in a semantic treatment of relevance logic for data-base query systems see Mitchell and O'Donnell 1986.

[11]Levesque 1984.

partial worlds in question; Lakemeyer[12] extends this to allow the embedding of operators. Patel-Schneider[13] extends Levesque's logic to the first-order case with an intuitionistic reading of the quantifiers, and with a decidable algorithm for determining tautological entailment through a skolemisation result for sentences in prenex normal form. But unlike Levesque's logic, it does not contain any operators, and so avoids the intricacies of scoping that might otherwise have arisen. Lakemeyer, however, has built upon this work on tautological entailment and has developed a first-order logic with a similar decidability result but which addresses the problems of quantifying in with operators for both explicit and implicit belief. Explicit belief is suitably weak, but preserves the distinction between 'knowing that' and 'knowing what' in a reasonably intuitive manner.

The position adopted here is that, whatever the practical or philosophical purpose of an epistemic logic, it is often an important methodological point to take into account the most basic system of reasoning it attributes to agents. We maintain that it is right to put the axioms before the structure of the model, but not in the narrow sense whereby sometimes, given the widespread acceptance of Kripke structures, debate was circumscribed to arguing about axiomatic issues of secondary importance, such as just which forms of introspective reasoning were legitimate. This will not be our concern: rather, that basic underlying system of reasoning should be given proper consideration before choosing how to model belief; and once such a system is chosen, we should be able to accommodate axiom systems within the resulting logic as freely as is possible in classical modal logic.

The methodological point is important if we want to avoid the problem of classical logical omniscience, and so adopting this perspective with respect to

---

[12]Lakemeyer 1987.

[13]Patel-Schneider 1985.

one particular idealisation of limited reasoning, the second half of the thesis examines models appropriate for a range of belief logics based of the four-valued inference system mentioned above. These include Kripke structures as a subsystem. Before that however, we describe a number of other logics avoiding Kripke structures which may be used to avoid problems of logical omniscience, including syntactic approaches which, whatever their philosophical credibility, risk inconsistency; then we examine and reinterpret within our methodological framework some logics which do use such structures and so where reasoning may be seen to play an important rôle.

Chapter 2

Solutions To The Problem Of Logical Omniscience I

In recent years there has been a great deal of interest in the formal study of epistemology, and many treatments have been proposed in the attempt to find a satisfactory treatment within a logical framework of the notions of knowledge and belief. Once the preserve of the philosophical community, interest in these concepts and their formal analysis has spread to other fields such as artificial intelligence and computer science where their interpretation and application have thrown up issues not previously discussed within this field of study. In spite of this diversification many important concerns remain common to researchers across these fields, and an examination of the attempted solutions to technical problems which arise in confronting formally the main conceptual issues in reasoning about knowledge would feature many of the interesting logics in the literature on formal epistemology. This is what we propose to do in this chapter with regard to the problem of logical omniscience. Thus, while the choice of model may, given its interpretation and ambitions, respect such issues as just what are the introspective and reasoning abilities of an individual, or what is the relation between knowledge and action, and perhaps also characterise the de re/de dicto distinctions and the problems of quantifying in and identity statements, these requirements are balanced by certain other conceptual constraints arising out of the mathematical implications of the standard models. A treatment that is both consistent and credible as a model of a human agent has been found to be less straightforward than it might appear, and much of the work in this field might be seen as attempting to meet both these criteria. While the Lōb-Montague results haunt the approach which treats the objects of knowledge as sentences, those who see the attitudes as relations between agents and propositions must avoid or explain the likely result that the agent's knowledge and beliefs are closed under logical consequence. Methods of remedying or alleviating these problems therefore serves as a useful classificatory device in describing attempts to model and reason about the knowledge or beliefs of agents, and for that reason will be used here to present the main types of theories as well as some of the methods of analysis that are to be found in this field of research.

2.1 The Standard Model

It is a commonplace that the study of knowledge and belief could not be carried out within an extensional language such as standard first-order logic. Since Frege it has been known that the intersubstitutability *salve veritate* of co-referring terms within these contexts does not generally hold, so epistemic and doxastic logic cannot be adequately carried out in traditional extensional first-order logic as it stands. The propositional operators of classical logic, being truth-functional, can be understood solely in terms of their truth tables, but the operators 'knows' and 'believes' cannot be understood in this way. A natural assumption at the beginning of the study of models for knowledge and belief was to treat these notions as modal operators on propositions, so permitting a non-truth-functional semantics. At the time of Hintikka's *Knowledge and Belief*[1],

---

[1]Hintikka 1962.

the first extensive treatment of the subject, formal modal logic had progressed from its purely syntactic beginnings with Lewis to the model-theoretic tradition initiated by Carnap. Modal model theory developed from unordered sets of state-descriptions together with a valuation function from pairs of state-descriptions and atomic propositions to truth values, into a more structured model containing in addition a binary relation on the set of possible worlds, as introduced by Kripke[2]. We present here Hintikka's propositional logics of knowledge and belief which adopted this framework and was the first such model in the semantic tradition of modelling these notions.

The idea that Hintikka starts with is that, given a set of propositional variables, a (partial) description of a state of affairs is to consist of a *model set*, or a downwardly saturated set of sentences. A downwardly saturated set Δ of sentences is one which satisfies the following syntactic conditions, where α and β are arbitrary formulae of the language.

$(C1)$ If $\alpha \in \Delta$ then $\neg\alpha \notin \Delta$.

$(C2)$ If $\alpha \wedge \beta \in \Delta$ then $\alpha \in \Delta$ and $\beta \in \Delta$.

$(C3)$ If $\alpha \vee \beta \in \Delta$ then $\alpha \in \Delta$ or $\beta \in \Delta$.

$(C4)$ If $\neg\neg\alpha \in \Delta$ then $\alpha \in \Delta$.

$(C5)$ If $\neg(\alpha \wedge \beta) \in \Delta$ then $\neg\alpha \in \Delta$ or $\neg\beta \in \Delta$.

$(C6)$ If $\neg(\alpha \vee \beta) \in \Delta$ then $\neg\alpha \in \Delta$ and $\neg\beta \in \Delta$.

The connectives $\supset$ and $\equiv$ may be understood in terms of these: $\alpha \supset \beta$ is defined as $\neg\alpha \vee \beta$, and $\alpha \equiv \beta$ is defined as $(\alpha \supset \beta) \wedge (\beta \supset \alpha)$.

---

[2]Bull and Segerberg 1984 contains a history of this development.

To *embed* a set Γ_0 of sentences into a downwardly saturated set Δ of sentences is (possibly) to add a number of formulae to Γ_0 to obtain $\Delta' \subseteq \Delta$, a set of sentences which satisfies the above conditions. Δ' need not be maximally consistent and so need not contain a *literal* of every sentence of the language: literals are atomic formulae and their negations. Without $(C1)$ these conditions together form an effective procedure. The presence of $(C1)$, the consistency condition, ensures that a set of sentences is inconsistent if and only if it cannot be embedded into a model set, so the possibility of such an embedding tells us that there is nothing incompatible with this set of sentences to be found in the model set. This is intuitive, since the consistency of Γ_0 just is the possibility of a state of affairs where all its members are true, and this is the case if and only if there is a consistent description of such a state which contains all members of Γ_0.

The extension to modal logic requires new conditions for K and for B, where K and B are propositional operators meaning "a knows that" or "a believes that", where subscripted by a. These conditions must be in harmony with the idea that the consistency of a set of sentences is the possibility of embedding it in a model set, even where this set contains sentences such as $K_a\alpha$ and $B_a\alpha$. Hintikka's idea here is not to make these new rules conditions on the construction of a model set, which would be to make the notion extensional, but rather to define them by invoking the idea of a *set* Ω of downwardly saturated sets of sentences, together with a binary alternativeness or accessibility relation R_a defined over this set for each agent a. An agent knows α if α is true in all alternative states of affairs he considers possible, or, to adopt Hintikka's rather idealistic terminology, the extent of a's information is to be such that he can restrict his attention to a particular subset of the set of model sets. This structure of model sets is called a model system, and it is within the model system that Hintikka introduces his conditions on epistemic operators. First he defines the operators P_a and C_a meaning 'it is possible, for all a knows, that' and 'it is compatible with everything a believes that' respectively. In what follows we shall give the rules for knowledge and its related concept; exact counterparts for belief may be formulated simply by substituting the operators in the definitions. The conditions that Hintikka gives are as follows:

$(C.P^*)$ If $P_a\alpha \in \Delta$ and if Δ belongs to a model system Ω, then there is in Ω at least one epistemic a-alternative Δ^* to Δ such that $\alpha \in \Delta^*$.

$(C.\neg K)$ If $\neg K_a\alpha \in \Delta$, then $P_a\neg\alpha \in \Delta$.

$(C.\neg P)$ If $\neg P_a\alpha \in \Delta$, then $K_a\neg\alpha \in \Delta$.

To capture knowledge we add to these conditions:

$(C.K^*)$ If $K_a\alpha \in \Delta$ and if Δ^* is an epistemic alternative to Δ (with respect to a) in some model system, then $\alpha \in \Delta^*$.

$(C.refl)$ The alternativeness relations R_a are reflexive[3].

$(C.trans)$ The alternativeness relations are transitive.

If we substitute B for K and C for P in $(C.P^*)$, $(C.K)$ and $(C.P)$ we get $(C.C^*)$, $(C.B)$ and $(C.C)$, the corresponding conditions on belief. Hintikka's notion of belief is captured by adding the following combination of conditions to these three rules:

---

[3]A relation R is *reflexive* if $(x,x) \in R$ for all x; is *euclidean* if $(y,z) \in R$ whenever $(x,y) \in R$ and $(x,z) \in R$; is *transitive* if $(x,z) \in R$ whenever $(x,y) \in R$ and $(y,z) \in R$. Another common move is to make the relation *serial*: for each x in the model system or frame, there is some y in the system such that $(x,y) \in R$. This last condition ensures the soundness of $\neg K$(**false**), where **false** is a propositional constant false everywhere. **true** is the constant defined by $\neg$**false**. It is expressed in Hintikka's semantics by $(C.b^*)$ below.

($C.B^*$) If $B_a\alpha \in \Delta$ and if Δ^* an doxastic alternative to Δ (with respect to a) in some model system, then $\alpha \in \Delta^*$.

($C.trans$) as above.

($C.b^*$) If $B_a\alpha \in \Delta$ in the model system Ω, then there is in Ω at least one doxastic alternative Δ^* to Δ (with respect to a) such that $\alpha \in \Delta^*$.

What is happening here is similar to what happens in many more familiar logics of knowledge and belief based on Kripke structures: to make more precise what is intended we may place conditions on the alternativeness relation. For example we may want to ensure that everything that the agent knows is true. To do this the actual state of affairs must always be considered by the agent as an alternative, and so the alternativeness relation is always reflexive. This is the effect of ($C.refl$). The corresponding constraint on model sets is that if $K_a\alpha \in \Delta$ then $\alpha \in \Delta$. Similarly, to require that agents always know what they know or believe what they believe - to allow them positive introspection - amounts to requiring that the ancestral of the alternativeness relation, the accessibility relation, be transitive, and so through ($C.trans$) if $K_a\alpha \in \Delta$ then $K_aK_a\alpha \in \Delta$. The model system condition required here is one that Hintikka calls ($C.KK^*$): if $K_a\alpha \in \Delta$ and Δ' is an a-alternative to Δ then $K_a\alpha \in \Delta'$. And if we were to award them negative introspection, knowledge of what they don't know, and so allow ignorance to lead in a non-monotonic fashion to knowledge through introspection, this would be achieved by making the relation euclidean,with a condition to the effect that if $\neg K_a\alpha \in \Delta$ then $K_a\neg K_a\alpha \in \Delta$. In the case of the single agent, the cumulative effect of adding these constraints to the arbitrary relation of K, the smallest normal logic, are known respectively as T, $S4$ and $S5$.

As should be clear, the relational semantics of Hintikka's model systems can be given an equivalent formulation within the more familiar Kripke semantics, which has since become one of the most common possible worlds tools for reasoning about knowledge and belief. Both can be regarded as a labelled and directed graph, with possible worlds treated as primitive notions as its nodes, a fact sometimes useful in matters relating to decision procedures; and both are possible worlds theories which, in the present context, model knowledge and belief as a relation between conceivable states of affairs. Worlds are not possible *simpliciter*, but rather are possible only relative to other worlds: the worlds possible relative to a given world are simply a subset of the set of all possible worlds.

2.1.1 The Problem Of Logical Omniscience

There is however an unpleasant consequence of capturing in this way an agent's knowledge at a world relationally as those sentences true in all the worlds he thinks possible. If this set of alternative worlds uniquely determines an agent's knowledge at a world, then we face the family of problems known as the problem of logical omniscience. An agent is logically omniscient if whenever he believes all the formulae in a set Σ and Σ logically implies the formula α, then the agent also believes α. Before looking at the most general case of the problem, we note that one example of how this might happen in some particular epistemic logic is where α is true under no assumptions and so is a valid sentence of the logic. This first case is brought about by the fact that a logically true sentence is one that is true in all logically possible worlds. Therefore all such sentences will be true in every member of every subset of these worlds, and so true in all worlds thought possible by any agent at any world, under the assumption that these worlds thought possible are sentence complete. All agents, then, know all the valid formulae, and clearly this problem also applies to belief if we capture that notion in the same manner. This first case is clearly unintuitive, but also equally damaging is following situation, perhaps the problem which has most concerned the epistemic logician favouring propositional theories: that agents are perfect reasoners in classical logic, unable to know and belief anything without being fully aware of its consequences. This generalisation of the previous result is brought about by the fact that if agent a knows α and if α logically implies β - or if a knows that $\alpha \supset \beta$ - then a must also know β. If a knows α then α is true in all worlds he considers possible, and if $\alpha \supset \beta$ is logically true then, as above, it is true in the same set of worlds. So β is true in all the worlds considered possible by a, and by definition of knowledge, a knows β. The two aspects of the problem mentioned are not co-extensive: in a partial semantics such as Hintikka's epistemic logic an alternative model set need only be downwardly saturated and not maximally consistent or sentence complete, and so the agent need not believe all valid formulae. But within a Hintikka-type model system, where knowledge is defined as truth in all alternatives, an agent's beliefs and knowledge are closed under consequence as the following proof shows. First assume that $K_a\alpha \in \Delta$ and $K_a(\alpha \supset \beta) \in \Delta$: then in all alternatives both α and $\neg\alpha \vee \beta$ are true. If β were *not* true in one of these alternatives then both $\neg\alpha$ and α would be true there, which is impossible. So β is true in all alternatives and so $K_a\beta \in \Delta$. No rules peculiar to the analysis of belief have been used, so by using the analogous conditions for belief a parallel argument would demonstrate that this form of omniscience holds for belief as well, and so agents know and believe all logical consequences of their knowledge and beliefs. There are, besides these results, corollaries to the problem of logical omniscience to the effect that agents cannot distinguish between logically equivalent formulae - they cannot have different epistemic attitudes to two logically equivalent sentences - and if an agent believes both a formula and its negation, then he cannot but have the reasoning ability to put these facts together and so he believes every formula, since a contradiction may entail anything.

These unintuitive consequences of this possible worlds approach to modelling propositional attitudes are calamitous if we intend our models to be to some degree faithful to our intuitions about these concepts. The history of science, for example, reveals that much effort has gone into working out the implications of sets of axioms; there was no shortcut available which involved simply *believing* those axioms. The problems caused by consequential closure reveal that this type of model as it stands cannot deliver an account of the epistemic attitudes capable of giving a plausible account of the knowledge and beliefs of agents which, as in the case of humans, have limited reasoning capabilities. And since to divorce knowledge from reasoning about the world and about one's knowledge itself would be almost completely uninformative, a solution to this problem is necessary to any epistemic logic which takes 'knows' and 'believes' to be propositional operators.

At this point, of course, one could reject the propositional attitude approach and advocate a model where knowledge is instead a relation between agents and sentences. If we take 'knows' to be a first-order predicate of terms that are names of sentences, then it will no longer be a built-in feature of our logic that agents are logically omniscient; there is nothing intrinsic to the naming process which guarantees any reasoning ability. There are nevertheless methods of dealing with the problem which retain the *semantic* approach to knowledge and belief where these are seen as relations between agents and propositions - the terminology does not refer exclusively to models based on Hintikka/Kripke-style possible worlds semantics, or even on possible worlds semantics. We now mention some logics which tackle the problem of logical omniscience within this broad tradition.

2.1.2 Reinterpreting The Operator

Perhaps the simplest idea is to re-interpret the meaning of the operator. Although the model is basically similar, for example, to that which Hintikka presents and so has the property of consequential closure, this situation is made acceptable by the claim that what is intended is not in fact a model of the knowledge or the beliefs of human agents with their cognitive and computational limitations and their restricted faculties of reasoning. Thus one may take oneself to be modelling the ideally rational agent with unlimited reasoning abilities and thus obviate any requirement that the model respect *human* powers of reasoning, and so be credible as a model of human agents. This idealisation lends itself to easy logical analysis at the expense of being unrealistic for human agents. Alternatively we may say, with Moore[4] that inferring that $K_a\beta$ from $K_a\alpha$ and $K_a(\alpha \supset \beta)$ should be treated as a default rule. The fact that something follows from a's knowledge is seen as a justification for concluding that a knows it, although this is a defeasible conclusion on which doubt may later be cast. A third strategy, first proposed by Levesque[5] and later taken up by Halpern and Moses[6] distinguishes between *implicit* and *explicit* belief, reinterpreting the operator as implicit belief. What is implicitly believed by an agent is what the world would be like if what he believed were true, and so closure under implication is appropriate to this concept. What distinguishes these proposals from other approaches which re-interpret the meaning of the operator, however, is that our everyday concept of belief is not ignored, but is incorporated into the model by different means. The technical means to this end is the introduction of inconsistent and incomplete worlds into the semantics. The type of approach of which this is an instance will be examined in greater detail in the next chapter.

We shall only mention briefly Hintikka's own, rather different proposal for impossible worlds, given that it is best understood within the context of his game-theoretical semantics. It borrows from Rantala[7] the idea of a kind of model, an

---

[4]Moore 1985.

[5]Levesque 1984.

[6]Halpern and Moses 1985.

[7]Rantala 1975.

urn model, the domains of which may change during investigation. These are to be the impossible worlds where some logically false sentences may be true. Invariant urn models, or invariant worlds, are those that are logically possible. But some worlds change so invidiously that, given one's logical abilities, they may appear invariant, and so be epistemically possible. This is just as Hintikka's argument for impossible worlds requires. It is unclear, however, whether the intuition behind the nature of these impossible worlds is quite so coherent or compelling outside game-theoretical semantics, for the dynamic way in which truth definitions are given within that theory is precisely what allows for the possibility of defining changing models. Outside the theory that intuitive step is not available: there is no generally plausible motivation for impossible worlds[8], and the interpretation of possible worlds semantics which underlies Hintikka's argument for allowing impossible worlds does not have general currency.

2.1.3 Non-Relational Modal Structures

There is a line of thought amongst some propositional attitude theorists which discerns the cause of the problem of logical omniscience to lie in the way in which epistemic notions are captured within Hintikka's or Kripke's relational possible worlds semantics, and so requires the rejection of this treatment as an important ingredient of a solution to the problem. A case in point is the view of Cresswell[9] whereby propositional attitudes do not satisfy modal conditions, and so take logically different objects from operators such as "it is necessary that". Thomason, on the other hand, puts forward an interesting proposal which treats the proposition as a primitive notion - rather than as a set of possible worlds - in

---

[8]Although there have of course been other definitions of impossible worlds; see, for example Rescher and Brandon 1979.

[9]Cresswell 1985.

fact as a basic type in a modification of Montague's intensional logic[10] but where logically equivalent sentences need no longer express the same proposition. There are, however, less radical alterations to the traditional theories of knowledge and belief which remain within the framework of modal logic. We now describe an alternative modification of a theory of Montague which retains intensions as the objects of propositional attitudes by capturing Hintikka's possible worlds semantics in a *set-theoretical* rather than a graph-theoretical way. Since it uses a possible worlds theory agents inevitably believe all those sentences with the same intension as α, if they believe α, but since the modal structure used is not that of relational semantics we can avoid all other cases of the closure of the agent's beliefs under logical consequence, while in addition allowing a great deal of sensitivity in modelling reasoning.

Models for epistemic logic[11] are to be based on neighbourhood or *Scott-Montague* semantics for modal logic, as first outlined by Montague[12] for the purpose of interpreting what he called *pragmatic* languages. With each world $w \in W$ is associated a set of propositions for each agent, which are to be the propositions believed by the agent at that world. Where P is the finite set of atomic formulae, we want to partition the language based on it into equivalence classes defined by semantic equivalence such that each sentence in a particular equivalence class will have the same intension. With this in mind the function $\Pi : P \rightarrow 2^W$ assigns to each atomic formula its intension, or the intension of its equivalence class, which is a subset of W. An agent's beliefs at a world

---

[10]See e.g. Montague 1970, Montague 1973 or Gallin 1975 for an exposition of Intensional Logic.

[11]See e.g. Fagin and Vardi 1985, Vardi 1985, Vardi 1986.

[12]Montague 1968. See also Chellas 1980, chapter 7, where they are known as minimal models.

are simply a collection of intensions, where intensions are subsets of W, and so a function $N: A \times W \rightarrow 2^{2^W}$ introduced for this purpose can be seen as assigning to each agent $a \in A$, the set of agents, a set of intensions at each world. Thus a Scott-Montague model is to be a triple $< W, N, \Pi >$ where W is a non-empty set and N and Π are functions as defined above. The atomic case for the definition of satisfaction is $M, w \models p_i$ if $w \in \Pi(p_i)$, where p_i is atomic, and the induction steps for the non-atomic formulae are the usual ones for the classical propositional connectives. Taking B_a to be the epistemic operator "a believes that", the modal case is defined by

$$M, w \models B_a\alpha \text{ if } \{u \mid M, u \models \alpha\} \in N(a, w).$$

That is, $B_a\alpha$ is satisfied in w if the intension of α is a *neighbourhood of w for a*. So at each world agents are assigned a collection of intensions, the *belief set* of the agent at that world. Since it is not a set of worlds which the agent thinks possible that determines what he believes - as is the case in relational models of epistemic notions - one form of logical omniscience is avoided: modes of reasoning are not built into the underlying frame until axioms for these are chosen to constrain in some way the construction of the frame. Agents still believe semantically equivalent formulae but they need not believe all valid formulae because no assumptions are made about the nature of belief. If $B_a\alpha$ and α semantically implies β, then it need not be the case that $B_a\beta$. For if $\{u \mid M, u \models \alpha\} \in N(a, w)$, for the latter to be the case at some world w, $\{u \mid M, u \models \beta\}$ would have to belong to $N(a, w)$. But the members of $N(a, w)$ are a collection of intensions for which no means has been described of how to produce new intensions out of the sets of worlds that make up the members of this collection, and we need such a procedure if we are to add $\{u \mid M, u \models \beta\}$ to $N(a, w)$. Of course if $\{u \mid M, u \models \alpha\} = \{u \mid M, u \models \beta\}$ then $\{u \mid M, u \models \beta\}$ is already in $N(a, w)$, for this is the case where α and β are semantically equivalent. But in the more general case, it might be that for all u such that $M, u \models \alpha$ it happened that $M, u \models \beta$. But β may also be true at worlds outside $\{u \mid M, u \models \alpha\}$ and in this way fail to be in $N(a, w)$. We have, then, the basis of a possible worlds theory for modelling epistemic notions founded - as it must be - on the notion of there being many possible states of affairs, but which, unlike Hintikka's theory, does not model belief as a relation between these states of affairs. Rejecting that model of epistemic notions leads naturally to a rejection of the philosophical interpretation Hintikka grafts onto it - although this interpretation is in no way obligatory for users of Kripke models - as well as permitting ipso facto an alternative role for possible states of affairs within epistemic logic, one that is informative with respect to the task at hand. In other words, we are free to choose the structure of these basic elements in the model, tailoring them, as we wish, to the requirements of the particular epistemic notion being modelled. According to this view, in modelling belief states we want our worlds to be recognizably *belief* worlds. The idea, then, is to modify the basic Scott-Montague semantics through outlining a construction process for W, rather than treating its elements as primitive notions. This should also alleviate the one problem with logical omniscience that remains by weakening the relation of semantic equivalence, since this is determined by the choice of W. In fact, it will be as weak as logical equivalence.

We now outline the semantics for epistemic logic presented by Vardi and others which is based on these insights. Assuming fixed sets P of atomic formulae and A of agents, they first proceed to define belief worlds, before relating these to the full belief structure, which is to be based on the Scott-Montague semantics extended to multiple modalities. Worlds are defined constructively to have not only truth assignments to the atomic formulae but also functions associating with each agent sets of propositions for each depth of embedding at which they may have beliefs. If John believes that Mary believes α, this will be reported at worlds with depth no less than two. Informally, any instance of the first level of construction is called a 1-ary world and tells us about 'nature'; it does not tell us about knowledge or belief. These worlds have the form $< f_0 >$ having a single function f_0 from the atomic formulae of the language to the truth values.

If we suppose that the language has only one atomic formula p_i then there are two possible 1-ary worlds since there are two possible assignments $f_0(p_i) = 1$ and $f_0'(p_i) = 0$ to the two truth values. The two worlds would then be $< f_0 >$ and $< f_0' >$ or more perspicaciously, $< p_i >$ and $< \neg p_i >$. The next level tells us about each agent's knowledge or beliefs about nature: a function f_1 assigns to each agent a collection of sets of 1-ary worlds. Since in the example we only have two such worlds, $< p_i >$ and $< \neg p_i >$, agents may be assigned the propositions $\{p_i\}$, $\{\neg p_i\}$, $\{p_i, \neg p_i\}$ and $\emptyset$. In the case of knowledge, for example, this would mean that the agent knows that p_i, knows that not p_i, does not know whether p_i, or 'knows' some contradictory formula. Combining the particular functions f_0 and f_1 give the 2-ary world $< f_0, f_1 >$: for example, such a world might be $< p_i, (\text{Bob} \rightarrow \{p_i, \neg p_i\}, \text{Chris} \rightarrow \{p_i\}) >$, where p_i is true, Chris believes it and Bob does not believe or disbelieve it. At the next level we are told about the agents' beliefs about their own and other agents' beliefs, this time assigning to each agent a collection of sets of 2-ary worlds of the form $< f_0, f_1 >$; this process may be continued to further levels. More formally, an 0-th order assignment is defined to be a truth assignment $f_0 : P \rightarrow \{1, 0\}$ to the atomic formulae. $< f_0 >$ is a unary world, since its length is 1, consisting of just the one function. For the purpose of induction assume that k-ary worlds of the form $< f_0, \ldots f_{k-1} >$ have been defined, and let W_k be the set of all k-ary worlds. Then a k-th order assignment is a function $f_k : A \rightarrow 2^{2^{W_k}}$, associating with each agent a set of propositions, where each of these is a set of k-ary worlds. $< f_0, \ldots f_k >$ is a $(k+1)$-ary world. Carrying out this process of construction to the limit gives us infinitary worlds: $f_\omega =< f_0, f_1, f_2, \ldots >$ is an infinitary world if each of its prefixes $< f_0, \ldots f_{k-1} >$ is a k-ary world. W_ω is the set of all infinitary worlds, and W is the set of all worlds.

W gives the set of all possible belief worlds, and thus is what is required as part of the Scott-Montague semantics. We can, however, restrict our attention to a subclass of W, but before seeing how, there are a number of restrictions on the construction process to be outlined. One such restriction is obligatory, since it ensures the compatibility of a function extending a world with the functions preceding it. That is, an agent's higher-order beliefs should extend his lower order beliefs with the result that, if we were to remove the last level of construction from the set of propositions at any given level, we should be left with the propositions of the previous level (It should be noted that this is very different from the converse: to ensure that the set of $(k-1)$-level propositions assigned to any agent were prefixes of any k-level propositions assigned to that agent would make the logic *normal*, which we do not necessarily want). The restriction that Vardi imposes is as follows:

$$f_{k-1}(a) = \{\mathbf{chop}(X) \mid X \in f_k(a)\}, \text{ for each } a \in A,$$

where for $X \subseteq W_k$, $\mathbf{chop}(X)$ is the set

$$\{< f_0, \ldots f_{k-2} > | < f_0, \ldots f_{k-2}, f_{k-1} > \in X\}.$$

Agents clearly have little reasoning power as things stand, which leaves us free to choose how much they should have. The following are examples of the constraints that may be put on the construction of belief worlds, some of which mirror the constraints put on the accessibility relation in Kripke semantics, which allow the worlds to be tailored to various modes of reasoning.

1. $B_a\alpha \supset B_aB_a\alpha$. For all $k \geq 1$, if $X \in f_k(a)$ then $\{< g_0, \ldots g_k > | \ X \in g_k(a)\} \in f_{k+1}(a)$.

2. $\neg B_a\alpha \supset B_a\neg B_a\alpha$. For all $k \geq 1$, if $X \notin f_k(a)$ then $\{< g_0, \ldots g_k > | \ X \notin g_k(a)\} \in f_{k+1}(a)$.

3. $B_a(\alpha \wedge \beta) \supset B_a\alpha$. For all $k \geq 1$, if $X \subseteq Y \subseteq W_k$ and $X \in f_k(a)$, then $Y \in f_k(a)$.

4. $B_a\alpha \wedge B_a\beta \supset B_a(\alpha \wedge \beta)$. For all $k \geq 1$, whenever $X, Y \in f_k(a)$ then also $X \cap Y \in f_k(a)$.

5. B_a**true**. For all $k \geq 1$, $W_{k-1} \in f_k(a)$.

6. $\neg B_a$**false**. For all $k \geq 1$, $\emptyset \notin f_k(a)$.

7. Knowledge. $< f_0, \ldots f_{k-1} > \in X$, for all $X \in f_k(a)$.

If 3 and 4 hold the logic is called a quasi-filter. If, further, 5 holds, the associated model contains the unit and so is a filter; the class of filters determine the smallest normal system K. It is clearly useful that K is not generally valid in a class of these models. A sound and complete axiomatisation is achieved through all substitution instances of propositional tautologies and the inference that $B_a\alpha \equiv B_a\beta$ from $\alpha \equiv \beta$.

Finitary belief worlds satisfy formulae as follows:

1. $< f_0, \ldots f_k > \models p_i$, where p_i is atomic, if $f_0(p_i) = 1$.

2. $< f_0, \ldots f_k > \models \neg\alpha$ iff $< f_0, \ldots f_k > \not\models \alpha$

3. $< f_0, \ldots f_k > \models \alpha \wedge \beta$ iff $< f_0, \ldots f_k > \models \alpha$ and $< f_0, \ldots f_k > \models \beta$.

4. $< f_0, \ldots f_k > \models B_a\alpha$ if $k \geq 1$ and $\{w \mid w \in W_k \text{ and } w \models \alpha\} \in f_k(a)$.

Because of the restriction on the construction of belief worlds which ensures that an agent's higher order beliefs are an extension of his lower order beliefs, determining satisfaction requires inspection only of a long enough prefix of the world. With the following definition:

1. $\mathbf{depth}(\alpha) = 0$ if α contains no B-operator.

2. $\mathbf{depth}(\neg\alpha) = \mathbf{depth}(\alpha)$.

3. $\mathbf{depth}(\alpha \wedge \beta) = \max\{\mathbf{depth}(\alpha), \mathbf{depth}(\beta)\}$.

4. $\mathbf{depth}(B_a(\alpha)) = 1 + \mathbf{depth}(\alpha)$.

Then where $\mathbf{depth}(\alpha) = k$ and $r \geq k$, $< f_0, \ldots f_r > \models \alpha$ iff $< f_0, \ldots f_k > \models \alpha$. Now we are almost ready to choose the set W of worlds for the Scott-Montague semantics. First, satisfaction for infinitary worlds is defined: where $f_\omega =< f_0, f_1, \ldots >$, $f_\omega \models \alpha$ if $< f_0, \ldots f_k > \models \alpha$, where $\mathbf{depth}(\alpha) = k$. Then it is shown that worlds can be extended conservatively. Worlds are said to be *equivalent* if they satisfy exactly the same formulae; and for every k-ary world $< f_0, \ldots f_{k-1} >$ there is an equivalent $k+1$-ary world $< f_0, \ldots f_{k-1}, f_k >$ which satisfies the same formulae. So it suffices in the model to consider W_ω, the set of infinitary worlds. To this effect define $\mathbf{prefix}_k(X)$ as $\{< f_0, \ldots f_{k-1} > | < f_0, \ldots f_{k-1}, f_k, \ldots > \in X\}$; define $N(a, w)$, for all $a \in A$, as

$$\{X \mid X \subseteq W_\omega \text{ and } \mathbf{prefix}_k(X) \in f_k(a), \text{ for all } k \geq 0\}.$$

Then where the intension of p_i is $\Pi(p_i) = \{w \mid w \models p_i\}$, the appropriate belief structure is the Scott-Montague model $M =< W_\omega, N, \Pi >$.

By working through the cases in the definition of satisfaction for the semantics it is easy to see that $M, f \models \alpha$ iff $f \models \alpha$ for any formula α and world f, and so the choice of W has the effect of making semantic equivalence identical to logical equivalence. Although agents cannot distinguish between semantically equivalent sentences, this is about the best we can achieve within a classical possible worlds framework. Vardi argues for the greater expressive power of this semantics over Kripke semantics[13], especially with regard to its modelling of both reasoning and context. Further concepts are explored by working out the exact correspondence he proves between the two: belief structures model particular possible worlds, while Kripke structures model collections of these possible worlds. Substitutivity of equivalents within epistemic contexts, however remains a problem here, but this seems inevitable for any propositional attitude

---

[13]As well as previous references, see also Fagin and Vardi 1986 for an example of its use in modelling information and communication in distributed systems.

theory which models epistemic notion by sets of propositions, where these are sets of possible worlds.

2.2 Syntactical Theories Of The Attitudes

The other main tradition within the field of epistemic logic may be characterised by its treatment of the attitudes as predicates of sentences, a suggestion of which Quine was one of the most committed of the original advocates since it accommodates the modalities within a version of first-order logic. Such systems remain first-order in their quantificational form because it is not really sentences but rather *names* of sentences which are assigned to the extension of the epistemic predicate; and for this reason they are also known as *quotational* theories. If such a programme were feasible then its advantages would be obvious, for non-extensional contexts would no longer be a problem and within first-order logic with identity a solution to many of the philosophical problems concerning belief statements would be at hand. However the Lõb-Montague results, later extended by Thomason, threaten many of these approaches with inconsistency, and so oblige their advocates to first diagnose then circumvent the cause of the problems. The results, which are in effect an extension of Tarski's theorem on the non-definability of a truth predicate, proceed by constructing an equivalent of the paradoxically self-referential Liar sentence within a syntactical treatment of modality. The paradoxical result Montague[14] following an earlier result by Lõb[15], arrives at is the following. First let T be a first order theory which is an

---

[14]Montague 1963.

[15]Lõb 1955.

extension of $Q^{(\delta)}$, Robinson's arithmetic[16] relativised to the formula δ whose only free variable is u[17]. Assuming some standard gōdelisation scheme, Let $< \alpha >$ stand for the numeral in $Q^{(\delta)}$ denoting the Gödel number of the sentence α. Then suppose for some one place predicate of expressions K, and for all formulae α and β of T, the following conditions hold:

$(K1)$ $\vdash_T K(< \alpha >) \supset \alpha$.

$(K2)$ $\vdash_T (K(< \alpha >) \wedge K(< \alpha \supset \beta >)) \supset K(< \beta >)$.

$(K3)$ $\vdash_T K(< \alpha >)$, if α is a theorem of logic.

$(K4)$ $\vdash_T K(< K(< \alpha >) \supset \alpha >)$.

Then T is inconsistent.

The proof involves the construction of a self-referential sentence of the form $\sigma = K(< \chi \supset \neg\sigma >)$, where χ is a single axiom for $Q^{(\delta)}$. Details may be found in the original paper. Montague concludes that:

> "if necessity is to be treated syntactically, that is, as a predicate of sentences... then virtually all of modal logic... must be sacrificed."

---

[16]See, for example, Tarski, Mostowski and Robinson 1953, or Boolos and Jeffrey 1980, chapter 14, for the theory Q.

[17]$\alpha^{(\delta)}$, the relativisation of α to δ, is defined recursively as follows: $\alpha^{(\delta)}$ is α; $(\neg\alpha)^{(\delta)}$ is $\neg\alpha^{(\delta)}$; $(\alpha \supset \beta)^{(\delta)}$ is $\alpha^{(\delta)} \supset \beta^{(\delta)}$, and similarly for the other sentential connectives; $(\forall\alpha)^{(\delta)}$ is $\forall u(\delta(u) \supset \alpha^{(\delta)})$, and $(\exists\alpha)^{(\delta)}$ is $\exists u(\delta(u) \wedge \alpha^{(\delta)})$. $T^{(\delta)}$, the relativisation of the theory T to δ, is that theory whose constants are the union of those of T and those occurring in δ, and whose valid sentences are the logical consequences within this language of the set of sentences $\alpha^{(\delta)}$, where α is a valid sentence of T. The crucial properties of Q are that it is finitely axiomatisable and permits the representation of all one-place recursive functions of natural numbers.

Since inspection shows that epistemic logics of idealised knowledge share the same axiomatic structure as those of necessity, it would seem that these also yield a contradiction when knowledge is represented syntactically in this way within a language rich enough to permit gödelisation. And undoubtedly, any general approach to epistemic logic should be able to handle idealised knowledge.

Thomason[18] shows that the result can be extended to languages without the first axiom, and so it is not the close tie between knowledge and truth which is the cause of the paradoxical result. Since condition $(K1)$ (at least) distinguishes between knowledge and belief, if it can be shown that logics of idealised belief lead to similar paradoxes then the contention that knowledge is somehow not a "purely psychological" concept since it involves the notion of truth could not then be used to sidestep the application of the results to certain psychologically motivated semantic theories[19]. Let T be a theory as above, but suppose instead that the following conditions hold for all formulae α and β and for some one-place predicate of expressions B:

$(B1)$ $\vdash_T B(<\alpha>) \supset B(<B(<\alpha>)>)$.

$(B2)$ $\vdash_T (B(<\alpha>) \wedge B(<\alpha \supset \beta>)) \supset B(<\beta>)$.

$(B3)$ $\vdash_T B(<\alpha>)$, if α is a theorem of logic.

$(B4)$ $\vdash_T B(<B(<\alpha>) \supset \alpha>)$.

Then by a proof similar to Montague's, for all formulae α, $\vdash_T B(<\chi>) \supset B(<\alpha>)$, where χ is as above, through the construction of a paradoxical

[18]Thomason 1980.

[19]Thomason regards the theories of those such as Fodor 1978, Jackendoff 1976 and Katz 1977 as being rather programmatic examples of this type of theory.

sentence of the form $\sigma = (\chi \supset B(<\neg\sigma>))$. So although T is not inconsistent, the belief of an apparently harmless tautology as well as $\neg B(<\alpha>)$ for any α would render T inconsistent. It would appear that idealised doxastic logic cannot coherently represent belief as a predicate of sentences.

2.2.1 The Consequences For Theory

Responses to the paradoxes from those theorists who wish to maintain the syntactical treatment of epistemic logic must all somehow obstruct the possibility of constructing a fixed point sentence for any formula in the language. It would however be useful for this approach first to establish just which feature of syntactic theories is responsible for the paradoxes, for then not only would a solution be simpler to find but we would also be able to determine more precisely which theories are at risk from these results if some arithmetic machinery is added to them. Unfortunately the identification of the culpable feature is a matter of some controversy. Thomason regards the Löb-Montague results as putting strict limitations on the cognitively orientated semantic theories that are popular in cognitive psychology and linguistics[20], since these theories make certain assumptions which he regards as parallel to those at the root cause of the paradoxical results. The semantic representation of these theories comprise a recursive set of structures, allowing complex representations to be composed in a combinatorial fashion via certain principles from the basic constituents of representation. Since the ideal user of language is able to store the interpretation of his language, this must be explained by the existence of a recursive relation between a sentence and its semantic representation. If we suppose that the language contains the necessary arithmetical syntax and axioms to contain Robinson's arithmetic, and conditions 1 – 4 are met, then belief cannot be a predicate of the semantics

[20] See previous footnote.

representations of sentences because this predicate can be converted into a predicate of sentences and the paradox ensues. So Thomason regards the paradox as threatening not only those syntactic theories where the epistemic predicate is satisfied by sentences of the same language - which is the way in which Montague first presented it - but rather to have the more general implication of threatening with inconsistency any theory which takes epistemic notions to be predicates on representations structured in this way, whether or not the theory identifies the objects of the attitudes with sentences of its own language.

2.2.2 An Alternative Diagnosis

The claim that the recursive character of the representations that are the objects of the attitudes is the reason why these theories are vulnerable to the paradox is strongly disputed by Asher and Kamp[21], who argue that the problem is more widespread. Although there are some situations where it is an essential element in the derivation of the paradox, they argue, there are many cases where the paradox goes through without this assumption. All that is required to force it through is the ability to define a predicate - related to the epistemic predicate of propositions - on Gödel numbers of sentences, which satisfies the four epistemic principles of Montague or Thomason. The aspects of a theory that are relevant to this ability are whether it can represent the relation which holds between a proposition and a sentence expressing that proposition; the nature of the structure of its propositions, and whether the theory has the machinery to talk about this structure; and the form in which the epistemic principles are expressed. To illustrate, they suppose that the axioms of Q are added to the valid sentences of Montague's system of intensional logic, and demonstrate that the presence in the language of the sentence forming operator $^\vee$ leads to paradox:

---

[21] Asher and Kamp 1986.

"Let H be some particular gödelization relation - i.e. n stands in the relation H to sentence ψ if n is the Gödel number according to some chosen gödelization scheme of ψ. This relation determines a second relation G between numbers and propositions which holds between n and p if n is the Gödel number of a sentence which expresses p. Semantically this relation is completely defined; i.e. its extension is fully determined in each of the models of this extended system of IL. It might therefore seem harmless to add to the given system a binary predicate to represent this relation; and to adopt as new axioms such intuitively valid sentences as a) $G(\mathbf{n}, \check{\psi})$, where $\mathbf{n}$ is the n-th numeral and n is the Gödel number of ψ. b) $(\forall u)(\mathrm{Sen}(u) \supset (\exists! p)G(u,p))$, where 'Sen' is the arithmetical predicate which is satisfied by just those numbers which are Gödel numbers of sentences, and c) $(\forall p)(G(\mathbf{n},p) \supset (\check{p} \equiv \psi))$, where $\mathbf{n}$ and ψ are as under a). However this addition renders the system inconsistent: for we can now define a 'truth' predicate T of Gödel numbers $T(u) = (\exists p)(G(u,p)$ and $p)$ for which we can easily show that $T(\mathbf{n})$ is valid whenever n is the Gödel number of ψ. The inconsistency then follows in the usual way." pp133-134.

So the impossibility of representing within IL - Montague's intensional logic - a semantically well-defined relation is independent of any assumptions about an attitudinal predicate, but the addition of G may be blocked in weaker systems by the presence of such a predicate governed by the usual epistemic principles: if we could define an attitudinal predicate B' of numbers as $B'(n) = (\exists p)(G(n,p)$ and $B(p))$ and show that the principles governing B also hold for B', then the contradiction would go through. In these cases it is not the recursiveness assumption that is critical, but rather the ability of the theory to relate the syntactic structure of its own sentences to whatever it assumes to be the objects of the attitudes. It appears that a framework for avoiding the paradox may be open to the advocates of syntactical theories, which takes the form of excluding the machinery which permits a relation of this type, perhaps via a strategy parallelling some solution to the Liar paradox, its extensional cousin.

2.2.3 Restricting The Syntax

A practical approach to avoiding the paradox is to prevent the construction of the fixed point in the derivation of the contradiction by restricting the extent to which the language can be used to talk about its own syntax, and so avoiding self-reference. One such proposal, from des Rivières and Levesque[22], is of some interest since it creates the foundations of a syntactic treatment of modality by identifying and then placing restrictions on another mainstay of Montague's proof. Then they are able to define a consistency-preserving translation function from modal languages to 'classical' first-order languages with the modal operator reinterpreted as a syntactic predicate. Part of such a process must involve the translation of modal schemata such as $\Box\alpha \supset \alpha$ into $L(<\alpha>) \supset \alpha$, the corresponding schema of what they call the *classical* language, and the strategy is to restrict the syntactic domain of the target language over which the schematic sentence letter should range. Sentences which are of the form $\exists x L(x)$, for example, have no equivalent in the modal language and so should not be seen as belonging to the set of sentences in the classical language over which the schematic letters may range in the translation process. The range of such a translation function on sentences of the modal language would be only a proper subset of the classical language, so any correct translation of a modal schema must be one whose schematic letters range over just this subset. Sentences belonging to this subset are called *regular*, and this restriction to regular sentences

[22] des Rivières and Levesque 1986.

blocks the application of the Lōb-Montague results to the logics based on the associated translation.

Sketching their results briefly, they define an *embedding* as a translation function which maps atomic formulae to themselves, distributes over the connectives and quantifiers of first-order logic and leaves unchanged the free variables of the source formula. A particular embedding of the formulae of the modal language $L(\Box)$ in the formulae of the classical language $L(C)$ is then defined in such as way that the mapping preserves derivability: first, $\bullet$ is defined to be a $1-1$ embedding of $L(\Box)$ in $L(C)$ which has the property $(\Box\alpha)^{\bullet} = L_n(<\alpha^{\bullet}>, x_1, x_2 \ldots x_n)$[23] where $(x_1, \ldots x_n)$ are the ordered free variables of the formula α, and the set of regular formulae is defined to be $L(\Box)^{\bullet}$. Then it is shown that for all formulae $\alpha \in L(\Box)$, and sets of formulae $\Gamma \subseteq L(\Box)$, letting $\Gamma^{\bullet} = \{\beta^{\bullet} \mid \beta \in \Gamma\}$, if the translation function $\bullet : L(\Box) \rightarrow L(C)$ is such that for every $\Gamma \subseteq L(\Box)$, $\Gamma \vdash \alpha$ iff $\Gamma^{\bullet} \vdash \alpha^{\bullet}$ then for all $\Gamma \subseteq L(\Box)$, Γ is consistent iff $\Gamma^{\bullet}$ is consistent. They further prove that satisfiability and theoremhood can be preserved by a particular embedding of $L(\Box)$ In $L(C)$, and so any consistent set of sentences in the modal language can be reduced in this way to a consistent set of sentences of the classical language, where a syntactical predicate takes the place of the modal operator. An attempt to push through the Lōb-Montague results on this language now fails: if we augment it, as before, with Robinson's arithmetic and the four schemata of Montague's or Thomason's proof, holding only for *regular* sentences, a counterexample to the theorem can be found. The construction of a fixed point sentence such as Thomason's $\sigma = (\chi \supset B(<\neg\sigma>))$ is not possible

---

[23]The subscript n varies according to the number of free variables in the formula being predicated. Quine 1979 shows how such a family of predicate symbols may be reduced to a single 2-place symbol L_ϵ taking as its second argument a finite sequence of variable-value pairings such that this becomes $(\Box\alpha)^{\bullet} = L_\epsilon(<\alpha^{\bullet}>, ((<x_1>, x_1), (<x_2>, x_2), \ldots (<x_n>, x_n)))$. Such matters help in constructing a finite axiomatisation.

because it is not regular, and the restriction to regular sentences fails to provide axioms that apply to σ. This strategy of restricting the range of formulae the names of which are allowed to satisfy the syntactical modal predicate is thus a good illustration of the Lōb-Montague results, delimiting more clearly the properties of theories which are dangerously self-referential. Inconsistency does seem to have been avoided in a first order formalisation of modality, and a tractable alternative to intensional logic may appear closer. But the reason why only a subset of the formulae of the language can be fully described as such - although all the formulae have encoding terms, for many of these this cannot be made explicit - has not really been properly motivated. The fact that this saves the theory from contradiction may justify the strategy from a practical perspective, but it does not explain *why* a contradiction would arise were we to talk about any of the sentences other than the regular ones. A plausible motivation is still required, which would involve explaining this curious and unintuitive fact, for why this particular subdivision of the language avoids a contradiction, but it is difficult to see why things should be so.

2.2.4 Truth Value Gaps

Given the standard treatment of truth as a predicate of sentences and the effort that has gone into developing theories to deal with the Liar paradox[24] it would not be surprising should similar treatments of syntactical theories of the attitudes provide a means of blocking the Lōb-Montague results. A number of new theories of truth have been developed in the past few years through dissatisfaction with the unintuitive restrictiveness of Tarski's 'hierarchy of languages' approach, and these suggest that parallel, though more complicated, treatments of the attitudes may be promising. Thus one such method might involve simply carrying over

[24] A useful collection of essays dealing with this subject is to be found in Martin 1984.

the model of truth and using it along with some notion of belief in the definition of knowledge. Tarski envisaged the separation of paradoxical sentences from non-problematic ones to lie in making truth language-external through a stage-by-stage construction of a hierarchy of languages, none of which could contain the means to predicate truth of its own sentences. L_0 is defined to be the usual language of first order predicate calculus, but with the means of discussing its own syntax and without its own truth predicate. A metalanguage L_1 for L_0 is the next stage, and in L_1 we define the predicate $true_1$, applying to sentences in L_0. The process of construction is iterated to give the sequence $\{L_0, L_1, L_2, \ldots\}$ of languages each with a truth predicate for the preceding language. Thus sentences such as 'α is $true_n$' may belong to the extension of $true_{n+1}$, but must fall outside the extension of $true_n$, assuming that α contains no predicate $true_m$, for any $m \geq n$.

Kripke[25], amongst others, has argued against an implicit subscript in our ordinary usage of the concept and against the possibility of a unique assignment of levels to certain general statements involving truth. Moreover, assignment of levels on a sentence-by-sentence basis is not possible in cases where the sentences refer to other sentences: Kripke's example is the situation where Dean asserts "All of Nixon's utterances about Watergate are false" and Nixon says "Everything Dean says about Watergate is false". It is impossible consistently to assign intrinsic levels simultaneously to these two sentences on Tarski's approach. These and other reasons led Kripke to develop an alternative theory of truth which retains something of the stage-by-stage process but with a single concept of truth, and which gives up the distinguishing classical principle of logical bivalence. Instead of having a series of different truth predicates, each defined at different stages and totally defined over the whole domain, there is to be one truth predicate which is given a progressively richer interpretation as

[25]Kripke 1975.

we proceed to higher stages until the process saturates, and then we can distinguish between paradoxical and what he calls *grounded* sentences. To allow the truth values of sentences to be undecided at stages in the evaluation process, a semantical scheme that permits partially defined predicates is required, and Kripke opts for Kleene's strong three-valued logic[26]. Given a non-empty domain D and an interpretation function I, $\neg\alpha$ is true (false) if α is false (true), and undefined if α is undefined. The truth of a disjunction requires the truth of at least one disjunct; it is false if both of its disjuncts are false, and is otherwise undefined. The other truth functions may be defined in terms of these. $\exists x P(x)$ is true if $P(x)$ is true for some assignment by I of an element of D to x; false if $P(x)$ is false for all such assignments to x, and undefined otherwise. $\forall x P(x)$ is defined as $\neg\exists x \neg P(x)$. Let $M =< D, I >$ be a model for a first-order language L, the variables ranging over D and n-ary predicates interpreted by totally defined n-ary relations on D. M is assumed to be fixed throughout the following definitions, and L is assumed to be enriched with some coding scheme allowing it to express its own syntax. L is then extended to include the partially defined monadic predicate T interpreted by the pair of sets (U, V), for $U, V \subseteq D$ and $U \cap V = \emptyset$. U is the extension of T and V is the anti-extension of T. Let κ be the valuation scheme for sentences in a model $M + (U, V)$, and let $\kappa_M(< U, V >) =< U', V' >$ where U' (V') is the set of sentences true (false) in the model $M + (U, V)$ according to κ_M. Say that $< U', V' >$ *extends* $< U, V >$, or $< U, V > \leq < U', V' >$, iff $U \subseteq U'$ and $V \subseteq V'$, and that κ is *monotonic* iff $\kappa_M(< U, V >) \leq \kappa_M(< U', V' >)$ if $< U, V > \leq < U', V' >$. Since Kleene's strong three-valued logic will be a monotonic operation on $\leq$, then although previously undefined sentences may receive a definite truth value in the process to be outlined, and so extend the interpretation of T, once a truth value is established it never changes or becomes undefined. In other words, $< U, V > \leq \kappa_M(< U, V >)$,

[26]Kleene 1952; section 64.

and so κ_M is an *increasing* function. We begin with the interpretation of truth as $< \emptyset, \emptyset >$. Applying κ to $M + (\emptyset, \emptyset)$ gives a set U of codes of sentences that are true in $M + (\emptyset, \emptyset)$, and a set V of codes of sentences that are false in $M + (\emptyset, \emptyset)$ or are not codes of sentences in $M + (\emptyset, \emptyset)$; and so at stage one we are to evaluate sentences in $M + (U, V)$, for here the interpretation of the truth predicate is $< U, V > (= \kappa_M(< \emptyset, \emptyset >))$. At the next stage the interpretation of the truth predicate is $\kappa_M(\kappa_M(< \emptyset, \emptyset >))$, and so on. Kripke proves that if this operation is performed often enough, perhaps transfinitely many times[27], there exists a minimal fixed point $\kappa_M(< U, V >) =< U, V >$ of κ_M, a point such that for any fixed point $< U', V' >$ of κ_M, $< U, V > \leq < U', V' >$. Here we are no longer able to assign truth values to any more of the statements, no matter how many more applications of κ_M we perform. A sentence α is grounded in a model M for L if for the minimal fixed point $< U, V >$ of κ_M , $\alpha \in U \cup V$; otherwise it is paradoxical. In particular the Liar sentence is paradoxical.

Using this construction it is now possible to define knowledge as true belief in such a way that the paradoxical knowledge sentences turn out to be ungrounded. To avoid Thomason's results, the concept of belief used to define knowledge must not satisfy some axiom of idealised belief, and so we reject the fourth axiom $B_a(B_a\alpha \supset \alpha)$. This resolution of the paradox therefore centres on the rôle of truth in the definition of knowledge, taking for granted the fact that the syntactic predicate 'believes' is always a well-defined relation between agents and codes of sentences, and does not lead to inconsistency. $K(a, < \alpha >)$ is then defined to be true at a stage in the construction if $B(a, < \alpha >)$ and α is in the extension of T at the previous stage; it is false if $B(a, < \alpha >)$ is false or if α is false at the

[27] At limit ordinals one takes the union of all sentences declared true or false at previous levels, so κ_M is a continuous function, and remains increasing. Such functions are called normal functions, and it is well known that all normal functions have arbitrarily large fixed points.

previous stage; and otherwise it is undefined. This construction is monotonic, and it eventually results in a fixed point for truth as well as for the set of what is known by any agent. The result is a model for knowledge where the paradoxical sentences are ungrounded.

2.2.5 A Classical Solution

A more sophisticated adaptation of a strategy for handling the Liar paradox has been proposed, one which also casts doubts on ($B4$) but may allow for its reinstatement in a weak sense. Asher and Kamp choose to adopt the method of Herzberger[28] and Gupta[29], and present their analysis within a traditional possible worlds framework. Put simply, the extensions of the predicate, according to this method, are classical rather than three-valued, and the stage-by-stage process that leads to the model of truth or belief is characterised in a quite different way to Kripke's strategy. Whereas Kripke used a cumulative procedure which assumed, according to Gupta, that truth had an *application procedure* associated with it which determined its extension in the world and then increased this through iteration, Gupta characterises the concept by a different procedure. It is not a cumulative procedure but rather a process of *revision* which underlies the concept; a rule allowing us to *improve* on any given proposal for the extension of truth by coming up with better candidates, rather than just bigger ones. Preservation of the genuine improvements brought about by repeated applications of the procedure is achieved by collecting through a stability property at limit ordinals the initial extension along with those sentences that eventually stop fluctuating in and out of the extension as we proceed through the previous stages of revision, subtracting from these those sentences that eventually never

---

[28]Herzberger 1982.

[29]Gupta 1982.

feature in any improvement. This decision is not final: at higher limit ordinals this collection of the effects of these improvements may turn out to be illusory, but in favourable conditions it can be shown that the revisions stabilise at a fixed point. Moreover this fixed point is the same no matter what initial extension was chosen - the arbitrary character of this set is made irrelevant by the revision rule.

This is the general method adopted by Asher and Kamp. Simplifying, their models for a first order language L with a one-place predicate B, interpreted as 'believes that' for a single agent, are of the form $M =< W, D, [\,], R >$: W is a non-empty set of worlds; D assigns to each world w a non-empty set, the universe of w - in fact M has a fixed universe, the universe of each of its worlds is the same; $[\,]$ assigns to each non-logical constant of L a classical extension at each world. For all $w, w' \in W$ and each individual constant c, $[c]_w = [c]_{w'}$, and, moreover, every sentence is in the fixed universe of M. They note that this fairly traditional-looking model allows for two methods of determining the truth value of $B\alpha$ at a world w: $B\alpha$ is true in M at w iff $< \alpha > \in [B]_{M,w}$; or alternatively, iff α is true at all w' such that Rww'. If these two methods are equivalent, the L-model is called *coherent*, but there are many models where this is not the case; the definitions can conflict, and then there is a problem about evaluating belief sentences. The method Asher and Kamp adopt is to evaluate sentences using the extension of the belief predicate. First, given a model M, they define for each ordinal γ the model M^γ, where $M^\gamma =< W_M, D_M, R_M, [\,]^\gamma >$, and where $[Q]^\gamma = [Q]_M$ for all non-logical constants Q other than B. Then:

1. $[B]^0_w = [B]_w$.

2. $[B]^{\gamma+1}_w = \{\alpha \mid \forall v(Rwv \Rightarrow [\alpha]_{M^\gamma,v} = 1)\}$.

3. For a limit ordinal $\gamma = \lambda$, $[B]^\lambda_w = \{\alpha \mid \exists\delta \leq \lambda \forall\gamma'(\delta \leq \gamma' < \lambda \Rightarrow \alpha \in [B]^{\gamma'}_w)\}$.

A ***model structure*** is a model in which B has not been interpreted. A model structure is called ***essentially incoherent*** if no coherent model can be formed from it. These clauses for B are the direct intensional analogue of Gupta's stability property of his revision procedure for collecting stably true sentences at limit ordinals, except for one thing: the initial extension of B is minimised in this definition, whereas Gupta's included its analogue in its definition.

Whether a model M becomes coherent after revision depends on three factors: the initial extension $[B]^0$; the constraints on R_M , and the forms of self-reference that are realised in M. Concentrating on the results of most relevance here, the following assignments to the constants σ and τ are the self-referential sentences to be considered: $[\sigma]_M = \neg B(\sigma)$ and $[\tau]_M = B(\tau)$. Let $<_M$ be the transitive closure of the relation holding between two constants c and d iff c names in M a sentence containing d. Then M is said to be *non-self-referential* iff $<_M$ is well founded. Given these definitions, the following propositions hold: if M is a non-self-referential model structure, then

1. There is an intension $[B]$ such that the model obtained by adding $[B]$ to M is coherent;

2. For any model obtained from M by adding intensions for B, there is a γ such that M^γ is coherent.

And secondly, if we suppose that a model is coherent and its relation is transitive, and also reflexive on its range[30], then all the instances of the axioms $1-4$ of Thomason's proof are true at all worlds in that model. These results are interesting with regard to how certain constraints can result in coherent models which contain a degree of self-reference. In a model structure M which contains τ but no other form of self-reference, it can be shown that if R is transitive and

[30]If xRy then yRy.

reflexive on its range, then any model obtained from M will be coherent after one revision. Without these assumptions on R, the model structure does not determine the truth value of τ, for this cannot here be determined without reference to the initial extension of B. What is required is an appropriate choice for $[B]$. It appears, then, that the axioms of idealised belief may be accommodated in a logic which contains a benign form of self-reference, although there remain important questions to be answered about the logic[31]. Noting that models with vicious self-reference of the form of the sentence σ start out incoherent and remain so upon revision, Asher and Kamp tentatively define a sort of stability which they can achieve, and this allows axioms $1-3$ to turn out valid, but yields counterinstances to 4. First they note the result that, for each L-model M there is a least ordinal γ such that

1. For each $w \in W_M$ and each sentence α, $\alpha \in [B]^{\gamma}_{M,w}$ iff $\forall \gamma' \geq \gamma, \alpha \in [B]^{\gamma'}_{M,w}$; and

2. There is an ordinal γ' such that for any $\delta_1, \delta_2 > \gamma$, if there is a $\xi < \gamma'$ and some π_1, π_2 such that $\delta_1 = \gamma'\pi_1 + \xi$ and $\delta_2 = \gamma'\pi_2 + \xi$, then for all $w \in W_M$, $[B]^{\delta_1}_w = [B]^{\delta_2}_w$.

Excluded from $[B]^{\gamma}_{M,w}$ are those sentences α such that $\alpha \notin [B]^{\gamma'}_{M,w}$ for *all* $\gamma' \geq \gamma$ - the anti-extension of B at w - as well as those that fluctuate in and out of $[B]_w$: those whose status never gets settled. The result captures the fact that the revision process becomes cyclical after γ with a fixed period. M^{γ} is called a *metastable model*, and the idea is to identify the valid sentences with those true in all metastable models. Making this assumption and constraining R to be transitive in the metastable model M^{γ}, then all instances of axioms $1-3$ are true at all worlds in $W_{M^{\gamma}}$. With further constraints 4 can be falsified. But

[31]The definition of logical validity is a case in point: see below.

constructing a partial model, perhaps using Kleene valuations, for the extension and anti-extension of B in M^{γ}, those sentences that continue to fluctuate after γ would not be assigned a truth value, and this allows a weak kind of reinstatement of the schema 4; the valid sentences could simply be identified as those that never come out false in any of these partial models associated with the metastable models, and this schema will never be false since it always lacks a truth value there.

2.2.6 Conclusions

This demonstration of how all four axiomatic schema may be reinstated in a logic where belief is a predicate of sentences, but where the semantics is partial and self-reference permitted, shows how little has to be given up to avoid the Lõb-Montague results - the absolute validity of the fourth axiom[32]. Given the logical framework used, it is an unavoidable consequence that logical validity must somehow be redefined: for example, it is not clear whether the notion of truth at all worlds in all models should include incoherent models, nor is it obvious how to motivate any particular choice of a distinguished subset of models. The strategy of modelling the structure of solutions to the problems revealed by the Lõb-Montague results on those techniques already developed to circumvent the Liar paradox is nevertheless a programme that promises to bear fruit. It remains the case that to treat semantic representations as syntactic objects leads, given certain assumptions, to a proof that the knowledge and beliefs of the ideal agent are inconsistent. And still, the assumptions opposed by Thomason, which regard the semantics so described as a theory of ideal semantic competence where this is presented as being in a psychological state which is the end process of systematic translation and calculation, are in need of

[32] And, for reasons not gone into here, Tarski's convention T.

urgent revision, in particular with regard to their close identification of theories of meaning with the ideal agent's semantic competence. The Lōb-Montague results continue to press home this philosophical point in spite of these technical changes which issue in a consistent theory.

There are many further issues within this field and areas of application for the models which we can only briefly mention. Some of the semantical theories, for example, have been given concrete interpretations as models of distributed systems. These consist of a set of processors connected by a communication network, the state of each processor being in part a function of its initial state and the messages it has received. By taking worlds in, for instance, a Kripke structure to be descriptions of the states of all processors at some particular time, the relation may then be defined for a processor i as holding between worlds if i is described as being in the same state in those worlds. Here the worlds of the Kripke structure are given a clear interpretation and the epistemic properties being modelled in this application might quite naturally include logical omniscience. So in this setting semantic possible worlds theories are very well suited to an analysis of reasoning and it is not difficult to find interpretations for more complex epistemic concepts which may also be applied in interpreting human reasoning. Common knowledge and mutual belief have a straightforward semantics in most of these propositional theories, and Vardi has shown how to accommodate within a Scott-Montague semantics comparative judgements of the knowledge states of an agent, as well as limits on his knowledge. Interpretations can be given to statements such as "I know more today than I did yesterday" and "I know that α, and that's *all* I know". Other issues we have not discussed include the work in artificial intelligence that has gone into exploring how knowledge affects action, where it is being recognised that finding models for agents who have partial knowledge of the problem domain and do not know all the axioms of the system will be necessary if agents are to have recognisably human powers of reasoning. Here care must be taken with regard to consequential closure in the theory, be it syntactic or semantic. And there are also problems that arise in characterising the state of knowledge of an agent after receiving information, especially if the communication medium allows the transmission of statements about knowledge.

These are just some of the ways in which it is possible to develop the epistemic logics presented here. Another obvious and superficially straightforward development is to extend the semantic theories to accommodate quantifiers. This move, however, introduces to the subject a host of new conceptual desiderata and, especially in the case of intensional theories, technical complications which would have demanded a lengthy separate treatment which space precludes. Identity and quantificational inference rules for quantified epistemic logic introduce questions, just to begin with, about what sort of entities are to be quantified over, and how to handle non-rigid terms and world-relative domains[33]. But having the means to express in a technically sound manner plausible forms of propositional reasoning promises to provide some useful conceptual and technical groundwork, as well as to give rise to a large and varied number of interpretations which are of worth in themselves.

---

[33]Garson 1984 gives a flavour of the kind of problems to be encountered in quantified modal logic even without worrying about problems such as logical omniscience; we sample some of these problems in chapter 3.

Chapter 3

Solutions To The Problem Of Logical Omniscience II

In this chapter are outlined several logics of belief, mainly designed for the specification of knowledge representation systems, which are based on modal logics and which, we argue, at least implicitly give *logically* motivated accounts of deductive reasoning in their attempts to avoid some of the problems of logical omniscience. Recall that these problems, traditionally inherent in epistemic logics, are that all valid formulae are believed, all the classical logical consequences of what is believed are also believed, and belief is closed under modus ponens in the sense that if α and $\alpha \supset \beta$ are believed, so is β. The thesis is that there are two identifiable logical tasks in providing a logic of belief with this objective. The first is to provide the logical principles of reasoning by means of which are generated the beliefs that a fairly rational agent must have, given certain others. This logic defining the valid argument schemata in reasoning shall be called the logic of commitment. Secondly, a language with a semantics which respects the logic of commitment must be provided in which the agent's beliefs can be exhibited, discussed, analysed logically for what they actually entail, and compared with what is actually the case. The first task works with what are known as the explicit beliefs of the agent, whereas the second objectifies this concept, and further introduces the concepts of implicit belief and truth. The first four logics presented are semantic accounts which may be regarded as recognising the importance of the first of these logical tasks; they make use of inconsistent and incomplete situations, using truth and falsity support relations in evaluating the sentences of the ground level language where beliefs are exhibited. With these, the objective occasioning the importance of the first task is that of a good computational performance in the intended knowledge base application of the logic, especially for sentences of the form $B\alpha \supset B\beta$ - B being the doxastic operator - where the consequences of a body of facts stored in a knowledge representation system must be retrieved in a reasonable amount of time. These requirements explain the emphasis on normalisation theorems for α and β, and the choice of relevance logic entailments between these normal forms which can be checked to hold be means of a simple and efficient decision procedure. This restriction to a proof-theoretic subsystem of classical logic is motivated in this context by the consideration that to retrieve the consequences of a body of facts stored in a knowledge representation system in a reasonable amount of time requires a weakening of the powerful deductive capabilities of classical logic in favour of the decidability and computational tractability of inference. The choice of this particular subsystem guarantees error-tolerance: that is, it ensures that the semantics does not sanction as correct any answer whatsoever to a query when the database contains contradictions, thus avoiding the intractable alternative of processing the whole database to eliminate contradictions before every query. Explicitly including a belief operator for explicit belief in the logic allows the notion of implicit belief to be modelled as long as there is also provided a definition of what it is for a situation or a world to be classical - that is, to be consistent and complete. Beliefs can then be analysed by means of classical logic, and it is straightforward to extend the logic to allow it to talk about non-epistemic truths. An aspect of this move essential for its success is the possibility of a classical definition of validity for modal as well as propositional formulae. Allowing embedding of operators further permits introspective reasoning, but given their intended use this aspect of the logics is widely ignored. The final logics also attempt to use explicit belief to avoid logical omniscience, but in two-valued systems which use a formal notion of awareness: it is shown how an important fragment of the first of these can be given a more revealing treatment when its semantics are standardised with those of the other logics; the chapter ends with a suggestion of a more semantically natural interpretation of the concept of awareness used in the second logic, which has a strongly syntactic flavour.

The general emphasis in these logics is on which beliefs must follow, given certain others - in other words on the logic of commitment for belief - given the fact that fairly straightforward deductions must correspond to the implications validated by the consequence relation for the logic of commitment. Thesishood is a far less important notion for logics of belief than is deducibility, or entailment - even if there are things *which* everybody knows, these are less interesting and less susceptible to logical treatment than is *how* they are known, that is the inference process by which new truths are discovered from old. This indicates one of the reasons why the usual conception of a logic of belief is not already realised in classical logic: if we took the material conditional to objectify the correct notion of inferential reasoning, then the deduction theorem would make dangerous connections between these two notions which a logic of belief ought to keep at a reasonable distance from each other. It should not in general be possible to use the material conditional as an object language correlate for the relation of consequence in the theory of deductive inference, where this is intended as a logic of commitment for belief. Thus epistemological considerations force a non-extensional characterisation of the deductive system here in the sense that its correct deductions cannot be characterised by, or in some sense reduced to, the set of formulae it validates: this set cannot serve as a criterion of identity, contrary to the case in classical systems of deduction. The aim should be to discover a weaker consequence relation and then most importantly, somehow to embed this metalinguistic relation into a classical or other base logic, so exhibiting the valid forms of reasoning. For if this were not possible we could not *use* the logic for the purposes in mind - the *discussion* of knowledge characterised as generated by this relation - for the relation would be flanked by either *mentioned* sentences or metavariables. This problem is of course not a new one; and since nearly all of the logics that are examined below can be seen, whether or not by design, as injecting a consequence relation into a classical object language, it is worth looking at the ideas underlying this procedure.

C. I. Lewis wanted his theory of strict implication to be a theory of deductive inference, where the implicative connective was flanked by 'unanalysed' propositional variables rather than by mentioned sentences or metavariables. Even if history judges that Lewis's analysis of deductive inference was faulty and that the concept is better handled by Gentzen's idea of conditional assertion, written $\vdash$, Lewis's approach is nevertheless still workable when corrected by Gentzen's analysis[1].

Leading up to the idea of using the relation $\vdash$ of Gentzen's conditional assertion as the implicative connective, the letters flanking $\vdash$ remain unanalysed propositional variables when it is allowed that they are potential bearers of truth values; that is, when we have in mind some set of intended valuations V from the language into a set of truth values, in terms of which the relation $\vdash$ is defined. In the general case, any set of truth values is permissible, and it need not be assumed either that variables *always* receive a value or that they receive a *unique* value. Given a particular definition with respect to some chosen set V, the properties of $\vdash$ may then be given in terms of true metalinguistic statements known as *inference rules*, these being true - or, perhaps more generally, designated - with respect to the choice of V. If a set of rules adequately axiomatises $\vdash$ in the sense that all other valid rules can be proved from them, then this provides the

---

[1]Scott 1971 is a very clear defence, against Quine, of this point of view, and on which much of the following relies. See especially his argument that any suggestion of use/mention conflation in the construction outlined below can be effectively countered.

basis for a general theory of deductive inference conceived as a metalinguistic activity. There is no prohibition on simultaneously defining two or more different consequence relations which meet the above specifications, although the metalanguage may have to state in which ways they relate to each other, and for theoretical simplicity as well as notational convenience this will nearly always require that there be one theory in terms of which others are defined. Given this, and instead of formalising the metalanguage, the Lewis approach is, *pace* Quine, to inject part of it back into the object language by means of non-Boolean operators, which are usually but not always unary: the language is extended so that it can itself reflect the permissible deductive inferences, according to the chosen relation $\vdash$. When the language is extended by the new syntax, there follows a second construction process whereby valuations are extended and the new operators are interpreted with reference to the values received by the operand under some (not necessarily proper) subset of the set of all the new valuations. At this stage there are many critical and subtle alternatives available. In the cases examined below, the syntax and semantics taken to be most convenient for the resulting system are those of modal logic, the formal calculus invented by Lewis for very similar purposes. And while of course Lewis's calculus has been developed into an extremely rich and productive area of semantic and philosophical research, it is possible to see the apparent innovations proposed in these cases as being merely the syntactic side-effects of grafting a chosen metalinguistic theory of consequence onto a classical language. The awkwardness of the models in their definition of validity certainly suggests this. And if this is so, the main source of value in studying them may be from the perspective that regards them as viable epistemological instances of a promising and more general approach to the disparate applications within this field of research; at the very least the simplicity of the higher level analysis provides a unified means of making sense of their various lower level complexities. If a logic of belief is seen in this way as built around a theory of reasoning, then the demands of both computational and conceptual motivations for avoiding logical omniscience could probably both be equally be satisfied, without serious loss of expressive power. As will be seen, the definition of belief could arise out of objectifying the semantics of a consequence relation chosen either for its syntactic association with proof-theoretic restrictions on the use of assumptions in a valid piece of reasoning, or because it modelled semantically the way in which the conceptual repertoire of an agent constrains his inferential abilities. These features of reasoning would then show up in the consequent definitions of belief, but the easiest point of comparison - and the most important given that they are designed to avoid logical omniscience - would be in terms of the consequence relation. This approach plays down the essential value of non-Boolean operators in understanding (as opposed to expressing) epistemic concepts, and so does not guarantee the meaningfulness of iterated beliefs, so these are for the most part omitted from this discussion.

3.1 Propositional Logics

Loosely following Barwise and Perry[2], Levesque[3] replaces possible worlds with *situations*, which are partial in the sense that they support not only the truth or falsity of sentences, but may support both of these or neither of them. In this way he can be seen to be building on the idea inherent in Hintikka that model sets need not be maximally consistent, but he goes further and abandons the constraint of downward saturation - and even that of consistency - which, as we saw, induces a form of logical omniscience in Hintikka's model systems. Explicit belief is to be identified with sets of situations rather than sets of possible worlds, where in Levesque's terminology a *possible world* is a complete situation, a situation which supports either the truth or falsity of every primitive formula, but

[2]Barwise and Perry 1983.

[3]Levesque 1984.

never both. He uses situations rather than possible worlds because $\{\alpha, \alpha \supset \beta\}$ and $\{\alpha, \alpha \supset \beta, \beta\}$ are true in exactly the same downwardly saturated possible worlds and so are semantically equivalent; belief would then be closed under implication and we have succumbed to one of the problems of logical omniscience. Identifying belief with sets of incomplete situations rather than sets of complete possible worlds is the method Levesque adopts to circumvent this problem, for sentences not relevant to what is explicitly believed, perhaps including tautologies, need not be assigned a truth value in partial situations. His logic is limited to one agent and does not allow for the embedding of operators. It defines the set L of formulae from the standard connectives $\wedge$, $\vee$ and $\neg$, as well as the unary modal operators B and L, standing for, respectively, explicit and implicit belief. Non-atomic formulae are formed from the set P of atomic formulas in the usual way, except that only formulae without a B or an L can occur within the scope of a B or an L. Levesque's model is a structure $M =< S, B, T, F >$ where S is a set of situations, $\emptyset \neq B \subseteq$S are the situations considered possible, and T and F are functions from P to 2^S. $T(p_i)$ (respectively $F(p_i)$) are the situations which support the truth (falsity) of p_i. Some comments on these structures are in order. Typically modal logics use a relational semantics based on the notion of a Kripke frame, where a binary relation R is defined over the universe S to be used in the valuation clause for sentences containing the modal operator, in such a way that the truth at some $s \in S$ of a modal formula depends in some way on the truth of the embedded formula at members of the set $\{t \mid sRt\}$. There are however two features of Levesque's logic which allow the set B to suffice. No embedding of operators is allowed, and so we may simply associate a set of situations with each situation: this restriction stipulates that there will be no need to look further at the set of situations associated with each member of that set. And since, in addition, the situations believed possible at all $s \in S$ are taken to to comprise the same set, it is technically sufficient to use the one set $B \subseteq S$ instead of a relation. If meta-beliefs were allowed into the logic without changing the model structures and a relational notation were introduced, then for any s the set $\{t \mid sRt\}$ of situations believed possible at s would always be the whole of the set B; the restriction of R to $\{t \mid sRt\}$ would then be an equivalence relation, which is the characteristic feature of the modal logic weak $S5$. For future reference, this logic is obtained by adding to the minimal normal modal logic K the axiom schemes $B\alpha \supset BB\alpha$, $\neg B\alpha \supset B\neg B\alpha$, and $\neg B(\textbf{false})$, where **false** is the false propositional constant. It is sound and complete for transitive, Euclidean and serial Kripke frames. This is in fact the preferred set of constraints on the relational models given later in this chapter.

To illustrate some features of the logic which will arise from its use of the two assignment functions, it may be helpful to look at one way in which the usual notion of a proposition might be expressed here. It clearly cannot be a function from the set S of situations to $\{T, F\}$, the set of truth values: there is no requirement that for any s and p_i, s must belong to one and only one of $T(p_i)$ or $F(p_i)$, and since the proposition determined by any function associated with p_i would take situations to unique truth values, this is too restrictive a notion to give what is required. Moreover, sentences may receive no value at situations. This may suggest that the range of the function should instead be $2^{\{T,F\}}$, allowing in addition for the assignment of both truth values and no truth value to sentences, but as it stands this approach is slightly misleading in this context, although it is easier to work with and so will be the perspective we adopt. It does however disguise the fact that if a sentence is assigned both-T-and-F (one of the four truth values), then it is both T and F (two of the other truth values). The point is that as things stand the functional approach creates four distinct truth values - the members of $2^{\{T,F\}}$ - whereas, at least from the perspective of these models, only two are required: the classical values without their classical assumptions of disjointness and exhaustiveness. One natural approach which satisfies these criteria is to see a proposition p_i as being *a relation* between the set S of situations and the set $\{T, F\}$, to be represented by the value pair, or as it is technically known the *polarity* $(T(p_i), F(p_i))$[4]. This fits in neatly with the idea that the support relation $\models$ for models to be defined below may be thought of as a three place relation between situations, formulae and truth values, in which pairs of situations and formulae need not be related to a unique truth value: thus ambivaluation, or two clauses for each connective, is used in this definition, and this is heuristically useful in explaining why logical omniscience fails.

A *possible world* s is defined to be a situation s such that $s \in T(p_i)$ or $s \in F(p_i)$, but not both, for every $p_i \in P$. A situation s is *compatible* with a situation s' if they agree wherever s is defined, that is:

1. If $s \in T(p_i)$, then $s' \in T(p_i)$, for all $p_i \in P$, and

2. if $s \in F(p_i)$, then $s' \in F(p_i)$, for all $p_i \in P$.

In particular, a possible world will be compatible with a situation if every sentence whose truth is supported by the situation comes out true in the possible world, and every sentence whose falsity is supported comes out false. We may now see possible worlds as the limiting case of situations where these are consistent and every sentence has a truth value. Some other logics which use similar semantic techniques exploit the fact that situation are partially ordered by the compatibility relation, which describes stable improvements in the specification of values associated with situations, in order to define a classical semantics. Under certain other assumptions the classical worlds compatible with a situation may be defined as the related complete value specifications, but this more

[4]This idea has appeared in a number of places; the references most appropriate in this context are perhaps Blamey 1986, in his analysis of the work of Barwise and Perry, and, especially, Dunn 1976 and 1986. It is in fact technically equivalent to the method that will eventually be adopted here, so elaboration is not mathematically necessary.

constructive-looking definition is not possible here: because of the possibility of inconsistent situations not every situation has a classical completion, so the definition here is less elegant. Levesque's application of the notion of compatibility is used in the following definition, which will be used to define implicit belief. $W(B)$ is the set of all possible worlds with which some situation in B is compatible; that is, the set of all those maximally consistent situations which agree with some situation in B wherever that situation is defined. Finally the support relations $\models_T, \models_F \subseteq S \times L$ are defined, given a model M, as follows:

$s \models_T p_i$ iff $s \in T(p_i)$	$s \models_F p_i$ iff $s \in F(p_i)$
$s \models_T \alpha \vee \beta$ iff $s \models_T \alpha$ or $s \models_T \beta$	$s \models_F \alpha \vee \beta$ iff $s \models_F \alpha$ and $s \models_F \beta$
$s \models_T \alpha \wedge \beta$ iff $s \models_T \alpha$ and $s \models_T \beta$	$s \models_F \alpha \wedge \beta$ iff $s \models_F \alpha$ or $s \models_F \beta$
$s \models_T \neg\alpha$ iff $s \models_F \alpha$	$s \models_F \neg\alpha$ iff $s \models_T \alpha$
$s \models_T B\alpha$ iff for all $s' \in B, s' \models_T \alpha$	$s \models_F B\alpha$ iff $s \not\models_T B\alpha$
$s \models_T L\alpha$ iff for all $s' \in W(B), s' \models_T \alpha$	$s \models_F L\alpha$ iff $s \not\models_T L\alpha$

Given a model M, α is true at a situation s if $s \models_T \alpha$. α is valid - $\models \alpha$ - if for any $M =< S, B, T, F >$ and any $s \in W(S)$ - any possible world - α is true at s. So although truth is defined for all situations, when checking for validity only complete situations or possible worlds are considered, and so all classically valid propositional sentences come out valid in Levesque's logic. Thus although the truth of $p_i \vee \neg p_i$ may not be supported by some situation in a model this does not affect its validity: the notion of partial situations turns out to be redundant in this respect. Belief is defined as truth in all members of B, and in the definition of failure to believe a connection is introduced between the truth and the falsity of these modal sentences so that $B\alpha \vee \neg B\alpha$ is valid for all α. There are however models which reject $B\alpha \vee B\neg\alpha$ and models which accept $B\alpha \wedge B\neg\alpha$, so the logic of commitment of the agent is very different from its external characterisation, the classical logic in which it has been embedded. The meta-theoretical nature of the B operator that this seems to suggest will be further considered below. To establish whether some formula is implicitly believed is to determine whether it holds in all members of $W(B)$: as noted above, this involves disregarding inconsistent members of B and blowing up the rest to Henkin sets, and has the effect of introducing as truths common to all $t \in W(B)$ all propositional tautologies and all classical consequences of the formulae true in all $s \in B$.

It can be shown that $\models B\alpha \supset L\alpha$ - if a sentence is a logical consequence of what is explicitly believed then it is implicitly believed - and implicit belief is closed under implication and all valid formulas are implicitly believed. But necessary truths which may not logically valid are also implicitly believed, so there may be sentences true in the correct set of worlds which are not in fact implied by what is explicitly believed. In other words the following hold in the logic:

1. If $\models \alpha$ then $\models L\alpha$;
2. $(L\alpha \wedge (\alpha \supset \beta)) \supset L\beta$, where $\alpha \supset \beta$ is a classically valid formula;
3. $(L\alpha \wedge L(\alpha \supset \beta)) \supset L\beta$.

This is not, however, the case with explicit belief. $B\alpha \wedge B(\alpha \supset \beta) \wedge \neg B\beta$ is satisfiable, and so explicit belief is not closed under implication; and $\neg B(\alpha \wedge \neg\alpha)$ is satisfiable, so valid sentences need not be believed. Moreover, $\{B\alpha, \neg B(\alpha \wedge (\beta \vee \neg\beta))\}$ is satisfiable, so classical logical equivalents to a belief need not be believed. These are possible because, through the use of partial situations the agent may not be *aware* of the concept involved: there may be some situation relevant to evaluating the sentence which supports neither the truth not the falsity of α.

The tactic of allowing non-classical or impossible worlds into the semantics has had some philosphical defence in the literature, and may be wanted here. For while $(B\alpha \wedge B(\alpha \supset \beta)) \supset B\beta$ is not valid in his logic, $(B\alpha \wedge B(\alpha \supset \beta)) \supset B(\beta \vee (\alpha \wedge \neg\alpha))$ *is* valid, so closure of beliefs under implication is avoided at the cost of allowing the agent to believe possible some incoherent situation - one that is a member of $T(p_i) \cap F(p_i)$, for some atomic formula p_i. This ploy is also used by others, such as Cresswell[5], Hintikka[6] and Rescher and Brandon[7] to avoid logical omniscience; to choose just one of these advocates, Hintikka argues for it as follows. Hintikka's philosophical interpretation of the situation in possible world semantics, whereby what an agent a knows in a world is uniquely determined by a set of "epistemic a-alternatives" to that world is an idealistic one. The set of worlds presupposed in talking of a's knowledge constitute the restriction of a's attention to a subset of all possible contingencies, as dictated by the amount of information he has, and as allowed by his logical and conceptual abilities, as far as he can determine. It is not all those situations in which a set of sentences is true, but rather the epistemic alternatives are delimited by those contingencies he *envisages* to be possibly true. Some contingency may look possible to him, and therefore not be eliminated by him due to any doubts, even though it contains an unseen contradiction. His failure to eliminate this from consideration makes it a legitimate member of the set of epistemic a-alternatives since it is indeed an epistemic alternative, given the amount of information he has. To suppose that every epistemic alternative to a given world must be logically possible is to presuppose that an agent is able to eliminate merely apparent possibilities, and this *assumes* logical omniscience on his part. $\alpha \supset \beta$ is valid if it holds in all members of a subset of possible worlds - all those that are not 'impossible possible worlds'. But the operator is defined in terms of all worlds that are epistemic alternatives to the world at which it is being evaluated, and these alternatives may be drawn from the wider set of *all* possible worlds, whether they are logically possible or impossible. Epistemic possibility applies to a larger class of worlds

---

[5]Cresswell 1973.

[6]Hintikka 1975. See chapter 2.

[7]Rescher and Brandon 1979.

than does logical possibility, and so to admit as epistemically possible worlds which are in fact not logically possible can now block the inference from the logically true $\alpha \supset \beta$ to $K_a\alpha \supset K_a\beta$. $K_a\beta$ may be false while $K_a\alpha$ is true as the following counterexample shows: if $\neg K_a\beta$ is true at w then there is some epistemic alternative w' where $\neg\beta$ holds, and supposing $K_a\alpha$ to be true in all epistemic alternatives, α is true in particular in w'. Now if $\alpha \supset \beta$ is logically true it is true in all logically possible worlds, but w', where α and $\neg\beta$ are true, may not be logically possible even though it is epistemically possible, so the counterexample is not inconsistent.

Returning to Levesque's logic, The fact that valid propositional formulae need not be true everywhere suggests that were we to look for the notion of logical consequence inherent in these model structures, it may not coincide with that of propositional logic. That is, if we introduce a two place relation $\models$ on propositional formulae with the intended meaning that $\alpha \models \beta$ iff for any $s \in S$ and any pair $< T, F >$ of valuation functions, whenever α is true at some $s \in S$ so is β and whenever β is false so is α, then this will not turn out to be a classical consequence relation; indeed, validities and semantic consequences turn out to have different logical bases, and an examination of the latter is an informative method of seeing how the semantics have been tailored to meet the requirements of the logic of commitment. An alternative way of describing what is required to determine the relation is as something to tell us how to determine the *theory* of a situation, taken as a set of formulae, in the same way that a classical consequence relation tells us something about the propositional fragment of the theory of a Kripke world. The other validities of the logic indicate the nature of this consequence relation. First note that the restriction to $W(S)$ in determining the validity of formulae which contain only truth functions of modal formulae is not strictly necessary: their valuation in any situation whatsoever refers only to the sets B and $W(B)$, so they will be true in all $s \in W(S)$ if and only if they are true in all $s \in S$. Their location is only one of convenience, since that is where the non-modal validities are true, and so the sense in which a formula of the language is true in a model appears to involve two distinct methods of evaluation, only one of which entails its truth at all situations in the model. For example, $B\alpha \supset B\beta$ iff for all structures $< S, B, T, F >$, $\forall t \in B(t \models_T \alpha \Rightarrow t \models_T \beta)$; and so by definition of the operator B, $s \models_T B\alpha \supset B\beta$ for all models M and all $s \in S$. Validities of this form are true at all situations in all models, and so would appear to point the way to the natural notion of consequence in the model, the idea that β is a consequence of α if every pair $< T, F >$ which make α true also make β true. Indeed, if $B\alpha \supset B\beta$ is a valid formula, then for any structure $M =< S, B, T, F >$ and any $s \in S$, if $s \models_T \alpha$ then $s \models_T \beta$ and if $s \models_F \beta$ then $s \models_F \alpha$. This is the natural consequence relation of the model given its use of inconsistent and incomplete situations. This parallel would of course break down if iterations were permitted in the logic: for then there would, for example, be no consequence relation corresponding to $B(B\alpha \supset B\beta) \supset B\alpha$, since this would illegitimately have to relate a propositional formula and another relation. Thus the ban on iteration is further evidence for the thought that valid formulae of the form $B\alpha \supset B\beta$ are in fact intended to reflect a pre-existing relation on the set of formulae, the natural consequence relation of the models. It is also of some interest, as well as further support for this view, that simple set-theoretic relations among the collection of propositions or value pairs of the form $(T(\alpha), F(\alpha))$ can equally be used to represent semantic behaviour of modal formulae of the type in question - technically, it can be shown that these set-theoretic operations together with the set of all polarities in any ring of subsets of S form what is called a field of polarities, and that every de Morgan lattice, which is the characteristic algebra for the logic of commitment, is isomorphic to a field of polarities. The least natural part of the logic - that which requires most tinkering with the model structures - is the means by which classical logic is recovered from this framework in order to define implicit belief and propositional validity. A parallel definition of classical consequence by means of $L\alpha \supset L\beta$ would not throw any light on the structure of situations in the model. By way of example of its theoretical naturalness, it is straightforward to verify that it is reflexive and transitive; that contraposition holds - if $\alpha \models \beta$ then $\neg\beta \models \neg\alpha$; and that $\alpha \models \beta$ iff $\alpha \models \alpha \wedge \beta$ and $\alpha \wedge \beta \models \alpha$ iff $\beta \models \alpha \vee \beta$ and $\alpha \vee \beta \models \beta$. Some consideration of this relation in fact shows that sentences of the form $B\alpha \supset B\beta$ are valid iff $\alpha \rightarrow \beta$ is valid in the logic $\mathbf{E}_{fde}$ of first degree entailment in relevance logic, so this logic completely characterises this fragment, the logic of commitment for belief[8]. The α and β may be put into normal form just as in classical logic since this logic shares the axioms for commutativity, associativity, distributivity, double negation and the De Morgan laws; we shall see later why these properties are indispensable for Levesque's purposes. Also all explicit beliefs are implicitly believed, and implicit belief is closed under classical modus ponens. Inconsistent beliefs are permitted since an opinion on some part of the world need have no relation to its negation. This is illustrated by an example showing that explicit belief is not closed under modus ponens. Suppose that $B\alpha$ and $B(\neg\alpha \vee \beta)$: this is satisfied by the model where for all $s \in B$ $s \models_T \alpha \wedge (\neg\alpha \vee \beta)$. But this does not mean that $s \models_T \beta$ for this sentence may be true at some s such that $s \models_T \alpha$ and $s \models_F \alpha$ but $s \not\models_T \beta$. It is also of note that if α is a classically valid propositional formula and if $B(p_i \vee \neg p_i)$ for every propositional variable p_i occurring in α, then $B\alpha$. $B(p_i \vee \neg p_i)$ may be regarded as stating that the reasoner is 'aware' of p_i. This assumption induces belief in all valid statements, but does not rule

---

[8]The deductive system may be found in many places, for example Anderson and Belnap 1975. Proof-theoretic characterisations of this relation and other relevance entailment relations exist in the literature: Dunn 1976 gives a simple and informative proof-theoretic restriction which matches entailment in this logic; Prawitz 1965 ch. 7 does the same for another relevance logic, Church's theory of weak implication (Church 1951); the λI calculus yields the relevance logic $\mathbf{R}$ (see, for example, Helman 1977), and additional restrictions on abstractions can give proof systems for other relevance logics, as in Mitchell and O'Donnell 1986. We give a sequent calculus formulation in chapter 4.

out belief in unsatisfiable statements or lack of closure under implication: as we have seen, these are due to the presence of incoherent rather than incomplete situations.

The technical comment which is now sketched without proof reinforces the point that it is the logic under which explicit belief is closed that is really important in this modal logic: disregarding the operator L, the theory of any model M can be given a non-modal characterisation. The reader would miss little conceptually if he omitted the following paragraph, but the definitions of the technical terms mentioned in this aside may be found in the later chapters.

First, suppose that the language is based on γ propositional variables, and note that in any model M there are eight different 'values' that a formula may have depending on whether or not it is true, whether or not it is believed and whether or not its negation is believed. Let $\mathbf{2}$ be the two element characteristic algebra for classical logic: the corresponding characteristic algebra for the relevance logic under which belief is closed in all models M is the four element algebra $\mathbf{4}$ with universe $\{0, a = \neg a, b = \neg b, 1\}$ such that $a \wedge b = 0$ and $a \vee b = 1$. Where $F_{\mathbf{4}}^{\gamma}$ is the Lindenbaum algebra of this logic, for any model M define a homomorphism $F_{\mathbf{4}}^{\gamma} \xrightarrow{h^M} \mathbf{2} \times \mathbf{4}$ as follows: where π_i for $i \in \{1,2\}$ are the projections

1. If $M \models \alpha$ then $\pi_1 \circ h^M([\alpha]) = 1$.

2. If $M \models B\alpha$ then $\pi_2 \circ h^M([\alpha]) \in \{a, 1\}$.

3. If $M \models B\neg\alpha$ then $\pi_2 \circ h^M([\alpha]) \in \{a, 0\}$.

Then it is straightforward to check that h^M is a well-defined surjective homomorphism and that the converses of the above conditions also hold. Conversely, any surjective $F_{\mathbf{4}}^{\gamma} \xrightarrow{h} \mathbf{2} \times \mathbf{4}$ determines a model M^h with universe the prime filters of $F_{\mathbf{4}}^{\gamma}$. A prime filter F is in B if for each $[\alpha] \in F$ we have $\pi_2 \circ h^M([\alpha]) \in \{a, 1\}$; $T(p_i)$ is the set $\{F \mid [p_i] \in F\}$ and $F(p_i)$ is the set $\{F \mid [\neg p_i] \in F\}$. Given these two constructions and some such h, we then have $h^{M^h} = h$.

3.2 Iterated Beliefs

Obvious extensions of this logic include allowing the embedding of modal operators, or introducing quantifiers into the language. This last task first requires extending the relevance logic under which beliefs are closed with the introduction of quantifiers. We now outline three such attempts, all of which were done with the same computational intention as Levesque, and so can be regarded as emphasising the logic of commitment.

Lakemeyer[9] conservatively extends this logic to permit embeddings, thus preserving as valid all the axioms in Levesque's logic. In context, this extension to accommodate meta-reasoning permits, for example, the logic to model planning actions. The decision procedure for determining entailments also extends that of Levesque, so to illuminate this context it will be described below. It also illustrates some of the complexities involved when attempting to accommodate both axioms governing iterated beliefs and a logic of commitment with a fast decision procedure in a logic. Certain obvious changes to Levesque's logic are necessary: accessibility relations are required, rather than simply a set of situations thought possible, and these are also used in the definition of a possible world. There are two different relations for evaluating positive and negative beliefs but which coincide from the point of view of a possible world, and implicit belief is defined by means of the possible worlds accessible rather than those compatible with the world of evaluation. It is also of note that the operator L cannot occur within the scope of the operator B because L is to be viewed as a purely external characterisation of an agent's beliefs and what follows from them.

---

[9]Lakemeyer 1987.

Lakemeyer's models for his language BLK based on the modal logic K are tuples $M = < S, T, F, R, R^- >$. S, T and F are as in Levesque's logic, and conditions on the binary relations R and R^- first require the definition of a classical situation or, as it is called here, a *world.* $w \in S$ is a world if and only if

1. $w \in T(p_i)$ iff $w \notin F(p_i)$ for all atomic formulae p_i, and

2. for all $s \in S$, wRs iff wR^-s.

The relations of a BLK-model are then constrained by the following versions of the transitivity and Euclidean relations, where w_1 and w_2 are worlds and s is a situation:

1. if w_1Rw_2 and w_2Rs then w_1Rs.

2. if w_1Rw_2 and w_1Rs then w_2Rs.

Only the clauses for the epistemic operators are different in BLK:

$$s \models_T B\alpha \text{ iff for all } t, \text{ if } sRt \text{ then } t \models_T \alpha.$$
$$s \models_F B\alpha \text{ iff for some } t,\ sR^-t \text{ and } t \not\models_T \alpha.$$
$$s \models_T L\alpha \text{ iff for all worlds } w, \text{ if } sRw \text{ then } w \models_T \alpha.$$
$$s \models_F L\alpha \text{ iff } s \not\models_T L\alpha.$$

As in Levesque's logic, validity is defined as truth in all models and all *worlds*, and without nested modal operators BLK reduces to that logic. In particular, it also has the feature that closure of explicit belief under modus ponens fails only where some inconsistent situation is considered possible. Although from the point of view of the (classical) worlds the accessibility relation is constrained, outside these classical worlds no relational structure is specified - there is no transitivity and so no introspective reasoning with explicit belief. To extend BLK to allow for this requires more than making R transitive, for this leaves negative beliefs untouched: for example, $BB\alpha \supset BBB\alpha$ would be valid, whereas

$B\neg BB\alpha \supset B\neg B\alpha$ could fail. R and R^- are suitably related by also imposing the mixed transitivity condition that for all $s, t, u \in S$, if sR^-t and tRu, then sR^-u. This logic he calls $BL4$, and like BLK it is sound and complete. Even this extension is not too powerful, however, as is shown by the satisfiability of Moore's paradoxical sentence $B(\alpha \wedge \neg B\alpha)$.

3.3 Variable-Sharing Decision Procedures

These logics are intended to serve as the external and autonomous model-theoretic semantics for knowledge representation systems, independent of the syntax of a particular implementation; they are the means by which the stored syntactic expressions of the system can be regarded as facts or putative facts, and all the operations it can perform on these correspond exactly to semantic procedures in the logic, for example logical consequence, and so are characterised exactly by the logic. When this is achieved the response of the system with a set of stored expressions to the input of an expression can be understood as an answer to a query on the basis of the facts it believes: where **KB** is the conjunction of the stored facts and α is the query, the procedure leading to the answer corresponds to $\models B\mathbf{KB} \supset B\alpha$. One way in which to make the knowledge-base system efficient is to provide it with a syntactic decision procedure in exact correspondence with implications of this form, and the equivalence of $\models B\mathbf{KB} \supset B\alpha$ to the entailment relation $\mathbf{KB} \longrightarrow \alpha$ in our form of relevance logic provides just such a procedure. Let $\mathcal{L}(\alpha)$ stand for the set of literals of the propositional formula α. Then where α is in disjunctive normal form, and β is in conjunctive normal form, Anderson and Belnap's[10] decision procedure for determining whether α

[10] Anderson and Belnap 1975.

entails β in this logic is as follows: where $\alpha = \alpha_1 \vee \ldots \vee \alpha_n$ and $\beta = \beta_1 \wedge \ldots \wedge \beta_m$,

$$\alpha \longrightarrow \beta \text{ iff for all } i,\ 1 \leq i \leq n, \text{ and all } j,\ 1 \leq j \leq m, \mathcal{L}(\alpha_i) \cap \mathcal{L}(\beta_j) \neq \emptyset.$$

In computational practice, conversion of formulas into two different normal forms is not desirable, especially in view of the following similar decision procedure for this entailment which is effectively equivalent. If α and β are both in conjunctive normal form and $\alpha = \alpha_1 \wedge \ldots \wedge \alpha_n$ and $\beta = \beta_1 \wedge \ldots \wedge \beta_m$, then

$$\alpha \longrightarrow \beta \text{ iff for all } j,\ 1 \leq j \leq m, \text{ there is some } i,\ 1 \leq i \leq n \text{ with } \mathcal{L}(\alpha_i) \subseteq \mathcal{L}(\beta_j).$$

This is the decision procedure used by Levesque[11], and extended by Lakemeyer. A high priority of Lakemeyer's extension of this logic to accommodate meta-reasoning, if not its raison d'être , is the retention of this algorithm in a modified form. Obviously this first requires a different normal form to cope with the presence of belief operators; this he calls extended conjunctive normal form, or $ECNF$, and is an analogue of one used by Dunn[12] in connection with the relevance logic $\mathbf{E}_{fdf}$: α is called an *extended clause* iff α is of the form $\alpha_1 \vee \ldots \vee \alpha_n$, where each α_i is a literal or is of the form $B\beta$ or $\neg B\beta$, where β is an extended clause. A sentence α is then said to be in $ECNF$ iff α is of the form $\alpha_1 \wedge \ldots \wedge \alpha_n$, where for each i, $1 \leq i \leq n$, α_i is an extended clause. Every formula of the language can be converted into $ECNF$. The key to the algorithm is essentially the following straightforward result: where α and β are extended clauses of the form

$$\alpha_1 \vee \ldots \vee \alpha_{k-1} \vee B\alpha_k \vee \ldots \vee B\alpha_{l-1} \vee \neg B\alpha_l \vee \ldots \vee \neg B\alpha_{m-1}$$

---

[11] In his logic $\models B\alpha \supset B\beta$ if and only if α entails β in this logic, so where **KB**, the knowledge base, and α, the query, are in conjunctive normal form, the computation $\models B\mathbf{KB} \supset B\alpha$ has a tractable $O(\mid \mathbf{KB} \mid, \mid \alpha \mid)$ algorithm, taking only time proportional to the product of the sizes of **KB** and α.

[12] See Dunn 1986.

and

$$\beta_1 \vee \ldots \vee \beta_{n-1} \vee B\beta_n \vee \ldots \vee B\beta_{o-1} \vee \neg B\beta_o \vee \ldots \vee \neg B\beta_{p-1}$$

respectively, then $\models B\alpha \supset B\beta$ if and only if

1. For all i, $1 \leq i < k$, there is some j, $1 \leq j < n$ such that $\alpha_i = \beta_j$;

2. for all i, $k \leq i < l$, there is some j, $n \leq j < o$ such that $\models B\alpha_i \supset B\beta_j$; and

3. for all i, $l \leq i < m$, there is some j, $o \leq j < p$ such that $\models B\beta_j \supset B\alpha_i$.

This can be immediately extended to a corollary about formulas in $ECNF$: where $\alpha = \alpha_1 \wedge \ldots \wedge \alpha_n$ and $\beta = \beta_1 \wedge \ldots \wedge \beta_m$ are in $ECNF$, then $\models B\alpha \supset B\beta$ if and only if for each β_j there is some α_i such that $\models B\alpha_i \supset B\beta_j$. Assuming that formulas are in normal form, as with Levesque's logic the same complexity result can be shown to hold both in BLK and $BL4$ as was found for Levesque's logic.

3.4 Quantification

The first extension of these ideas to a similarly decidable first-order language to be considered does not attempt to embed the consequence relation into a base classical language: the sentences are simply considered as *believed* sentences, as if there were a suppressed B indicating this fact before each of them. This inevitably leads to a loss of expressive power, for since the inferential process of reasoning cannot be expressed in the language it cannot be characterised externally and so, for example, compared with the actual state of affairs. The logic that is given is nothing more than what we have been calling the logic of commitment for belief. Features of this logic are retained in the second logic of belief considered here, which attempts the process of objectification by injecting this logic of commitment into the classical language: the operators required for this purpose introduce the familiar semantic obligation to provide answers to question about the relation between de re and de dicto expressions.

Patel-Schneider[13] extends Levesque's logic to the first order case in the following way. A situation s now consists of a non-empty set D, the domain of the situation, mappings $\mathbf{f}_s$ and $\mathbf{t}_s$ from n-ary predicate letters P_n into n-ary relations on D and the mapping h from n-placed function letters into functions from D^n to D. Where $d \in D$, x is a variable and v a variable map into D, $v(x/d)$ is defined by $v(x/d)(y) = d$, if $y = x$, and $v(x/d)(y) = v(y)$ otherwise. Given a situation s and a variable map v, a mapping v_s^* from terms into the domain of s is defined as:

$$v_s^*(x) = v(x) \text{ if } x \text{ is a variable;}$$
$$v_s^*(f_n(t_1, \ldots t_n)) = (h(f_n))(v_s^*(t_1), \ldots v_s^*(t_n)) \text{ otherwise..}$$

For the base case, the support relations are defined by

$$s, v \models_T P_n(t_1, \ldots t_n) \text{ iff } (v_s^*(t_1), \ldots v_s^*(t_n)) \in \mathbf{t}_s(P_n)$$
$$s, v \models_F P_n(t_1, \ldots t_n) \text{ iff } (v_s^*(t_1), \ldots v_s^*(t_n)) \in \mathbf{f}_s(P_n)$$

where s is a situation, v is a variable map into the domain of s, P_n is a predicate letter and t_i is a term. These are extended to the truth functions in the following way:

$s, v \models_T \alpha \vee \beta$ iff $s, v \models_T \alpha$ or $s, v \models_T \beta$	$s, v \models_F \alpha \vee \beta$ iff $s, v \models_F \alpha$ and $s, v \models_F \beta$
$s, v \models_T \alpha \wedge \beta$ iff $s, v \models_T \alpha$ and $s, v \models_T \beta$	$s, v \models_F \alpha \wedge \beta$ iff $s, v \models_F \alpha$ or $s, v \models_F \beta$
$s, v \models_T \neg\alpha$ iff $s, v \models_F \alpha$	$s, v \models_F \neg\alpha$ iff $s, v \models_T \alpha$

So the meaning of the connectives is familiar from the other logics. Patel-Schneider then considers perhaps the most natural definitions for the quantifiers,

---

[13]Patel-Schneider 1985.

which result in first-order analogues of Levesque's situations:

$$s, v \models_T \forall x\alpha \text{ iff for all } d \in D,\ s, v(x/d) \models_T \alpha.$$
$$s, v \models_F \forall x\alpha \text{ iff for some } d \in D,\ s, v(x/d) \models_F \alpha.$$
$$s, v \models_T \exists x\alpha \text{ iff for some } d \in D,\ s, v(x/d) \models_T \alpha.$$
$$s, v \models_F \exists x\alpha \text{ iff for all } d \in D,\ s, v(x/d) \models_F \alpha.$$

But he discovers that a decidable entailment algorithm is not available if $Pa \vee Pb$ entails $\exists xPx$, so to block this inference the classical assumption of the material equivalence of the existential quantifier and an infinite disjunction should not be carried over into this logic, although it is plausible to suppose that the former should entail the latter. The same of course applies to universal quantification and infinite conjunction. This inequality can be effected without altering the form of the truth definition, but by changing certain assumptions underlying that definition. What is required is to allow for the possibility that the infinite disjunction of instantiations of $\exists xPx$ be true at a situation without $\exists xPx$ itself being true, while of course still permitting the existential generalisation inference from Pa to $\exists xPx$. To clarify the problem first consider for a situation the theory that is the set of quantifier-free formulae holding true at the situation. As with Levesque's logic, the theory thus identified is prime, but it is just this property, given existential generalisation, that enables the inference from the infinite conjunction to the existential, and so effectively identifies the two. So the first step in separating the two notions is to follow Levesque and abandon the assumption that the formulae believed at a situation form a prime theory. The second step is to reintroduce a restricted form of $\vee$ -elimination in order to re-identify the two notions in precisely the cases where a theory containing an infinite disjunction also happens to contain at least one of its disjuncts, for then existential generalisation applies. Note that a theory contains one of these disjuncts just in case every one of its prime extensions in the same domain also contains it, so the separation of infinite disjunction and the existential quantifier can be effected if we no longer assume that the theories of the situations with respect to which we evaluate are prime, and if quantification introduces reference to prime extensions of those non-prime theories. Both these conditions are met if formulae are evaluated with respect to *sets* of situations, which is exactly parallel to Levesque's treatment of belief, and the truth of an existential $\exists xPx$ at that set of situations requires the truth of Pn_i in all of the situations, for some element n_i of the common domain. A consequence of Patel-Schneider's decidable solution to this problem, and a point of diversion from Levesque, is that the theory of a point of evaluation in a model can no longer be seen as the intersection of its prime extensions: for example, applying his truth definition for the existential to each member of a set of situations (the set S in each of these applications of the rule is of course a singleton set) could result in $\exists xPx$ being true in each situation without it being true at the set of these[14].

So Patel-Schneider extends this definition to a support relation with respect to sets S of *compatible* situations. A compatible set of situations is a set of situations with the same domain and the same mapping of function letters to functions; they differ only in their assignments of truth and falsity. For S such a set, the definition is now changed to

$$S, v \models_T \alpha \text{ iff for all } s \in S,\ s, v \models_T \alpha$$
$$S, v \models_F \alpha \text{ iff for all } s \in S,\ s, v \models_F \alpha$$

where α is quantifier-free. Quantified sentences are interpreted as follows:

$$S, v \models_T \forall x\alpha \text{ iff for all } d \in D,\ S, v(x/d) \models_T \alpha.$$
$$S, v \models_F \forall x\alpha \text{ iff for some } d \in D,\ S, v(x/d) \models_F \alpha.$$
$$S, v \models_T \exists x\alpha \text{ iff for some } d \in D,\ S, v(x/d) \models_T \alpha.$$
$$S, v \models_F \exists x\alpha \text{ iff for all } d \in D,\ S, v(x/d) \models_F \alpha.$$

---

[14]This property is however retained in Fine 1988, a first-order extension of Fine 1974 emphasising generality in modelling relevance logics rather that decidability, where, very roughly, the relevant equivalence of quantification and disjunction or conjunction is avoided by defining it by means of the properties of an arbitrary individual outside the domain of point of evaluation.

Thus for the truth of an existential $\exists x\alpha$ to be supported in a set S of situations there must be some element d of the common domain D such that $\alpha(x/d)$ has true support in every situation in S. Lakemeyer calls this ***global existential quantification***. Although the semantics here is defined only on prenex form formulae, there is a similar one allowing for formulae of arbitrary complexity which possesses a powerful normalisation theorem to return to this state of affairs.

The definition of entailment in the logic of tautological entailment states that $\alpha \longrightarrow \beta$ iff β is true whenever α is and α is false whenever β is. In the logics considered above these two conditions have coincided for the analogous notion in belief logics, but in Patel-Schneider's logic this breaks down, forcing a decision about how the first-order extension of tautological entailment is to be defined. In Patel-Schneider's logic he chooses to define entailment by the first condition, which he calls t-entailment, written $\longrightarrow_t$; the alternatives are the second condition, or f-entailment ($\longrightarrow_f$), and both conditions, or tf-entailment ($\longrightarrow$). These are some features of the different definitions, when taken as relations between sentences of the above logic:

$$\begin{array}{ll}
\forall xPx \longrightarrow Pa & Pa \longrightarrow \exists xPx \\
\forall xPx \longrightarrow_t Pa \wedge Pb & Pa \vee Pb \not\longrightarrow_t \exists xPx \\
\forall xPx \not\longrightarrow_f Pa \wedge Pb & Pa \vee Pb \longrightarrow_f \exists xPx \\
\forall xPx \not\longrightarrow Pa \wedge Pb & Pa \vee Pb \not\longrightarrow \exists xPx
\end{array}$$

Given Patel-Schneider's intentions for the logic f-entailment and t-entailment are preferable to tf-entailment because they are stronger, but t-entailment is preferable to f-entailment because the inferences it allows are more appropriate for knowledge representation. Moreover, the type of reasoning in the invalid inference $Pa \vee Pb \not\longrightarrow_t \exists xPx$ is a form of reasoning from cases of a kind not generally valid in relevance logics, which would, if allowed in the logic, prevent the creation of an entailment algorithm along similar lines to those for propositional first degree entailments. Since both the knowledge base application and the existence of such an algorithm rate high in his priorities, Patel-Schneider opts to use $\longrightarrow_t$.

The decidable algorithm for determining entailment in this logic rests on a Skolemisation theorem. Let $\alpha_{S\exists}$ be α with all existentially quantified variables Skolemised and let $\beta_{S\forall}$ be β with all universals Skolemised. Then for all three versions of entailment, $\alpha \longrightarrow \beta$ iff $\alpha_{S\exists} \longrightarrow \beta_{S\forall}$. The following theorem, which strongly echoes the tautological entailment theorem for the propositional case, then holds:
Let α and β are sentences in Skolemised prenex conjunctive normal form: that is, $\alpha = \forall z_1 \ldots z_k(\alpha_1 \wedge \ldots \wedge \alpha_l)$ and $\beta = \exists x_1 \ldots x_m(\beta_1 \wedge \ldots \wedge \beta_n)$. Then $\alpha \longrightarrow_t \beta$ iff there exists a substitution θ for $x_1, \ldots x_m$, such that for each β_i there exists some α_j and some substitution ψ for $z_1, \ldots z_k$ such that $\alpha_j\psi \subseteq \beta_i\theta$ (where $\alpha_j\psi$ and $\beta_i\theta$ are treated as sets of literals).

Then an algorithm can be found to show that t-entailment is decidable. Let α and β be as above. For each α_j and β_i compute a set of most general substitutions Θ_{ij} such that for $\theta \in \Theta_{ij}$, $\alpha_j\theta \subseteq \beta_i\theta$. For each element of Θ_{ij} define a new substitution by systematically replacing the variables $z_1, \ldots z_k$ by others occurring nowhere else. Let Ψ_{ij} be this new set and let $\Psi_i = \bigcup_{1 \leq j \leq l} \Psi_{ij}$. Then $\alpha \longrightarrow_t \beta$ iff there is some substitution ϕ which is the most general unifier of some $\{\psi_i \mid \psi_i \in \Psi_i\}$. Although the worst case behaviour of this algorithm is exponential in the size of α and β it will always terminate and will be quite fast in normal conditions[15].

[15]See, for example, Apt and van Emden 1982 for a theoretical examination of the semantics of logic programming languages which use such methods.

3.5 Embedding The Logic

The clause for the interpretation of existential quantification, which is essential for the decidability of Patel-Schneider's logic and captures a 'relevant' or constructive mode of reasoning, is retained in a modified form by Lakemeyer[16] in an extension of Levesque's logic to first-order logic which contain operators for explicit and implicit belief, but which does not allow nested beliefs. In this context, however, the treatment of quantification must be modified to bring out the distinction between rigid and non-rigid designators by sometimes allowing that in different members of the belief set - as in Levesque's logic, the non-relationally specified set of situations with respect to which belief formulae are evaluated - different individuals may satisfy a given existential, and, again following Levesque, by taking into account the way things are in the 'actual' world in the interpretation of logical connectives and quantifiers not within the scope of B or L. Patel-Schneider's logic of commitment avoided such considerations, but as soon as a belief operator is explicitly introduced, the logic must be able to handle formulae that do not lie within its scope and also specify the conditions under which they are classically valid: as in Levesque's logic, an actual situation is required for these purposes, while the conditions enabling global existential quantification need be satisfied only when the existential is within the scope of an explicit belief operator. For example, $Pa \vee Pb \supset \exists x Px$ should turn out to be valid since the existential quantifier is not within the scope of a modal operator. The language L consists of a set of variables V, a set of explicitly represented parameters or rigid designators N, and function and predicate symbols, as well as the modal operators B and L. A term is either a variable, a rigid designator or a function symbol whose arguments are themselves terms. A closed term is a

[16]Lakemeyer 1986.

term not containing any variables, and the set of closed terms is denoted T. The set N is also used as the fixed universe of discourse in all models. An atomic formula is a predicate symbol with terms as arguments. Given the atomic formulae, the logical connectives $\neg$, $\vee$ and the existential quantifier $\exists$ as well as the modal operator symbols B and L, we can generate all the well-formed formulae of L by using the standard formation rules and equivalences.

The next three definitions introduce the key idea of the logic, to be exploited in the definition of explicit and implicit belief, which is an intermediate stage in valuations and in the realisation of terms in a model. They can be regarded as meeting in a novel fashion the adequacy conditions for a quantified intensional logic, concerned with what should determines the range of the quantifiers: namely that each situation should have a domain of individuals over which the quantifiers are to range, and that one and the same individual can be identified across situations. Distinctive features of Lakemeyer's solution are that, unlike most other such logics, this is achieved without admitting identity to the syntax of the language, thus sidestepping the other issues that this would have introduced; and that, as will become apparent, because of the use of closed terms in their definition the quantifiers cannot be categorised absolutely as either objectual or conceptual, as is usually possible with such logics. The logic is also committed to the view that there are no situations in which distinct parameters refer to the same thing, but that different situations may have different domains of actual individuals; on the other hand, every term of the language is defined in every situation. Let $\Pi = \{v \mid v : V \rightarrow T\}$ be the set of variable maps; for $n \in N$ $v(x/n)$ is the variable map identical to v except that x gets mapped to n, and note that for convenience it is assumed that, as in Patel-Schneider's logic, the domain of a variable map is extended to include terms and sequences of terms in the obvious way, assigning values to the variables in them. The individuals that serve as the values of bound variables are terms rather than parameters, and so in all contexts the terms are crucial to the logic of the quantifiers: further work is required to ensure that the domain of extensional quantifiers is objectual. Also let $\Xi = \{\xi \mid \xi : \Pi \times V \rightarrow T\}$ be the set of choice functions, which are to take care of the dependencies on the leading quantifiers in belief sentences, given the interpretation of quantification below. In formulae of the form $B\exists x\alpha$ an individual global to the entire belief set is picked with respect to which the rest of the formula is evaluated. Closed terms which are not rigid designators can be picked, so the individual of which some property is believed to hold need not in fact refer to the same object in all situations of the belief set, and so the belief set need not be compatible in the sense of Patel-Schneider, since this requirement only makes sense where de re beliefs cannot be expressed. But since the domain of quantification is to be the set of terms, and all terms have some meaning in every situation, Patel-Schneider's prerequisite for global existential quantification is in fact met. So since terms can be seen as denoting locally existing individuals, there is a cognitive means of identification of these individuals across situations which does not require that they are actually the same individual. To ensure both that this is possible and that the interpretation of $\exists xB\alpha$ is intuitively correct, an aspect of a situation concerning the reference of terms is introduced, with respect to which members of a belief set may differ. Whereas in Patel-Schneider's logic the elements of the belief set differed only in the interpretation of predicates, here the meaning of terms is also defined to be situation relative. So $\equiv_s$ is introduced as a coreference relation mapping every closed term into a rigid designator, its meaning with respect to the situation s. The following are properties of $\equiv$: it is an equivalence relation; no two rigid designators corefer, and so actual individuals cannot split or merge across alternatives; every closed term has a coreferring rigid designator - there are no non-referring terms - and the relation is preserved by uniform substitutions in a formula. Note that 0-place function symbols, which are in the range of choice functions, are interpreted as non-rigid designators according to the above definitions. Now it can be seen how this can be used in conjunction with the notion of a choice function so that $\exists xB\alpha$ is interpreted intuitively. In sentences of the form $\exists xB\alpha$ the function chooses a term during the interpretation of B, that is while evaluating with respect to the actual situation; it is then required in the truth clause for B not that the the term but that the *meaning* of that term in the actual situation has the stated property in every member of the belief set, and so the reference of that term is global to the entire set, and in each member of the set the property holds of that same object. v_s is the variable map such that $v_s(x) = n$, where $n \in N$ and $n \equiv_s v(x)$. Where T^s and F^s are the mappings associated with s by which the predicate letters are interpreted, a situation s is a triple $< T^s, F^s, \equiv_s >$, and the collection of all situations is S.

Thus Patel-Schneider's logic differs technically from this one in a number of ways. For example, here there is the ability to convert the values of variables into rigid designators in the context of explicit or implicit belief; sentences are evaluated with respect to a belief set consisting of *arbitrary* first-order situations and distinguished 'actual' situation; and variables may be interpreted as closed terms rather than rigid designators. Before the truth definitions, he defines, following Levesque, the set $W(s)$ of possible worlds compatible with a situation s. $W(s)$ is the set of situations $s' \in S$ such that for every k-tuple $< n_1, \ldots n_k > \in N^k$ and every k-ary predicate P,

1. $< n_1, \ldots n_k >$ is in exactly one of $\{T^{s'}(P), F^{s'}(P)\}$,

2. if $< n_1 \ldots n_k > \in T^s(P)$ then $< n_1 \ldots n_k > \in T^{s'}(P)$,

3. if $< n_1 \ldots n_k > \in F^s(P)$ then $< n_1 \ldots n_k > \in F^{s'}(P)$ and

4. $\equiv_{s'} = \equiv_s$.

$W(S)$, the set of possible worlds, is the union of all $W(s)$ for $s \in S$. Then where $U \subseteq S$ is arbitrary, $s \in S$, and α' is as α except with the obvious precautionary renaming of variables, the support clauses are defined as follows:

$U, s, v \models_T \alpha$ iff $\exists \xi \in \Xi$ such that $U, \xi, s, v \models_T \alpha'$

$U, s, v \models_F \alpha$ iff $\exists \xi \in \Xi$ such that $U, \xi, s, v \models_F \alpha'$

$U, \xi, s, v \models_T P(t_1, \ldots t_n)$ iff $v(t_1, \ldots t_k) \in T^s(P)$

$U, \xi, s, v \models_F P(t_1, \ldots t_n)$ iff $v(t_1, \ldots t_k) \in F^s(P)$

$U, \xi, s, v \models_T \neg\alpha$ iff $U, \xi, s, v \models_F \alpha$

$U, \xi, s, v \models_F \neg\alpha$ iff $U, \xi, s, v \models_T \alpha$

$U, \xi, s, v \models_T \alpha \vee \beta$ iff $U, \xi, s, v \models_T \alpha$ or $U, \xi, s, v \models_T \beta$

$U, \xi, s, v \models_F \alpha \vee \beta$ iff $U, \xi, s, v \models_F \alpha$ and $U, \xi, s, v \models_F \beta$

$U, \xi, s, v \models_T \exists x\alpha$ iff $U, \xi, s, v(x/\xi(v, x)) \models_T \alpha$

$U, \xi, s, v \models_F \exists x\alpha$ iff $\forall t \in T$ $U, \xi, s, v(x/t)) \models_F \alpha$

$U, \xi, s, v \models_T B\alpha$ iff $\exists \xi' \in \Xi$ $\forall s' \in S$ $U, \xi', s', v_s \models_T \alpha$

$U, \xi, s, v \models_F B\alpha$ iff $U, \xi, s, v \not\models_T B\alpha$

$U, \xi, s, v \models_T L\alpha$ iff $\forall s' \in W(S)$ $\exists \xi' \in \Xi$ $U, \xi', s', v_s \models_T \alpha$

$U, \xi, s, v \models_F L\alpha$ iff $U, \xi, s, v \not\models_T L\alpha$

A formula α is then said to be valid ($\models \alpha$) iff for all $U \subseteq S$, all $s \in S$ and all $v \in \Pi$, $U, s, v \models_T \alpha$.

If the language is restricted to its quantifier-free sentences, and as a consequence variable maps and choice functions become redundant, and $\equiv_s$ remains unexploited, then what we have is Levesque's logic of explicit and implicit belief; the set U in the truth definition is of course the set B of a Levesque model. But the features of the contained propositional fragment carry over to the full first-order case, for the following are some theorems of the logic: implicit belief is closed under modus ponens, all valid first-order formulae are implicitly believed, and every explicit belief is also implicitly believed. Also, where $\alpha^{x/a}$ ($\alpha^{(x/n)}$) is the formula obtained from α by replacing all occurrences of the free variable x

by the closed term a (by the rigid designator n),

$$\models \exists x B\alpha \supset B\exists x\alpha \text{ and } \models \exists x L\alpha \supset L\exists x\alpha;$$
$$\models B\alpha^{x/n} \supset \exists x B\alpha \text{ and } \models L\alpha^{x/n} \supset \exists x L\alpha;$$
$$\models B\alpha^{x/a} \supset B\exists x\alpha \text{ and } \models L\alpha^{x/a} \supset L\exists x\alpha;$$
$$\models B\forall x\alpha \supset \forall x B\alpha \text{ and } \models L\forall x\alpha \supset \forall x L\alpha;$$
$$\models \forall x L\alpha \supset L\forall x\alpha.$$

It can also be shown that there are formulae α and non-rigid terms a such that $\not\models B\alpha^{x/a} \supset L\exists x\alpha$, that $\not\models B\exists x\alpha \supset \exists x B\alpha$ and that $\not\models L\exists x\alpha \supset \exists x L\alpha$; that explicit belief is not closed under modus ponens, and that there is no universal generalisation for explicit belief ($\not\models \forall x B\alpha \supset B\forall x\alpha$). There is also a decidability result for the logic of commitment - that is for entailments of the form $B\alpha \supset B\beta$ - very similar, not surprisingly, to that of Patel-Schneider, where α and β are in normal form.

3.6 Classical Logics

Fagin and Halpern[17] propose a different logic based on Levesque's distinction between explicit and implicit belief, and which also accommodates multiple agents and embedding of beliefs, which they call a logic of general awareness. As in Levesque's logic, implicit beliefs are represented within a possible worlds semantics and these are all the logical consequences of an agent's explicit beliefs. But in order to capture resource-bounded reasoning, which Levesque's logic did not do in its semantic treatment of explicit belief, they first introduce an awareness operator into a standard Kripke structure which is syntactic in nature. $A_i(s)$ is the set of sentences of which agent i is aware at world s, and an awareness operator is defined as holding precisely for members of this set. So while implicit

[17]Fagin and Halpern 1988.

belief in a proposition at a world is defined as the truth of that proposition at all accessible worlds, explicit belief is defined by this semantics restricted to those sentences permitted by awareness at the world of evaluation. Agents are thus in a sense perfect reasoners, logically omniscient restricted to an arbitrary syntactic class of sentences. But agents do not explicitly believe all valid formulas: it will turn out that $\neg B_i(p_j \vee \neg p_j)$ is satisfiable since perhaps $(p_j \vee \neg p_j) \notin A_i(s)$. And $B_i p_j \wedge B_i(p_j \supset p_k) \wedge \neg B_i p_k$ is satisfiable since, although aware of $p_j \supset p_k$, agent i may not be aware of p_k - there is no necessary restriction on the syntactic filter that it be closed under subformulas. This introduction of a syntactic element to the model does, however, seem to give rise to a too fine-grained notion of belief: $B(\alpha \wedge \beta)$ is not equivalent to $B(\beta \wedge \alpha)$ - it is possible that only one of these sentences may be in the awareness set - and so the order in which the conjuncts appear is made semantically significant. This inclusion within the model structure of explicit sets of sentences may, like previous similar attempts[18] avoid the essentially semantic problem of logical omniscience, but such an injection of syntax makes the arbitrary seem significant. There do appear to be a large number of restrictions which may be put on the awareness set: we may stipulate that $\alpha \wedge \beta \in A_i(s)$ iff $\beta \wedge \alpha \in A_i(s)$; or $\alpha \in A_i(s)$ iff $\neg\alpha \in A_i(s)$; or if $\alpha \in A_i(s)$ then $A_i\alpha \in A_i(s)$, where the A_i in $A_i\alpha$ is the operator defined by the predicate A_i. But such ad hoc restrictions leave no essential rôle to possible world semantics in analysing belief, for no explanatory condition on the structure of the accessibility relation is associated with these moves - the semantics of belief is given by ad hoc restrictions on the syntax which neither are motivated by the semantics nor correspond formally to it. The attempt to marry semantic and syntactic approaches to belief would appear in this case to have resulted in a sentential semantics[19]. However, for the two important versions of awareness

---

[18]e.g. Konolige 1983 and Moore and Hendrix 1979.

[19]See Konolige 1986 for arguments to this effect.

now examined, the logic of commitment is able to provide an endorsement of the logics.

Fagin and Halpern's propositional "logic of awareness" dispenses with partial and incoherent situations in order better to accommodate nesting of explicit and implicit beliefs. Multiple agents are also allowed into the logic, and they also make use of awareness operators in the definition of explicit belief. These will allow some of the effects of partial situations into the logic, but come into play only in the context of explicit belief. So in contrast to the previous logics examined which did precisely the reverse, the classical structure of a possible world is taken as the fundamental tool, relative to which different partial situations and so explicit belief may be defined. Models are Kripke structures with a set of classical worlds, a truth assignment to the finite set of atomic formulae for each world, a serial, transitive and Euclidean relations for each agent as well as a function A_i for each agent i which associates with each world the set of atomic formulae of which i is aware at that world. As before, there is a special propositional constant **true**, and a model is a structure $M = (S, \pi, A_1, \ldots A_n, R_1, \ldots R_n)$; S is a set of worlds, π a truth assignment to the values $\{0, 1\}$ for each word, R_i a serial, transitive and Euclidean relation on S, and A_i a function associating with each world a set of atomic formulae, those of which i is aware at that world. Then the support relations and the truth definition are as follows: where Ψ is a subset of Θ, the set of all atomic formulae

$$M,s \models_T^\Psi \mathbf{true}$$
$$M,s \not\models_F^\Psi \mathbf{true}$$
$$M,s \models_T \mathbf{true}$$
$$M,s \models_T^\Psi p_i \text{ iff } \pi(s,p_i) = 1 \text{ and } p_i \in \Psi$$
$$M,s \models_F^\Psi p_i \text{ iff } \pi(s,p_i) = 0 \text{ and } p_i \in \Psi$$
$$M,s \models p_i \text{ iff } \pi(s,p_i) = 1$$
$$M,s \models_T^\Psi \neg\alpha \text{ iff } M,s \models_F^\Psi \alpha$$
$$M,s \models_F^\Psi \neg\alpha \text{ iff } M,s \models_T^\Psi \alpha$$
$$M,s \models \neg\alpha \text{ iff } M,s \not\models \alpha$$
$$M,s \models_T^\Psi \alpha \vee \beta \text{ iff } M,s \models_T^\Psi \alpha \text{ or } M,s \models_T^\Psi \beta$$
$$M,s \models_F^\Psi \alpha \vee \beta \text{ iff } M,s \models_F^\Psi \alpha \text{ and } M,s \models_F^\Psi \beta$$
$$M,s \models \alpha \vee \beta \text{ iff } M,s \models \alpha \text{ or } M,s \models \beta$$
$$M,s \models_T^\Psi B_i\alpha \text{ iff } M,t \models_T^{\Psi \cap A_i(s)} \alpha \text{ for all } t,\ sR_it$$
$$M,s \models_F^\Psi B_i\alpha \text{ iff } M,t \models_F^{\Psi \cap A_i(s)} \alpha \text{ for some } t,\ sR_it$$
$$M,s \models B_i\alpha \text{ iff } M,s \models_T^\Theta B_i\alpha$$
$$M,s \models_T^\Psi L_i\alpha \text{ iff } M,t \models_T^\Psi \alpha \text{ for all } t,\ sR_it$$
$$M,s \models_F^\Psi L_i\alpha \text{ iff } M,t \models_F^\Psi \alpha \text{ for some } t,\ sR_it$$
$$M,s \models B_i\alpha \text{ iff } M,s \models_T^\Theta B_i\alpha$$

The definition of validity is then standard. Implicit belief satisfies the axioms of weak $S5$, and all valid formulae and all the logical consequences of one's beliefs are believed. As in Levesque's logic, $\neg B_i(\alpha \vee \neg\alpha)$ is satisfiable, and $\alpha \supset \beta$ does not entail $B_i\alpha \supset B_i\beta$; also, if α is a classical theorem, then $\models A_i\alpha \supset B_i\alpha$, where $A_i\alpha$ is an abbreviation for the conjunction of $B_i(p_j \vee \neg p_j)$ for all atomic formulae p_j that appear in α. Fagin and Halpern prove a theorem to the effect that any sentence α containing operators B_i is equivalent to another sentence α^* in which B_i occurs only in the context $B_i(p_j \vee \neg p_j)$. In effect then, explicit belief is definable in terms of implicit belief and awareness Also $B_i\alpha \supset B_i(\alpha \vee \beta)$ is valid - it does not matter whether $\beta \in A_i(s)$, for any s. This suggests a different intuitive reading for awareness from their logic presented later, where this is not valid. Unlike Levesque, however, inconsistent beliefs are not allowed and an agent's set of beliefs are closed under implication. So logical omniscience in this sense fails only because of lack of awareness of some proposition on the part of the agent.

Also valid in this logic is the axiom $B_iL_i\alpha \equiv B_i\alpha$, so i explicitly believes that he implicitly believes α exactly if he explicitly believes it. This suggests that $L_i\alpha$ should perhaps be read as "α is a (classical) logical consequence of what i believes" rather than an as external characterisation of i's beliefs. A notable feature of the logic is the fact that though $\neg A_i\alpha$ (i.e. $\neg B_i(\alpha \vee \neg\alpha)$) is of course satisfiable, sentences of this form cannot be the object of some agent's belief, since $\neg B_iB_j(\alpha \vee \neg\alpha)$ is a theorem of the logic. Lack of awareness in an agent cannot be recognised by others or by himself. Other features of the logic involving embedded operators are not immediately apparent: the only complete axiomatisation that Fagin and Halpern have found is the axioms of weak $S5$ together with $\alpha^* \equiv \alpha$, where $*$ is as in the theorem mentioned above, and this is not highly informative with regard to the relations between the operators.

Fagin and Halpern tell us that beliefs are closed under classical implication, but this is only so in the sense that if $B\alpha$ and $B(\alpha \supset \beta)$ then $B\beta$, where the awareness of the conclusion is implicit in the premises. It would be interesting to discover the logic of commitment for explicit belief in this logic, the consequence relation under which beliefs are closed, and perhaps also a workable deductive system to match this logic. This would tell us when $B\beta$ could be inferred from $B\alpha$ without the use of $B(\alpha \supset \beta)$, and so would have a similar status to that of first degree entailment in Levesque's logic. From here on we drop the propositional constant from the language. Much of the interest comes from the way in which this illustrates the concept of awareness by pushing it into the background logic, and also allowing us to see a logic of awareness and explicit belief as arising in a sense from a normal modal logic of implicit belief simply by changing the underlying propositional logic while retaining the definition of modality. If formulae $< \alpha, \beta >$ were in this relation then for all models M and all situations s, if $M, s \models_T^{\Psi} B_i\alpha$ then $M, s \models_T^{\Psi} B_i\beta$, so requiring that the logic of commitment treat the partial states of Fagin and Halpern's logic rather than the complete possible worlds they use to define validities. There are some features of Fagin and Halpern's logic that dictate aspects of the consequence relation that is to be defined: for example, there can be no validities, since there are no formulae α such that $B\alpha$ is true in all models, and the role of awareness in the definition of belief requires that the propositional letters of the conclusion of a logical consequence are in some systematic way dependent on those of the premises. For simplicity the logic is restricted to depth one beliefs and multiple agents as well as the modal operator L are ignored. Although much of what makes the logic interesting is subsequently passed over, its basic concepts should be better illustrated in this way.

What is required is a relation which tells us what follows from a given belief, relative to an arbitrary model and partial situation. Since belief in a formula is defined as its having true support in each member of the set of accessible situations, evaluated relative to a set of propositional letters of which the agent is aware, the problem can be reduced to discovering a consequence relation for sentences holding true at situations, which are partial with respect to this awareness set. For if we have a relation $\models_K$ such that $\alpha \models_K \beta$ iff for all M, s, Ψ, if $M, s \models_T^{\Psi} \alpha$ then $M, s \models_T^{\Psi} \beta$, then belief is closed under this relation. If $\alpha \models_K \beta$ and $M, s \models_T^{\Theta} B_i\alpha$ then $M, t \models_T^{\Theta \cap A_i(s)} \alpha$ for all t such that $(s, t) \in R_i$, and so $M, t \models_T^{\Theta \cap A_i(s)} \beta$ for all t such that $(s, t) \in R_i$, that is $M, s \models_T^{\Theta} B_i\beta$. This also gives the biggest relation with this property, since it makes the minimal requirement of preservation of truth at situations.

The relation $\models_K$ is, trivially, a consequence relation for sentences holding true at situations in Fagin and Halpern's logic, but this is hardly illuminating. It is however not difficult to see which logic it resembles. Since we are here concerned with partial and consistent situations the relations $\models_T^\Psi$ and $\models_F^\Psi$ in the truth definition can also be seen as partial truth assignments, where Ψ is the set of sentence letters on which these are defined, and total refinements - that is, completely specified extensions - of which are classical assignments, corresponding to the classical relation $\models$ in the truth clauses. From the perspective which saw meta-beliefs as important, it was much easier to take classical assignments as basic and define belief in terms of partial sub-assignments of these, rather than to do things the other way round as in the logics previously examined. But to illustrate the restriction of their logic under consideration here which disregards truth and embedded and implicit beliefs, it is best to revert to regarding partial situations as the fundamental tool of the logic. Once it is recognised that $\models_T^\Psi$ and $\models_F^\Psi$ together form a consistent but partial truth assignment to the formulae of the language, it is readily seen that according to Fagin and Halpern's truth definitions for the connectives this truth assignment is a partial valuation from the propositional language into $\{1,0\}$, the interpretation of the connectives being the partial n-ary functions from $\{1,0\}^n$ to $\{1,0\}$ given in the matrix for Kleene's strong three-valued logic. Since it is the logic of the *positive* commitment for belief, and so the relation $\models_T$, that is at issue, the consequence relation under which belief is closed is the rather unrestrictive one which requires only preservation of truth from premises to conclusion: t-entailment, in Patel-Schneider's terminology. Unlike classical logic, if it is also required that some premise must be false whenever the conclusion is, then this would define a different consequence relation. So if $\models_K$ is defined as, for Γ a set of sentences, $\Gamma \models_K \beta$ iff for all M, s, Ψ, if $M, s \models_T \alpha$ for all $\alpha \in \Gamma$ then $M, s \models_T \beta$, then $\Gamma \models_K \beta$ iff every partial assignment from the propositional letters into $\{1,0\}$ - extended to complex formulae according to Kleene's strong three-valued system - which assigns 1 to each $\alpha \in \Gamma$, also assigns 1 to β. There are no tautologies: dually, no sentences β such that $\emptyset \models_K \beta$. For example we have $\emptyset \not\models_K \alpha \vee \neg\alpha$. This corresponds to the fact that in Fagin and Halpern's logic no formulae are always believed. In their logic, however, beliefs are closed under classical implication in the sense explained above: this can be seen as correct because $\models_K$ respects disjunctive syllogism and so β is a logical consequence of α and $\alpha \supset \beta$ according to $\models_K$. It is also easy to check that for all α and β, $\alpha \wedge \neg\alpha \models_K \beta$, since all assignments are consistent; this explains why contradictory sets of beliefs are not possible.

There is a natural deduction system S such that β is a sentence deducible from the set of sentences Γ by means of the rules of S iff $\Gamma \models_K \beta$. Some care is needed in setting up S, given that logical consequence is defined with respect only to preservation of truth from antecedent to consequent, and not preservation of falsity from consequent to antecedent. So reductio ad absurdam cannot form part of the system, for if a contradiction is derivable from α it need not be false. Also, since $\alpha \models_K \alpha$ but not $\models_K \alpha \supset \alpha$, the fact that β is derivable from α should not allow us to deduce that $\alpha \supset \beta$ is true. As has already been argued, this is a desirable feature in a logic of belief. The rules for S, which is the propositional fragment of a system of Kearns[20], are as follows:

$$
\begin{array}{ll}
\alpha/\alpha \vee \beta & \beta/\alpha \vee \beta \\
\alpha \vee \beta, \gamma^{[\alpha]}, \gamma^{[\beta]}/\gamma & \alpha, \neg\alpha/\beta \\
\alpha/\neg\neg\alpha & \neg\neg\alpha/\alpha \\
\neg(\alpha \vee \beta)/\neg\alpha & \neg(\alpha \vee \beta)/\neg\beta \\
\neg\alpha, \neg\beta/\neg(\alpha \vee \beta) &
\end{array}
$$

$\gamma^{[\alpha]}$ means that the bracketed hypothesis β from which γ was derived has been cancelled. $\alpha_1, \ldots \alpha_n/\beta$ is a theorem of S if β is the conclusion of a proof tree whose uncancelled assumptions are among $\alpha_1, \ldots \alpha_n$. This may be written as $\alpha_1, \ldots \alpha_n \vdash_S \beta$. The proof of the strong completeness of $\models_K$ with respect to this system may be found in Kearns 1979; it uses the standard technique due to Lindenbaum.

---

[20]Kearns 1979.

S can now be seen as the deductive system that gives the positive logic of commitment for explicit belief for this fragment of Fagin and Halpern's logic which does not allow embedding of operators. With this restriction in place we have at hand a concise means of expressing the difference between the logics of explicit and implicit belief. Consider the characterisation of minimal normal modal logics as those closed under the rule K:

$$\text{from } \alpha_1, \dots \alpha_n \vdash \beta \text{ infer } B\alpha_1, \dots B\alpha_n \vdash B\beta,$$

where $n \geq 0$. Then where $\vdash$ is consequence in classical propositional logic, the operator B defines the logic of commitment for Fagin and Halpern's implicit belief operator and completely characterises their system without embeddings permitted; where $\vdash$ is read as our chosen notion of consequence for Kleene's three-valued logic, the same is done for the notion of explicit belief. This perspective on awareness as a change in the propositional logic underlying a unitary modal definition might break down if applied in the same fashion as an alternative description to the multiple modalities of the rest of Fagin and Halpern's logic, but nevertheless it appears to offer an interesting generalisation of their awareness operator, and to suggest a very natural progression to a first-order language, where awareness of objects would seem to have a straightforward formal counterpart in Kleene's semantics for predicate logic.

The main feature of their second general proposal, called the logic of general awareness, is a syntactic awareness function for each agent which assigns to it at each world an arbitrary set of formulae which need not be primitive. No structure at all is assumed for this set. Then where M is as before except for this change to A_i, the truth definition is:

$M, s \models \mathbf{true}$

$M, s \models p_i$ iff $\pi(s, p) = 1$

$M, s \models \neg\alpha$ iff $M, s \not\models \alpha$

$M, s \models \alpha \vee \beta$ iff $M, s \models \alpha$ or $M, s \models \beta$

$M, s \models A_i\alpha$ iff $\alpha \in A_i(s)$

$M, s \models L_i\alpha$ iff $M, t \models \alpha$ for all t such that $(s, t) \in R_i$

$M, s \models B_i\alpha$ iff $\alpha \in A_i(s)$ and $M, t \models \alpha$ for all t such that $(s, t) \in R_i$

L_i is again the classical belief operator, and explicit beliefs are the restriction of implicit beliefs those that are also in the awareness set, and so given the appropriate relational conditions on frames, the logic may be axiomatised by the axioms of weak $S5$ together with $B_i\alpha \equiv L_i\alpha \wedge A_i\alpha$. Any interest in the logic must come in the restriction placed on the awareness function, and Fagin and Halpern suggest several. Order of presentation of conjuncts in a conjunction may not matter, so an axiom could be added to this effect; other examples include closing off under subformulae, and restricting the set to any sentences generated from a set of atomic formulae. If awareness is closed under subformulae, however, then explicit belief is closed under implication. Agents themselves may be put into or left out of awareness sets with the result that some others may not be aware of any sentences that mentioned them, and to ensure that agents know precisely which formulae they are aware of we require that if $(s, t) \in R_i$ then $A_i(s) = A_i(t)$. In this case, $B_i\alpha \wedge A_iB_i\alpha \supset B_iB_i\alpha$, so positive introspection holds for explicit belief if an agent is aware of the belief in question.

We conclude this chapter by looking at one of the more systematic proposals made in this context, a logic of 'general awareness' where explicit belief is the restriction of implicit belief to the upward closure of a set of atomic sentences, those of which the agent is 'aware' at a world. Although the authors note that its theorems closely resemble those of Levesque's logic of explicit and implicit belief and despite the fact that it is one of the more plausible of their proposals from the point of view of human beliefs, it compares unfavourably with it because of its syntactic flavour. This is not irreparable, for there already exist as standard techniques in logics of analytic implication semantic models for dealing with issues similar to those that arise when the notion of awareness is introduced into a logic of explicit belief[21]. In fact, the logic of commitment induced here by this type of awareness belongs to an existing class of relevance logics which differ from Levesque's in a number of ways: in particular, lack of awareness apart, agents are logically omniscient in the classical sense, and of course agents do not consider possible inconsistent situations. To emphasise this alternative perspective, we now give the alternative models as well as the axiomatisation for which they can be checked to be complete. The point is that the models standardly given for the logic under which belief is closed can be used to illuminate the models for this logic of awareness.

A model is a tuple $M = (W, I, \cup, \gamma, \pi, R_1, \ldots R_n, A_1, \ldots A_n)$ such that where P is the set of atomic formulae of the language the language L, W is a non-empty set; for each $w \in W$, I_w is a non-empty set; for each $w \in W$, $\cup_w$ is an associative, commutative and idempotent operation on I_w; for each $w \in W$, γ_w is a function from P into I_w; π and R_i are as above; and $A_{i,w} \subseteq P$ for each i and for each $w \in W$. It is also a condition that whenever $(w, v) \in R_i$ then $(I_w, \cup_w, \gamma_w)$ extends - can be embedded in - $(I_v, \cup_v, \gamma_v)$, and $A_{i,v} \subseteq A_{i,w}$.

For each semilattice $(I_w, \cup_w, \gamma_w)$ with $a, b \in I_w$, $\leq_w$ is defined by $a \leq_w b$ iff $a \cup_w b = b$. γ_w may then be extended to all formulae such that where $\mathcal{L}(\alpha)$, the set of sentence letters occurring in α, is $\{p_1, \ldots p_n\}$, $\gamma_w(\alpha) = \gamma_w(p_1) \cup_w \ldots \cup_w \gamma_w(p_n)$. Then relative to a model $M \models$ is defined as above, except that

$$w \models A_i\alpha \text{ iff } \gamma_w(\alpha) \leq_w \gamma_w(\bigwedge A_{i,w}).$$

---

[21]Dunn 1972, Urquhart 1973, Deutsch 1984 and Fine 1986, for example, use models very similar to the one below.

The axioms of the logic are all classical propositional tautologies; for the L-operator the axioms of the modal logic weak S5, and the following axioms concerning B. $B_i\alpha \equiv B_i\alpha'$, where α' is got from α by the associativity, commutativity and distribution rules for the connectives $\wedge$ and $\vee$; and

$$\begin{array}{ll} B_i\neg(\alpha \wedge \beta) \equiv B_i(\neg\alpha \vee \neg\beta) & B_i\neg(\alpha \vee \beta) \equiv B_i(\neg\alpha \wedge \neg\beta) \\ B_i\neg\neg\alpha \equiv B_i\alpha & B_i\alpha \wedge B_i\beta \equiv B_i(\alpha \wedge \beta) \\ B_i(\alpha \vee (\beta \wedge \neg\beta)) \supset B_i\alpha & \end{array}$$

Essentially, these axioms simply state that the relevance logic of analytical implication is the logic of commitment. For example, the last axiom above ensures that belief is closed under modus ponens, and also if α is a propositional tautology, then $A_i\alpha \supset B_i\alpha$. From this it can be seen that this logic provides another example of an epistemic logic best seen as representing the consequence relation under which belief is closed.

Chapter 4

Frames And Algebras

4.1 Frames

Previously we saw two different approaches to weakening epistemic logic in order to avoid attributing to agents the ability to reason perfectly in classical logic: the descriptive approaches of chapter 2, which permitted the attribution of beliefs to agents without the simultaneous attribution of reasoning ability; and the normative approaches argued for in chapter 3, which maintained but weakened the reasoning ability of agents. For these logics we saw the value of viewing their logics of commitment as a consequence relation, whether or not this was intended in the construction of the logic. Adopting this perspective we shall now develop four-valued modal logic, based on the well-known four-valued logic used by Levesque and Lakemeyer, but shall take a more standard approach towards the subject - in particular with regard to the definition of validity, and by making use of the notion of *frame*.

The relational constraints in Lakemeyer's models depend on the presence of valuations, because these are required to identify the classical worlds of the models, which in turn are used to define the relational constraints. Similarly, validity is defined by means of classical worlds, and there is no useful concept of a frame - a model without valuations. Here we choose to revert to the more standard picture in which the frame is a fundamental notion and where *all* worlds are used in defining validity; also, the operator for implicit belief - which was defined by means of classical worlds - is dropped. However, we add to frames a relation $*$ which allows a valuation-free definition of a classical world and so, if wished, a means of recovering the restrictive definition of validity and the operator for implicit belief. At the end of this section this definition of validity will be introduced temporarily to compare Lakemeyer's models with our own.

For the remainder of the thesis we work out the semantics for the modal logic determined by adding the rule K to four-valued logic, noting the similarities and differences with the classical case. In other words, we consider the most general basic four-valued modal logic $\vdash_K$ and all its extensions: the value of some of these extensions has been noted, but a picture of the more general semantics for modelling both these and classical modal logic has not yet been drawn. This is what we now do. First are introduced the relational semantics for our four-valued modal logic, which will be based on the usual notion of Kripke frames; as has been seen, these are well suited to modelling normative conceptions of belief. The intuitive idea on which these semantic structures are based is that we are given a set of possible worlds between which certain relations of accessibility hold. Given a fixed world x, the worlds accessible from x are those "considered epistemically possible" at x; and a proposition is believed at x if it is true in all worlds considered possible at x.

The classical definition of Kripke frames, and of models based on these, is designed to have the following effect on this intuitive picture. Possible worlds are classical worlds: the formulae true at a world are closed under classical entailment, with the result that for any formula α of the language, exactly one member of the set $\{\alpha, \neg\alpha\}$ is true there. This applies equally to modal formulae, allowing the belief in α at a world to be defined as the truth of α in all accessible worlds, and disbelieving α to be defined simply as the failure to believe α.

To adapt the same intuitive idea to a semantic structure for four-valued modal logic, where the truth of the negation of a formula is not defined as the failure of that formula to be true, changes are required in the classical definitions. The formulae true at a world are to be closed under four-valued entailment - they form a theory in this weaker logic. The change to the classical picture is that we can no longer insist that $\alpha \vee \neg\alpha$ is true at each world and no longer insist that $\alpha \wedge \neg\alpha$ fails to be true at each world. Letting α be a modal formula, this change forces a dissociation between the definitions of disbelief and failure to believe, and so an additional accessibility relation is required, by means of which beliefs are falsified. This will allow propositions to be both believed and disbelieved, and further, the epistemic agent is able to contemplate the failure on his part to have one of these attitudes towards a proposition. For example, we can allow the agent truthfully to profess his agnosticism on a given matter and to have inconsistent beliefs, but we force him to believe that he either has a given belief or he fails to have it.

These considerations give rise to the following formal definition of a frame for four-valued modal logic; the exact relation of these frames to four-valued logic will be given at the end of this section.

Definition 1 *A frame is a structure* $\mathbf{C} = < X, *, R >$, *where* X *is a non-empty set,* R *is an arbitrary binary relation over* X, *and* $*$ *is a functional and symmetric binary relation on* X. □

X may be regarded as a set of worlds, and R as the relation used to verify what is believed. We write the unique y such that $x * y$ as x^*; observe that $x^{**} = x$. Intuitively, whatever is true at x fails to be false at x^*, and so conversely if $\neg\alpha$ is true at x^* then its negation should fail to be false at x^{**}. In other words, since $x^{**} = x$ and $\neg\neg\alpha = \alpha$, α is true at x if and only if $\neg\alpha$ is not true at x^*. If $x = x^*$, it is natural to call x a classical world, given this intended role of $*$.

Finally we define R' over X by $xR'y$ iff x^*Ry. This relation will be used to define disbelief in a formula: it gives the worlds used to check whether a belief is false. In a classical world it can be seen that these are precisely the same worlds that are used to verify belief, which is what is wanted to maintain bivalence there, but in general this is not the case. The classical worlds will not here be given the privileged rôle they were assigned in chapter 3, but this could of course be introduced at will, given that they are defined.

Given a denumerable set $\{p_i \mid i \in I\}$ of propositional variables, the formulae of the modal language are defined in the following way.

Definition 2 *The formulae of the language are:*

1. *Propositional variables are formulae;*
2. *If α and β are formulae then so are $\alpha \vee \beta$, $\alpha \wedge \beta$, $\neg\alpha$, $\Box\alpha$ and $\bigcirc\alpha$;*
3. *Nothing else is a formula.* $\Box$

The natural interpretation of the modal operators $\Box$ and $\bigcirc$ are:

$$\begin{aligned} \Box\,\alpha &: \text{ believes } \alpha; \\ \neg\Box\,\alpha &: \text{ disbelieves } \alpha; \\ \neg\bigcirc\alpha &: \text{ fails to believe } \alpha; \\ \bigcirc\alpha &: \text{ fails to disbelieve } \alpha. \end{aligned}$$

This choice of interpretation for the modal operators is arbitrary in the sense that in the most general logic their technical behaviour is identical - $\Box$ and $\bigcirc$ may be interchanged there in there natural interpretation. Given a frame $\mathbf{C} =< X, *, R >$ and $Z \subseteq X$ let $-Z$ denote the complement of Z and Z^* denote $\{x^* \mid x \in Z\}$.

Definition 3 *A valuation v on the frame $\mathbf{C} =< X, *, R >$ is a mapping from the propositional variables $\{p_i \mid i \in I\}$ to $\mathcal{P}(X)$, the powerset of X, which extends to other formulae by*

1. $v(\neg\alpha) = -(v(\alpha)^*)$

2. $v(\alpha \vee \beta) = v(\alpha) \cup v(\beta)$

3. $v(\alpha \wedge \beta) = v(\alpha) \cap v(\beta)$

4. $v(\Box\alpha) = \{x \in X \mid \forall y(xRy \Rightarrow y \in v(\alpha))\}$

5. $v(\bigcirc\alpha) = \{x \in X \mid \forall y(xR'y \Rightarrow y \in v(\alpha))\}$ $\Box$

A world $x \in X$ is said to ***satisfy*** α in the ***model*** $< \mathbf{C}, v >$ if $x \in v(\alpha)$. This is written as $\mathbf{C}, v, x \models \alpha$. A model $< \mathbf{C}, v >$ satisfies α , or $\mathbf{C}, v \models \alpha$ if for all $x \in X$ we have $\mathbf{C}, v, x \models \alpha$. Where the context is clear, this is sometimes abbreviated to $x \models \alpha$. And α is ***valid*** in a frame $\mathbf{C}$ if for all valuations v on $\mathbf{C}$ we have $\mathbf{C}, v \models \alpha$. This is written as $\mathbf{C} \models \alpha$. In the usual manner $\not\models$ is used to denote the failure of these concepts. It is important to see that it follows from the definitions that $\mathbf{C}, v, x \models \neg\alpha$ iff $\mathbf{C}, v, x^* \not\models \alpha$, and $\mathbf{C}, v, x \models \alpha$ iff $\mathbf{C}, v, x^* \not\models \neg\alpha$.

For example, it can be shown that for an arbitrary frame $\mathbf{C}$, $\Box\alpha \vee \neg \bigcirc \alpha$ is valid in $\mathbf{C}$. Let v be any valuation on $\mathbf{C}$ and let $x \in X$. Then $\mathbf{C}, v, x \models \Box\alpha \vee \neg\bigcirc\alpha$ iff $\mathbf{C}, v, x \models \Box\alpha$ or $\mathbf{C}, v, x \models \neg \bigcirc \alpha$ iff $\forall y(xRy \Rightarrow y \models \alpha)$ or $x^* \not\models \bigcirc\alpha$, that is $\exists y(x^*R'y$ and $y \not\models \alpha)$. If the first disjunct fails, then $\exists y(xRy$ and $y \not\models \alpha)$ and so $x^*R'y$, so the second disjunct is true. It can be shown similarly that $\bigcirc\alpha \vee \neg\Box\alpha$ is valid in all frames. As another useful example, let $< \mathbf{C}, v >$ be an arbitrary frame model. Then

Theorem 4 $\mathbf{C}, v \models \Box\alpha$ ***iff*** $\mathbf{C}, v \models \bigcirc\alpha$.

Proof. Suppose $\mathbf{C}, v \not\models \bigcirc\alpha$ because $\mathbf{C}, v, x \not\models \bigcirc\alpha$ with $xR'y$ and $y \not\models \alpha$. But then x^*Ry, so $x^* \not\models \Box\alpha$ and $\mathbf{C}, v \not\models \Box\alpha$. The converse clearly holds by a similar proof. $\Box$

Corollary 5 $\mathbf{C} \models \neg\Box\alpha$ ***iff*** $\mathbf{C} \models \neg \bigcirc \alpha$; *if* $\mathbf{C} \models \Box\alpha$ *then* $\mathbf{C} \not\models \neg\Box\alpha$; *if* $\mathbf{C} \models \bigcirc\alpha$ *then* $\mathbf{C} \not\models \neg \bigcirc \alpha$. $\Box$

The classical notion of frame morphism or *p-morphism* is here extended to these frames.

Definition 6 *Given frames* $\mathbf{C}_1 =< X, *, R >$ *and* $\mathbf{C}_2 =< Y, *, S >$ *a frame morphism* $\mathbf{C}_1 \xrightarrow{f} \mathbf{C}_2$ *is a mapping* $f : X \mapsto Y$ *such that:*

1. *If* xRx' *then* $f(x)Sf(x')$;
2. $f(x^*) = f(x)^*$;
3. *If* $f(x)Sy$ *then* $\exists x' \in X$ *with* xRx' *and* $f(x') = y$. □

Given a frame morphism $\mathbf{C}_1 \xrightarrow{f} \mathbf{C}_2$, clauses 1 and 3 also hold for the relations R' and S'. This is because, treating $*$ as a relation, R' can be regarded as the relational composition of $*$ with R, and these properties both hold for $*$ and R. This point is important because it shows that any validity preserving properties of frame morphisms regarding the operator $\Box$ also hold for $\bigcirc$.

A frame morphism is *injective* if it is injective as a function - if $x \neq y$ implies that $f(x) \neq f(y)$ - and *surjective* if surjective as a function: if for all $y \in Y$ there is some $x \in X$ with $f(x) = y$. If it is both injective and surjective then it is an *isomorphism.* If $\mathbf{C}_1 \xrightarrow{f} \mathbf{C}_2$ is an isomorphism, then so is its inverse, and $\mathbf{C}_1$ and $\mathbf{C}_2$ are said to be *isomorphic*: from a logical point of view they are indistinguishable. If $\mathbf{C}_1 \xrightarrow{f} \mathbf{C}_2$ is injective then $\mathbf{C}_1$ is isomorphic to the *subframe* $f(\mathbf{C}_1)$ of $\mathbf{C}_2$; and if $\mathbf{C}_1 \xrightarrow{f} \mathbf{C}_2$ is surjective then $\mathbf{C}_2$ is a *p-morphic image* of $\mathbf{C}_1$. It is often convenient to regard a frame that is isomorphic to a subframe of $\mathbf{C}_2$ as though it were that subframe, though it should always be borne in mind that strictly the former notion applies.

Let $\mathbf{C}_1 \xrightarrow{f} \mathbf{C}_2$ be injective. Then the image of f is what is often known as a generated subframe of $\mathbf{C}_2$, because the conditions for a frame morphism guarantee that the image $f(\mathbf{C}_1)$ of f is a subset of $\mathbf{C}_2$ closed under the join of the relations $*$, S and S'. For $f(x)$ in $\mathbf{C}_2$, any path from $f(x)$ in relations from these three is also a path from x in the corresponding relations of $\mathbf{C}_1$ if, for $x' \in X$ we regard x' and $f(x')$ as the same elements.

Theorem 7 *If* $\mathbf{C}_1 \xrightarrow{f} \mathbf{C}_2$ *is injective and* $\mathbf{C}_2 \models \alpha$ *then* $\mathbf{C}_1 \models \alpha$.

Proof. Let $\mathbf{C}_1$ and $\mathbf{C}_2$ be as above. For a valuation v on $\mathbf{C}_1$ with $\mathbf{C}_1, v, x \not\models \alpha$ let v' be *any* valuation such that for all p_i and for all $x' \in X$ $\mathbf{C}_1, v, x' \models p_i$ iff $\mathbf{C}_2, v', f(x') \models p_i$. Clearly such a valuation v' exists. Then $\mathbf{C}_2, v', f(x) \not\models \alpha$ so $\mathbf{C}_2, \not\models \alpha$. □

Now let $\mathbf{C}_1 \xrightarrow{f} \mathbf{C}_2$ be surjective and let v be a valuation on $\mathbf{C}_2$. Then where $f^{-1}(Z) =_{def} \{x \mid f(x) \in Z\}$, define a valuation v' on $\mathbf{C}_1$ by $v'(p_i) = f^{-1}(v(p_i))$. This is indeed a valuation: clearly for $Z, Z' \in Y$ we have $f^{-1}(Z \cap Z') = f^{-1}(Z) \cap f^{-1}(Z')$ and $f^{-1}(Z \cup Z') = f^{-1}(Z) \cup f^{-1}(Z')$. Defining $\Box Z =_{def} \{y \in Y \mid \forall z(yRz \Rightarrow z \in Z)\}$, then $f^{-1}(\Box Z) \subseteq \Box f^{-1}(Z)$ is a restatement of the first part, and $\Box f^{-1}(Z) \subseteq f^{-1}(\Box Z)$ is a restatement of the third part, of the definition of frame morphism; so $f^{-1}(\Box Z) = \Box f^{-1}(Z)$. Finally, $f^{-1}(-Z^*) = -f^{-1}(Z^*) = -f^{-1}(Z)^*$. This shows v' to be a valuation on $\mathbf{C}_1$ with $\mathbf{C}_1, v', x \models \alpha$ if $\mathbf{C}_2, v, f(x) \models \alpha$.

Theorem 8 *If* $\mathbf{C}_1 \xrightarrow{f} \mathbf{C}_2$ *is surjective and* $\mathbf{C}_1 \models \alpha$ *then* $\mathbf{C}_2 \models \alpha$.

Proof. Suppose $\mathbf{C}_2 \not\models \alpha$ because $\mathbf{C}_2, v, y \not\models \alpha$. Then with v' as above $\mathbf{C}_1, v', x \not\models \alpha$, where $f(x) = y$. □

For $\{\mathbf{C}_i \mid i \in I\}$ a collection of frames with $\mathbf{C}_i =< X_i, *, R_i >$, their *disjoint union* $\sum_{i \in I} \mathbf{C}_i$ is the frame $< \sum_{i \in I} X_i, *, R >$ where $\sum_{i \in I} X_i = \{(x, i) \mid x \in X_i, i \in I\}$, $(x, i)R(x', j)$ if $i = j$ and xR_ix', and $(x, i)^* = (x^*, i)$. Then for $f_j(x) = (x, j)$, it is straightforward to check that $\mathbf{C}_j \xrightarrow{f_j} \sum_{i \in I} \mathbf{C}_i$ is an injective frame morphism.

Theorem 9 *If for all* $i \in I$ $\mathbf{C}_i \models \alpha$, *then* $\sum_{i \in I} \mathbf{C}_i \models \alpha$.

Proof. Suppose $\sum_{i\in I} \mathbf{C}_i \not\models \alpha$ because $\sum_{i\in I} \mathbf{C}_i, v, (x, j) \not\models \alpha$; then v restricted to $\mathbf{C}_j$ is a valuation on $\mathbf{C}_j$, so $\mathbf{C}_j, v, x \not\models \alpha$. □

These preservation results are all classical results which transfer directly to four-valued frames. There is another preservation result, as well as a further conditional one concerning *canonical extensions*, which will be given in chapter 6.

Before turning to algebraic models for the logic, we shall close this section by making explicit the four-valued logic on which our modal logic is based, as well as the comparison between this logic and that of Lakemeyer's outlined in the previous section. The four values of the four-valued logic we are using may be named T, F, TF and $*$; or only-true, only-false, both-true-and-false and neither-true-nor-false respectively. The last value should not of course be confused with the frame relation.

The truth-functional behaviour of the connectives with regard to these values is as follows: in the four element truth-function lattice with $* \wedge TF = F$ and $* \vee TF = T$, the connective $\wedge$ is meet, the connective $\vee$ is join and $\neg$ is the involution with fixed points at $*$ and TF. These are *monotonic* connectives: in the *degree-of-definedness* lattice $\mathcal{T}$ with $T \wedge F = *$ and $T \vee F = TF$, the truth function defined by each of these n-ary connectives is a monotonic function from $\mathcal{T}^n$ to $\mathcal{T}$.

In a model, the value $v(s, \alpha)$ of a formula α at a world s is T if $s \models \alpha$ and $s \not\models \neg\alpha$, F if $s \not\models \alpha$ and $s \models \neg\alpha$, TF if $s \models \alpha$ and $s \models \neg\alpha$, and $*$ if $s \not\models \alpha$ and $s \not\models \neg\alpha$. If we define the non-monotonic truth function u by $u(T) = T$, $u(F) = F$, $u(TF) = *$ and $u(*) = TF$, then it is easy to see that in all models $v(s^*, \alpha) = u(v(s, \alpha))$. This *truth function* u could be defined by using the frame relation $*$ to define a modal operator in the usual way, but as we have seen it is not really modal in character. To add this to the language would be equivalent to adding a classical negation: if u is also the language connective defined modally by $*$, then we have $s \not\models \alpha$ iff $s \models u(\neg\alpha)$. So this will not be done here, although models appropriate for this extension are considered in the final chapter.

It is routine to check that the propositional logic defined by our frames is indeed this four-valued logic. For example, if $v(s, \alpha) = TF$ and $v(s, \beta) = *$, then $v(s, \alpha \wedge \beta) = F$: clearly $s \not\models \alpha \wedge \beta$ because $s \not\models \beta$; but also $s \models \neg(\alpha \wedge \beta)$, because $s \models \neg\alpha$ and so $s^* \not\models \alpha$, implying that $s^* \not\models \alpha \wedge \beta$ and so $s \models \neg(\alpha \wedge \beta)$. So, by definition, $v(s, \alpha \wedge \beta) = F$. That this is also the underlying four-valued logic we saw in the previous chapter is most easily seen by treating the values T, F, TF and $*$ as sets $\{T\}$, $\{F\}$, $\{T, F\}$ and $\emptyset$, and then noting that the definition of satisfaction for each of the proposition connectives there corresponds to the truth functions above.

To compare our structures with those of Lakemeyer with respect to their common language we will, as noted at the beginning of the chapter deal only with models and will temporarily use the restricted, classical definition of validity. First with any of Lakemeyer's BLK models $M =< W, T, F, R, R^- >$ we associate another such model M' which has the same logic. This model M' is then easily shown to have the same logic as one of our own models. A converse association is even easier to demonstrate. Let $C \subseteq W$ be the classical worlds of the model M - those $w \in W$ such that for all p_i, $w \in T(p_i)$ iff $w \notin F(p_i)$. Augment W to W' by adding fresh worlds w^* for each $w \in W - C$, so letting $W' = W \cup \{w^* \mid w \in W - C\}$. Augment the valuation pair $< T, F >$ to cover these new worlds by setting $w^* \in T(p_i)$ iff $w \notin F(p_i)$ and $w^* \in F(p_i)$ iff $w \notin T(p_i)$. Finally, augment R and R^- by $w^* Rv$ iff wR^-v and $w^* R^- v$ iff wRv. Let this extended model be M', and note that for no $v \in W'$ and $w^* \in W' - W$ do we have vRw^* or vR^-w^*: in particular, identifying the worlds of M naturally with worlds in M', it can be seen that the two models share precisely the same classical worlds; that in both cases each of these is related to precisely the same set of worlds; and their valuations coincide for the worlds W. So the worlds $W' - W$ are superfluous in the definition of validity. From this it follows that M' has the same logic as its submodel M.

Obviously, M' was constructed to look like one of our own four-valued models. Define the frame operation $*$ by $(w^*)^* = w$ and for $w \in C$, $w^* = w$. Then because R and R^- coincide at classical worlds in Lakemeyer's models, and because of the way we defined R and R^- in M', we clearly have wRv iff w^*R^-v; so $\mathbf{C} =< W', *, R >$ is a frame in which R^- coincides with our defined relation R'. Then we define a valuation v on this frame by $v(p_i) = T(p_i)$: the way in which T and F were extended to M' shows that $F(p_i) = -v(p_i)^*$, and so, given that the relational structure is the same in the two models, we have $M', w \models \alpha$ iff $\mathbf{C}, v, w \models \alpha$ for any α in the common language. It follows that any BLK model determines one of our models with the same logic.

The converse is clear: for any frame $\mathbf{C} =< W, *, R >$ such that all $w \in W$ with $w^* = w$ satisfy the relational constraints imposed on classical worlds in BLK models, and for any valuation v on $\mathbf{C}$, a BLK model with the same logic as $< \mathbf{C}, v >$ may be defined: use the underlying birelational structure $< W, R, R' >$, together with the valuations defined by $T(p_i) = v(p_i)$ and $F(p_i) = -v(p_i)^*$. Verification is then immediate, showing that Lakemeyer's models may in this way be put in logical correspondence with that class of our models satisfying the appropriate relational constraints.

4.2 Algebras

Definition 10 *A modal algebra* $\mathbf{A}$ *is a structure* $< A, \wedge, \vee, \neg, \nu, \mu, 0, 1 >$ *of type* $< 2, 2, 1, 1, 1, 0, 0 >$ *where for* $a, b, \in A$

1. $< A, \wedge, \vee, 0, 1 >$ *is a bounded distributive lattice;*

2. $\neg(a \wedge b) = \neg a \vee \neg b$, $\neg(a \vee b) = \neg a \wedge \neg b$ *and* $\neg\neg a = a$;

3. $\nu(a \wedge b) = \nu a \wedge \nu b$, $\nu 1 = 1$ *and* $\nu a \vee \neg\mu a = 1$;

4. $\mu(a \wedge b) = \mu a \wedge \mu b$, $\mu 1 = 1$ *and* $\mu a \vee \neg\nu a = 1$. □

Note that we also have $\neg 0 = 1$ and $\neg 1 = 0$, $\nu a \wedge \neg\mu a = 0$ and $\mu a \wedge \neg\nu a = 0$. As usual we sometimes write $a \leq b$ for $a \wedge b = a$. Associated with each formula α of the language in n propositional variables is an n-ary polynomial function p^α on any modal algebra $\mathbf{A}$: $\Box$ is interpreted as ν and $\bigcirc$ is interpreted as μ, and the other connectives get the obvious interpretations. If this function is constantly equal to 1 for any sequence of elements of the algebra as arguments, then α is *valid* in $\mathbf{A}$, or $\mathbf{A} \models \alpha$. Viewing p^α as a term t^α, this is equivalent to requiring that the algebra satisfies the identity $t^\alpha = 1$.

Recall that in algebra a mapping $\mathbf{A}_1 \xrightarrow{h} \mathbf{A}_2$ from the universe A_1 of $\mathbf{A}_1$ to the universe A_2 of $\mathbf{A}_2$ is a *homomorphism* if it preserves the operations: if $a_1, \ldots a_n$ are in A and ω is an n-ary operation derived from those of the signature, then $h(\omega(a_1, \ldots a_n)) = \omega(h(a_1), \ldots h(a_n))$. An isomorphism is defined to be an injective and surjective homomorphism, and similarly to frames, isomorphic structures are algebraically indistinguishable. If $\mathbf{A}_1 \xrightarrow{h} \mathbf{A}_2$ is injective, then $\mathbf{A}_1$ is isomorphic to a *subalgebra* of $\mathbf{A}_2$ - sometimes it is convenient to regard it as in fact a subalgebra; and if it is surjective then $\mathbf{A}_2$ is a *homomorphic image* of $\mathbf{A}_1$. For a class K of algebras, $S(K)$ is the class of algebras isomorphic to subalgebras of members of K, and $H(K)$ is the class of algebras isomorphic to homomorphic images of members of K.

Theorem 11 *If* $\mathbf{A}_1 \xrightarrow{h} \mathbf{A}_2$ *is injective and* $\mathbf{A}_2 \models \alpha$ *then* $\mathbf{A}_1 \models \alpha$.

Proof. We may regard the universe A_1 as a subset of A_2, and this is closed there under the operations of $\mathbf{A}_2$. So suppose that for some α with n propositional variables and for some $a_1, \ldots a_n \subseteq A_1 \subseteq A_2$ we have $p^\alpha(a_1, \ldots a_n) \neq 1$. But then $p^\alpha(a_1, \ldots a_n) \in A_1$ and h is injective, so $p^\alpha(a_1, \ldots a_n) \neq 1$ in $\mathbf{A}_2$. That is, $\mathbf{A}_2 \not\models \alpha$. □

Theorem 12 *If* $\mathbf{A}_1 \xrightarrow{h} \mathbf{A}_2$ *is surjective and* $\mathbf{A}_1 \models \alpha$ *then* $\mathbf{A}_2 \models \alpha$.

Proof. If the antecedent holds then for any $a_1, \ldots a_n \subseteq A_1$, $p^\alpha(a_1, \ldots a_n) = 1$. Choose any $b_1, \ldots b_n \subseteq A_2$ with $h(a_i) = b_i$. Then because $h(1) = 1$,

$$h(p^\alpha(a_1, \ldots a_n)) = p^\alpha(h(a_1), \ldots h(a_n)) = 1.$$

So $p^\alpha(b_1, \ldots b_n) = 1$. □

For algebras $\{\mathbf{A}_i \mid i \in I\}$ define their *product* $\prod_{i \in I} \mathbf{A}_i$ to be the algebra with universe the cartesian product $\prod_{i \in I} A_i$ of the universes A_i, and with operations defined pointwise: for n-ary operator ω and $a_1, \ldots a_n \in \prod_{i \in I} A_i$, define $\omega(a_1, \ldots a_n)(i)$ to be $\omega(a_1(i), \ldots a_n(i))$.

Theorem 13 *If* $\{\mathbf{A}_i \mid i \in I\}$ *are such that for all* $i \in I$, $\mathbf{A}_i \models \alpha$; *then we have* $\prod_{i \in I} \mathbf{A}_i \models \alpha$.

Proof. If p^α is n-ary and $a_1, \ldots a_n \in \prod_{i \in I} A_i$, then for all $i \in I$ $p^\alpha(a_1(i), \ldots a_n(i)) = 1$. So for all i $p^\alpha(a_1, \ldots a_n)(i) = 1$, then $p^\alpha(a_1, \ldots a_n) = 1$ and $\prod_{i \in I} \mathbf{A}_i \models \alpha$. □

If K is a class of algebras then $P(K)$ is the class consisting of those algebras isomorphic to products of algebras in K. A class of algebras closed under H, S and P is called a *variety*, and the smallest variety containing a class K of algebras was shown by Tarski to be $HSP(K)$. From a theorem by Birkhoff, it is known that varieties are precisely equational classes: for any variety V there is a set of identities such that V is the class of all algebras satisfying all of them; and conversely the class of all algebras satisfying a given set of identities is a variety.

4.3 Completeness

We now spell out the logic $\vdash_K$ by giving the following axiom schema and rules, where Γ and Δ vary over finite sets of propositional formulae.

$$p_i \vdash p_i \qquad \frac{\Gamma \vdash \alpha, \Delta}{\Gamma \vdash \neg\neg\alpha, \Delta}\neg\neg I \qquad \frac{\Gamma, \alpha \vdash \Delta}{\Gamma, \neg\neg\alpha \vdash \Delta}\neg\neg E$$

$$\frac{\Gamma \vdash \alpha, \Delta \quad \Gamma \vdash \beta, \Delta}{\Gamma \vdash \alpha \wedge \beta, \Delta}\wedge I \qquad \frac{\Gamma, \alpha \vdash \Delta}{\Gamma, \alpha \wedge \beta \vdash \Delta}\wedge E1 \qquad \frac{\Gamma, \beta \vdash \Delta}{\Gamma, \alpha \wedge \beta \vdash \Delta}\wedge E2$$

$$\frac{\Gamma \vdash \neg\alpha, \Delta \quad \Gamma \vdash \neg\beta, \Delta}{\Gamma \vdash \neg(\alpha \vee \beta), \Delta}\neg \vee I \qquad \frac{\Gamma, \neg\alpha \vdash \Delta}{\Gamma, \neg(\alpha \vee \beta) \vdash \Delta}\neg \vee E1 \frac{\Gamma, \neg\beta \vdash \Delta}{\Gamma, \neg(\alpha \vee \beta) \vdash \Delta}\vee E2$$

$$\frac{\Gamma, \alpha \vdash \Delta \quad \Gamma, \beta \vdash \Delta}{\Gamma, \alpha \vee \beta \vdash \Delta}\vee E \qquad \frac{\Gamma \vdash \alpha, \Delta}{\Gamma \vdash \alpha \vee \beta, \Delta}\vee I1 \qquad \frac{\Gamma \vdash \beta, \Delta}{\Gamma \vdash \alpha \vee \beta, \Delta}\vee I2$$

$$\frac{\Gamma, \neg\alpha \vdash \Delta \quad \Gamma, \neg\beta \vdash \Delta}{\Gamma, \neg(\alpha \wedge \beta) \vdash \Delta}\neg \wedge E \qquad \frac{\Gamma \vdash \neg\alpha, \Delta}{\Gamma \vdash \neg(\alpha \wedge \beta), \Delta}\neg \wedge I1 \frac{\Gamma \vdash \neg\beta, \Delta}{\Gamma \vdash \neg(\alpha \wedge \beta), \Delta}\neg \wedge I2$$

$$\frac{\Gamma \vdash \alpha}{\Box\Gamma \vdash \Box\alpha}K^{\Box} \qquad \frac{\Gamma \vdash \bigcirc\alpha, \Delta}{\Gamma, \neg\Box\alpha \vdash \Delta}\neg\Box E \qquad \frac{\Gamma, \bigcirc\alpha \vdash \Delta}{\Gamma \vdash \neg\Box\alpha, \Delta}\neg\Box I$$

$$\frac{\Gamma \vdash \alpha}{\bigcirc\Gamma \vdash \bigcirc\alpha}K^{\bigcirc} \qquad \frac{\Gamma \vdash \Box\alpha, \Delta}{\Gamma, \neg\bigcirc\alpha \vdash \Delta}\neg\bigcirc E \qquad \frac{\Gamma, \Box\alpha \vdash \Delta}{\Gamma \vdash \neg\bigcirc\alpha, \Delta}\neg\bigcirc I$$

together with the structural rules of Exchange, Weakening, Contraction and Cut. The propositional rules can be seen to be equivalent to:

1. $\alpha \vdash \alpha$;

2. $\alpha \vdash \beta \wedge \gamma$ iff $\alpha \vdash \beta$ and $\alpha \vdash \gamma$;

3. $\alpha \vee \beta \vdash \gamma$ iff $\alpha \vdash \gamma$ and $\beta \vdash \gamma$;

4. $\neg\neg\alpha \dashv\vdash \alpha$;

5. $\alpha \vdash \beta$ implies $\neg\beta \vdash \neg\alpha$.

In addition to these we also require axioms to the effect that $\vee$ and $\wedge$ are commutative and associative and that they distribute over one another. The rules $\neg\Box I$, $\neg\Box E$, $\neg\bigcirc I$ and $\neg\bigcirc E$ allow the derivations of $\vdash \Box\alpha\vee\neg\bigcirc\alpha$, $\vdash \bigcirc\alpha\vee\neg\Box\alpha$, $\Box\alpha \wedge \neg\bigcirc\alpha \vdash$ and $\bigcirc\alpha \wedge \neg\Box\alpha \vdash$. For example, $\vdash \Box\alpha \vee \neg\bigcirc\alpha$ can be shown to result from $\Box\alpha \vdash \Box\alpha$ by successive applications of $\neg\bigcirc I$, $\vee I1$, Exchange, $\vee I2$ and then a contraction. The other theorems are shown similarly. Moreover, without further axioms added to $\vdash_K$, this is all that these four rules can prove by themselves, since they simply say that certain formulae act as classical negations of each other - compare the sequent rules for classical negation. Finally the rules $K^{\Box}$ and $K^{\bigcirc}$ have the same properties as in classical modal logic:

$$\Box(\alpha\wedge\beta) \dashv\vdash \Box\alpha\wedge\Box\beta \text{ and } \vdash \Box\alpha \text{ if } \vdash \alpha\text{; and } \bigcirc(\alpha\wedge\beta) \dashv\vdash \bigcirc\alpha\wedge\bigcirc\beta \text{ and } \vdash \bigcirc\alpha \text{ if } \vdash \alpha.$$

As in the classical case this formulation is equivalent to the sequent calculus formulation. So, comparing this axiomatisation to the equations used to define algebras, informally we have shown:

Theorem 14 *The algebraic semantics is sound and complete for* $\vdash_K$. $\Box$

We will usually think of a modal logic L as a set of formulae, called axioms, closed under the substitution of formulae and closed under the rules of $\vdash_K$, but formally we must use sequents in place of these formulae. Thus all axioms of $\vdash_K$ are included in L. Choosing to present a logic as a set of formulae rather than a set of sequents would allow us to describe only one propositional extension of $\vdash_K$, namely that given by $\vdash \alpha \vee \neg\alpha$ which is classical modal logic. It might be thought that this also fails to describe the extension given by the sequent $\alpha \wedge \neg\alpha \vdash$ to give a three-valued base to the logic; but having recast the system with the rule that $\alpha \vdash \beta$ implies $\neg\beta \vdash \neg\alpha$, the presence of modal theorems shows this extension to be the same as that by $\vdash \alpha \vee \neg\alpha$. In fact, a three-valued modal logic is given by adding the sequent $\alpha \wedge \neg\alpha \vdash \beta \vee \neg\beta$, which has no formula equivalent; but in most cases the axiom presentation is a convenient shorthand. This blunter approach will generally be adopted because these propositional matters are not our concern here; but because of the above example, and because in general there are no propositional theorems, we must take note of the sequent formulation in showing completeness below.

The collection of all modal logics ordered by inclusion is a lattice, with meet in this lattice being intersection - the classical modal logics are a sublattice of this lattice. If Σ is a set of formulae, let $V(\Sigma)$ be the class of modal algebras determined by Σ: the algebras in which all formulae of Σ are valid. Now it is easily shown that $\alpha \vdash \beta$ iff $\alpha \wedge \beta \dashv\vdash \alpha$, so for a set of sequents Σ, $V(\Sigma)$ is defined to be the class of all algebras $\mathbf{A}$ such that $\mathbf{A} \models t^\alpha = t^\beta$ for α and β such that $\alpha \dashv\vdash \beta \in \Sigma$. The sequent $\alpha \dashv\vdash \beta$ is then said to be valid in an algebra if it satisfies this equation. Either way $V(\Sigma)$ is an equational class and so, by Birkhoff's theorem, a variety. Now for an algebra $\mathbf{A}$ let $L(\mathbf{A})$ be the *logic* of $\mathbf{A}$, the sequents valid in $\mathbf{A}$. Because a modal algebra is closed under the rules of $\vdash_K$, $L(\mathbf{A})$ is indeed a logic. Extending this definition to a class C of algebras, let $L(C) = \bigcap\{L(\mathbf{A}) \mid \mathbf{A} \in C\}$, which is a logic, because logics are closed under arbitrary intersection. Clearly, varieties determine logics.

Now every variety is defined by equations, and every equation is of the form $t^\alpha = t^\beta$, for α, β formulae in the modal language: let $\mathbf{A} \models t^\alpha = t^\beta$, for every $\alpha \dashv\vdash \beta \in \Sigma$ - then it is not difficult to check that $\mathbf{A} \in V(\Sigma)$. It follows that every variety is of the form $V(\Sigma)$ for some set of sequents Σ. Now let L^Σ be the smallest logic containing Σ and let $\mathbf{A} \in V(\Sigma)$. Then $\Sigma \subseteq L(\mathbf{A})$ so $L^{\Sigma} \subseteq L(\mathbf{A})$; that is, all the sequents of Σ are valid in $\mathbf{A}$. So $\mathbf{A} \in V(L^{\Sigma})$. The converse is trivial: if $\mathbf{A} \in V(L^{\Sigma})$ then $\mathbf{A} \in V(\Sigma)$. Consequently Σ and L^{Σ} determine the same variety, so V is a surjective mapping from modal logics to varieties of modal algebras.

Let L be a logic; then so is $L(V(L))$. But $V(L)$ is the equational class determined by the sequents of L, so every sequent of L is valid in each $\mathbf{A} \in V(L)$, which means that these are valid in $L(V(L)) = \bigcap\{L(\mathbf{A}) \mid \mathbf{A} \in V(L)\}$. To show that the reverse inclusion holds, it is shown that if $\alpha \vdash \beta$ is not in L, then it is not valid in any *Lindenbaum algebra* of L, which is in $V(L)$; and then $\alpha \vdash \beta$ is not in $\bigcap\{L(\mathbf{A}) \mid \mathbf{A} \in V(L)\} = L(V(L))$. Assume that the language has some fixed infinite set of propositional variables: a logic then induces an equivalence relation on its formulae given by $\dashv\vdash$. For a given logic L, let $[\alpha]$ be the equivalence class of α according to this relation, and let the universe of the Lindenbaum algebra $\mathbf{A}^L$ be the set of all such equivalence classes. Define the algebraic operations by $[\alpha] \wedge [\beta] = [\alpha \wedge \beta]$, $[\alpha] \vee [\beta] = [\alpha \vee \beta]$, $\neg[\alpha] = [\neg\alpha]$, $\nu[\alpha] = [\Box\alpha]$, $\mu[\alpha] = [\bigcirc\alpha]$, $0 = [\Box\alpha \wedge \neg\bigcirc\alpha]$ and $1 = [\Box\alpha \vee \neg\bigcirc\alpha]$. This is well-defined, because if $[\alpha] = [\gamma]$ and so $\alpha \dashv\vdash \gamma$, the logic shows that $\alpha \wedge \beta \dashv\vdash \gamma \wedge \beta$, $\alpha \vee \beta \dashv\vdash \gamma \vee \beta$ etc. So nothing depends on the the choice of $\alpha' \in [\alpha]$, and $\mathbf{A}^L$ is a modal algebra with $\alpha \dashv\vdash_L \beta$ iff $[\alpha] = [\beta]$.

That $\alpha \dashv\vdash \beta \in L$ iff $\mathbf{A}^L \models t^{\alpha} = t^{\beta}$ is then shown in the standard way. Let σ be a uniform substitution of variables for formulae: $\sigma(\alpha)$ is the result of replacing each variable p_i in α by ψ_i, where $\sigma(p_i) = \psi_i$. Now we know already that the valid equivalences of L are closed under the application of any such substitution to all formulae in that equivalence, so $\alpha \dashv\vdash \beta$ iff $\forall\sigma(\sigma(\alpha) \dashv\vdash \sigma(\beta))$. Let p^{α} be the term function of α. In other words, for a sequence $\vec{a} = (a_1, a_2, \ldots)$ of elements of the algebra, if α is a variable p_i then $p^{\alpha}(\vec{a}) = a_i$; if ω is a n-ary connective and $\alpha = \omega(\beta_1, \ldots \beta_n)$ then we have $p^{\alpha}(\vec{a}) = \omega(p^{\beta_1}(\vec{a}), \ldots p^{\beta_n}(\vec{a}))$, where ω is here the operation in the algebra corresponding to the language connective.

Let $\vec{v}$ be the sequence $([p_0],[p_1],\ldots)$ in $\mathbf{A}^L$. Also define $\sigma(\vec{v})$ to be the sequence $([\sigma(p_0)],[\sigma(p_1)],\ldots)$. Clearly every sequence of elements of $\mathbf{A}^L$ is of the form $\sigma(\vec{v})$ for some σ. Given the construction of $\mathbf{A}^L$ an easy induction on the length of formulae shows that $p^\alpha(\vec{v}) = [\alpha]$. The proof is little more than a restatement of the definition of p^α. Also, by considering the definition of σ and of $\sigma(\vec{v})$, it can be seen that $p^{\sigma(\alpha)}(\vec{v}) = p^\alpha(\sigma(\vec{v}))$. So it follows that $[\sigma(\alpha)] = p^{\sigma(\alpha)}(\vec{v})$. So altogether is can be seen that $\alpha \dashv\vdash \beta \in L$ iff $\forall\sigma(\sigma(\alpha) \dashv\vdash \sigma(\beta) \in L)$ iff $\forall\sigma([\sigma(\alpha)] = [\sigma(\beta)])$ iff $\forall\sigma(p^{\sigma(\alpha)} = p^{\sigma(\beta)})$ iff $\forall\sigma(p^\alpha(\sigma(\vec{v})) = p^\beta(\sigma(\vec{v})))$ iff $p^\alpha = p^\beta$. But this is just to say that $\mathbf{A}^L \models \alpha \dashv\vdash \beta$ as promised. So we have shown that $L = L(V(L))$; as a mapping, V has an inverse L. Finally, suppose that $L \subseteq L'$; then any algebraic model for L' is also a model for L, that is, $V(L') \subseteq V(L)$. This shows that the lattice of modal logics is anti-isomorphic to the lattice of varieties of modal algebras.

4.4 A Representation Theorem

The following structure is in places similar to one of Goldblatt[1], which is based on a construction by Priestley[2], so parts of the following proof may be found there. These are reproduced here for completeness and intelligibility.

We begin with some definitions. Given a set X, a *topology* $\mathcal{T}$ on X is a set of subsets of X containing $\emptyset$ and X and closed under finite intersection and arbitrary union: that is, if $U, V \in \mathcal{T}$, then $U \cap V \in \mathcal{T}$, and if $\{U_i \mid i \in I\} \subseteq \mathcal{T}$, then $\bigcup_{i\in I} U_i \in \mathcal{T}$. $< X, \mathcal{T} >$ is then called a *topological space.* A *base* for $\mathcal{T}$ is a subset $\mathcal{B} \subseteq \mathcal{T}$ such that for every $U \in \mathcal{T}$, U is the union of elements of $\mathcal{B}$. A

[1]Goldblatt 1989.

[2]Priestley 1970.

subbase for $\mathcal{T}$ is a subset $\mathcal{C} \subseteq \mathcal{T}$ the finite intersections of which form a base. If $U \in \mathcal{T}$ then U is called *open* and $-U$ *closed*; if U is both open and closed then it is ***clopen***. A subset $U \subseteq X$ is ***compact*** if for any $\{V_i \mid i \in I\} \subseteq \mathcal{T}$ with $U \subseteq \bigcup_{i \in I} V_i$ then there is a finite $J \subseteq I$ with $U \subseteq \bigcup_{j \in J} V_j$. Any such $\{V_i \mid i \in I\} \subseteq \mathcal{T}$ is called an *open cover* of U. A space $< X, \mathcal{T} >$ is compact if X is. Note that if $< X, \mathcal{T} >$ is compact then any closed $U \subseteq X$ is compact: for if $U \subseteq \bigcup_{i \in I} V_i$ with each V_i open and with U closed, then $-U$ is open and $-U \cup \bigcup_{i \in I} V_i = X$. Since X is compact we have some finite $J \subseteq I$ with $-U \cup \bigcup_{j \in J} V_j = X$, and so $U \subseteq \bigcup_{j \in J} V_j$.

Given a partial ordering $\leq$ on a set X, a ***cone*** $Y \subseteq X$ is an upward closed subset of X: that is, if $y \in Y$ and $y \leq z$ then $z \in Y$. A ***cocone*** is a subset the complement of which is a cone: it is a downward closed set. An ***ordered topological space*** $< X, \leq, \mathcal{T} >$ consists of a topological space $< X, \mathcal{T} >$ and a partial ordering $< X, \leq >$. An ordered topological space is ***totally order separated*** if for any $x, y \in X$, whenever $x \not\leq y$ then there is some clopen cone Y with $x \in Y$ and $y \notin Y$. If, in addition, an ordered topological space is compact, then it is called a ***Priestley space***.

Theorem 15 *If $< X, \leq, \mathcal{T} >$ is a Priestley space then the clopen cones and their complements form a subbase.*

Proof. We have to show that every open set U is the union of finite intersections of clopen cones and their complements. So let U be open and let $x \in U$. Then $-U$ is closed and so it is compact. For $y \in -U$, we show that there is a V_y with $y \in V_y$ and $x \notin V_y$, such that either V_y or $-V_y$ is a clopen cone. Now either $x \not\leq y$ or $y \not\leq x$. If $y \not\leq x$ then by total order separation there is a clopen cone containing y but not x: call it V_y. If $x \not\leq y$ but $y \leq x$, then again we can find a clopen cone containing x but not y, so its complement contains y but not x: let this complement be V_y. So $\{V_y \mid y \notin U\}$ is an open cover of $-U$, which is compact and so has finite subcover, say $-U \subseteq V_{y_1} \cup \ldots \cup V_{y_n}$. So $\forall i, 1 \leq i \leq n$

we have $x \in -V_{y_i}$, so $x \in -V_{y_1} \cap \ldots \cap -V_{y_n} \subseteq U$. U is of course the union of its elements, so U is the union of finite intersections of the set of clopen cones and their complements. □

Now if $< X, *, R >$ is a frame let $Rx = \{y \mid xRy\}$, and for $Z \subseteq X$ let

$$\nu_R(Z) = \{y \mid \forall z(yRz \Rightarrow z \in Z)\}.$$

$\mu_{R'}$ is defined similarly. Then

Definition 16 *A modal space is a structure* $< X, \leq, *, R, \mathcal{T} >$ *such that*

1. $< X, \leq, \mathcal{T} >$ *is a Priestley space;*
2. $< X, *, R, >$ *is a frame;*
3. $x \leq y \Rightarrow y^* \leq x^*$*;*
4. *if* $x \leq y \leq z$ *and* yRw*, then* xRw *and* zRw*;*
5. *If* U *is a clopen cone then* U^* *and* $\nu_R(U)$ *are clopen;*
6. *For any* $x \in X$, $-Rx$ *is a union of clopen cocones.* □

A modal space without the topology, or with the discrete topology, is called an *ordered frame*. Observe that if U is a clopen cone then $\nu_R(U)$ is in fact a clopen *cone*. For suppose $x \leq y$; if $y \notin \nu_R(U)$ then $\exists z(yRz$ and $z \notin U)$. But then by 4 we have not xRz so $x \notin \nu_R(U)$. Similarly it can be shown that $\nu_R(U)$ is also a clopen cocone if U is a clopen cone. Property 4 also holds for R': for if $yR'w$ and $x \leq y \leq z$ then y^*Rw and $z^* \leq y^* \leq x^*$, so by 4 x^*Rw and z^*Rw; that is $xR'w$ and $zR'w$. It can also be seen that $\mu_{R'}(U)$ is clopen if U is a clopen cone: $x \in \mu_{R'}(U)$ iff $\forall y(xR'y \Rightarrow y \in U)$ iff $\forall y(x^*Ry \Rightarrow y \in U)$ iff $x^* \in \nu_R(U)$ iff $x \in (\nu_R(U))^*$. So since if U is clopen $\nu_R(U)$ is clopen and so is $(\nu_R(U))^*$, by 5., we have $\mu_{R'}(U)$ is clopen.

Because $x \leq y \Rightarrow y^* \leq x^*$, if Z is a cone then Z^* is a cocone and $-Z^*$ is a cone. So if Z is a clopen cone, then so is $-Z^*$, since Z^* is clopen. We now show

Theorem 17 *If $\mathcal{C} =< X, \leq, *, R, \mathcal{T} >$ is a modal space, let*

$$\mathcal{C}^+ =< \mathbf{cl}(X), \cap, \cup, \neg, \nu_R, \mu_{R'}, \emptyset, X >,$$

where $\mathbf{cl}(X)$ are the clopen cones of $\mathcal{C}$ and for $U \in \mathbf{cl}(X)$, $\neg U = -U^*$. Then $\mathcal{C}^+$ is a modal algebra.

Proof. $\mathbf{cl}(X)$ was shown immediately above to be closed under $\neg$, ν_R and $\mu_{R'}$. Let $U, V \in \mathbf{cl}(X)$; then $U \cup V$ and $U \cap V$ are clearly cones. Since clopen elements are closed under finite intersection and union we have $U \cup V$, $U \cap V \in \mathbf{cl}(X)$. $\emptyset, X \in \mathbf{cl}(X)$ because for any clopen U, $-U$ is clopen and thus so are $U \cup -U$, $U \cap -U$; clearly they are cones. Distributivity follows from the obvious fact that $\cup$ and $\cap$ distribute over one another, and it is clear that $\nu_R(X) = \mu_{R'}(X) = X$, as required by the definition of the modal operators. The other requirements are:

1. $\nu_R(U \cap V) = \nu_R(U) \cap \nu_R(V)$. In general we have $A \subseteq B \cap C$ iff $A \subseteq B$ and $A \subseteq C$, so $x \in \nu_R(U \cap V)$ iff $Rx \subseteq U$ and $Rx \subseteq V$ iff $x \in \nu_R(U) \cap \nu_R(V)$. The clause for $\mu_{R'}$ is similar.

2. $\nu_R(U) \cup \neg\mu_{R'}(U) = X$. Note that for $x \in X$, $x \in \neg\mu_{R'}(U)$ iff $x \notin (\mu_{R'}(U))^*$ iff $x^* \notin \mu_{R'}(U)$; so to show that $x^* \in \mu_{R'}(U)$ implies that $x \in \nu_R(U)$. But we saw above that $(\mu_{R'}(U))^* = \nu_R(U)$. From this it can also be seen that $\mu_{R'}(U) \cup \neg\nu_R(U) = X$.

3. $\neg\neg U = U$. $x \in \neg\neg U$ iff $x \in -(-U^*)^*$ iff $x \notin (-U^*)^*$ iff $x^* \notin (-U^*)$ iff $x^* \in U^*$ iff $x \in U$.

4. $\neg(U \cap V) = \neg U \cup \neg V$. $x \in \neg(U \cap V)$ iff $x \notin (U \cap V)^*$ iff $x \notin U^* \cap V^*$ iff $x \in -(U^* \cap V^*) = -U^* \cup -V^*$, that is, iff $x \in \neg U \cup \neg V$.

5. $\neg(U \cup V) = \neg U \cap \neg V$ is proved similarly. $\square$

The aim is to show that *every* modal algebra can be represented in this way as the clopen cones of a modal space. So let $\mathbf{A} =< A, \wedge, \vee, \neg, \nu, \mu, 0, 1 >$ be any modal algebra. A subset $\{1\} \subseteq F \subseteq A$ is a *filter* if for any $a, b \in A$, $a \wedge b \in F$ iff $a \in F$ and $b \in F$. It is a *proper* filter if $0 \notin F$, and is a *prime filter* if it is proper and for any $a, b \in F$, $a \vee b \in F$ iff $a \in F$ or $b \in F$. Let $X_{\mathbf{A}}$ be the set of prime filters of $\mathbf{A}$, and for $F \in X_{\mathbf{A}}$ let $F^* =_{def} \{a \in A \mid \neg a \notin F\}$. F^* is a prime filter, because $a \vee b \in F^*$ iff $\neg(a \vee b) = \neg a \wedge \neg b \notin F$ iff $\neg a \notin F$ or $\neg b \notin F$, because F is a filter, iff $a \in F^*$ or $b \in F^*$. F^* is proper, for if $0 \in F^*$ then $\neg 0 = 1 \notin F$, which is contrary to the assumption that F is a prime filter and therefore proper. Further define the relation R_ν on $X_{\mathbf{A}}$ by $FR_\nu G$ iff $\nu^{-1}(F) \subseteq G$, where $\nu^{-1}(F) =_{def} \{a \mid \nu a \in F\}$. Because $\nu(a \wedge b) = \nu a \wedge \nu b$ and $\nu 1 = 1$, it can be seen that $\nu^{-1}(F)$ is a filter. A relation R'_μ may be defined similarly. Finally, for $a \in A$, let $r_{\mathbf{A}}(a) =_{def} \{F \in X_{\mathbf{A}} \mid a \in F\}$.

For reference we now state without proof a version of the classic Birkhoff-Stone prime filter theorem for lattices, which will often be used below.

Theorem 18 *Let $F, G \subseteq A$ be such that for all finite sets I, J with $I \subseteq F$ and $J \subseteq G$ we have $\bigwedge I \not\leq \bigvee J$. Then there is a prime filter H with $F \subseteq H$ and $H \cap G = \emptyset$.* $\square$

Now let $\mathbf{A}$ be a modal algebra. Define $\mathbf{A}_+$ to be $< X_{\mathbf{A}}, \subseteq, *, R_\nu, \mathcal{T} >$ where $\subseteq$ is the inclusion relation on the prime filters $X_{\mathbf{A}}$, $*$ and R_ν are defined as above, and $\mathcal{T}$ is given by declaring $\{r_{\mathbf{A}}(a), -r_{\mathbf{A}}(a) \mid a \in A\}$ to be a subbase. Observe that every element of the subbase is clopen. Then $\mathbf{A}_+$ will turn out to be a modal space. First it is shown that

Theorem 19 $\mathbf{A}_+$ *is an ordered frame and a Priestley space.*

Proof. We verify the appropriate criteria and some of those defining a modal space;

1. $F^{**} = F$, because $a \in F^{**}$ iff $\neg a \notin F^*$ iff $\neg\neg a = a \in F$. So $*$ is symmetric.

2. To show that $\mathbf{A}_+$ is a frame it must still be shown that $FR_\nu G$ iff $F^* R'_\mu G$. If $\nu^{-1}(F) \subseteq G$ and $a \in \mu^{-1}(F^*)$, then $\mu a \in F^*$ and so $\neg\mu a \notin F$. But since $1 = \nu a \vee \neg\mu a \in F$ and F is prime we have $\nu a \in F$; so $a \in \nu^{-1}(F)$ and so $a \in G$. The converse is shown similarly, using the fact that $\mu a \vee \neg\nu a \in F^*$, which is prime.

3. For part 3 of the definition it must be shown that $F \subseteq G \Rightarrow G^* \subseteq F^*$. Suppose $F \subseteq G$ and $a \in G^*$. Then $\neg a \notin G$ so $\neg a \notin F$ so $a \in F^*$.

4. Let $H \subseteq F \subseteq H'$ and $FR_\nu G$. Then $\nu^{-1}(H) \subseteq \nu^{-1}(F) \subseteq G$, so $HR_\nu G$. To show $H'R_\nu G$, let $a \in \nu^{-1}(H')$. Then $\nu a \in H'$, and because H' is proper, $0 = \nu a \wedge \neg\mu a \notin H'$ so $\neg\mu a \notin H'$. So $\neg\mu a \notin F$, and because $\nu a \vee \neg\mu a \in F$ and F is prime we have $\nu a \in F$. So $a \in \nu^{-1}(F) \subseteq G$, showing that $\nu^{-1}(H') \subseteq G$. So part 4 is satisfied.

5. $\mathbf{A}_+$ is totally order-separated. For if $F \not\subseteq G$ then there is some $a \in A$ with $a \in F$ and $a \notin G$. So $F \in r_{\mathbf{A}}(a)$ and $G \notin r_{\mathbf{A}}(a)$.

6. $\mathbf{A}_+$ is compact. By a theorem known as Alexander's Lemma, we need only show that every subbasic cover of $X_{\mathbf{A}}$ has finite subcover, so first we verify this lemma. Suppose that X is a topological space with subbase $\mathbf{B}$ such that every subbasic cover of X has finite subcover, and let Ξ be an arbitrary cover of X. Suppose, for contradiction that Ξ has no finite subcover. Then for any finite n and $U_1, \ldots U_n \in \Xi$, we have $-U_1 \cap \ldots \cap -U_n \neq \emptyset$, and so $\{-U \mid U \in \Xi\}$ can be extended to a prime filter F - an ultrafilter - over X. Now suppose that there is no $x \in X$ such that for any open U with $x \in U$ we also have $U \in F$. Then for each $x \in X$ there is a basic open set W_x with $x \in W_x$ and $W_x \notin F$. Now for some $B_1, \ldots B_m \in \mathbf{B}$, $W_x = B_1 \cap \ldots \cap B_m$; also, $W_x \notin F$, $-W_x \cup W_x \in F$ and so because F is prime, $-W_x = -B_1 \cup \ldots \cup -B_m \in F$. Again using primeness, we have

$-B_i \in F$, say. But $x \in W_x = B_1 \cap \ldots \cap B_n$, so $x \in B_i$ and $B_i \notin F$ since $-B_i \in F$. This shows that for each $x \in X$ we may assume that W_x is in fact subbasic. So $\{W_x \mid x \in X\}$ is a subbasic cover of X with finite subcover, say $\{W_x \mid x \in K\}$. So $\bigvee\{W_x \mid x \in K\} = X \in F$, which is prime, so for some $x \in K$ we must have $W_x \in F$, contradicting the supposition about W_x: so there is indeed some $x \in X$ such that any open U with $x \in U$ has $U \in F$. Now Ξ covers X, so for this x there is some $U \in \Xi$ with $x \in U$: by definition of F we have $-U \in F$ but consideration of this x shows that $U \in F$. So their meet $\emptyset$ is also in F. But we assumed that F was proper, so Ξ does indeed have a finite subcover.

So let Ξ be such a subbasic cover of $X_{\mathbf{A}}$, and let $Y = \{a \mid -r_{\mathbf{A}}(a) \in \Xi\}$ and $Z = \{a \mid r_{\mathbf{A}}(a) \in \Xi\}$. Suppose that for finite $I \subseteq Y$ and $J \subseteq Z$ we have $\bigwedge I \leq \bigvee J$. Let $F \in \bigcap\{r_{\mathbf{A}}(a) \mid a \in I\}$; then $\forall a \in I(a \in F)$, so $\bigwedge I \in F$ because I is finite and F is a filter. Since $\bigwedge I \leq \bigvee J$, then $\bigvee J \in F$, and J is finite and F is prime so for some $b \in J$ we have $b \in F$; that is $F \in r_{\mathbf{A}}(b)$. So $F \in \bigcup\{r_{\mathbf{A}}(a) \mid a \in J\}$. This shows that for finite $I \subseteq Y$,$J \subseteq Z$,

$$\bigwedge I \leq \bigvee J \text{ implies } \bigcap\{r_{\mathbf{A}}(a) \mid a \in I\} \subseteq \bigcup\{r_{\mathbf{A}}(a) \mid a \in J\}.$$

Now if Ξ has no finite subcover, this means that for any $I \subseteq Y$, $J \subseteq Z$,

$$\bigcup\{-r_{\mathbf{A}}(a) \mid a \in I\} \cup \bigcup\{r_{\mathbf{A}}(a) \mid a \in J\} \neq X_{\mathbf{A}},$$

and so

$$\bigcap\{r_{\mathbf{A}}(a) \mid a \in I\} \not\subseteq \bigcup\{r_{\mathbf{A}}(a) \mid a \in J\}.$$

and so, by the above, $\bigwedge I \not\leq \bigvee J$. Applying the prime filter theorem, there is some $F \in X_{\mathbf{A}}$ with $Y \subseteq F$ and $F \cap Z = \emptyset$. Because $Y \subseteq F$, for any $a \in Y$ we have $F \in r_{\mathbf{A}}(a)$, and so $F \notin -r_{\mathbf{A}}(a)$. Because $F \cap Z = \emptyset$, for any $b \in Z$ we have $F \notin r_{\mathbf{A}}(b)$. But then this F is a counterexample to the fact that Ξ is a cover of $X_{\mathbf{A}}$. So Ξ does have a finite subcover. This, together with total order-separation, shows that $\mathbf{A}_+$ is a Priestley space. $\square$

This leaves the last two clauses in the definition of modal space to be verified.

Theorem 20 *$-R_\nu F$ is a union of clopen cocones.*

Proof. Let $G \in -R_\nu F$. Then not $FR_\nu G$ so $\nu^{-1}(F) \not\subseteq G$. Then for some $a \in A$, $\nu a \in F$ and $a \notin G$, and consequently $G \in -r_{\mathbf{A}}(a)$, which is a clopen cocone. Denote this clopen cocone by $\mathcal{N}_G$. But $\mathcal{N}_G = -r_{\mathbf{A}}(a)$ is disjoint from $R_\nu F$: for if $H \in -r_{\mathbf{A}}(a)$ then $a \notin H$ and so $\nu^{-1}(F) \not\subseteq H$; then not $FR_\nu H$ and so $H \notin R_\nu F$. So for any $G \in -R_\nu F$, we have $\{G\} \subseteq \mathcal{N}_G \subseteq -R_\nu F$. But $\mathcal{N}_G$ is a clopen cocone, and therefore $-R_\nu F = \bigcup\{\mathcal{N}_G \mid G \in -R_\nu F\}$ is a union of clopen cocones. □

To complete the proof that $\mathbf{A}_+$ is a modal space, it must still be shown that if U is a clopen cone then U^* and $\nu_R(U)$ are clopen. But first we show that $r_{\mathbf{A}}$ is an isomorphism, and so that every modal algebra can be regarded as the clopen cones of a modal space. That is, we show that $\mathbf{A} \cong (\mathbf{A}_+)^+$. In the course of this proof, the final property for modal spaces will be demonstrated to hold of $\mathbf{A}_+$. By definition of prime filter we have $r_{\mathbf{A}}(0) = \emptyset$ and $r_{\mathbf{A}}(1) = X_{\mathbf{A}}$. Furthermore, $r_{\mathbf{A}}(a \wedge b) = r_{\mathbf{A}}(a) \wedge r_{\mathbf{A}}(b)$ because the elements of $X_{\mathbf{A}}$ are filters, and $r_{\mathbf{A}}(a \vee b) = r_{\mathbf{A}}(a) \vee r_{\mathbf{A}}(b)$ because they are prime filters. We have $r_{\mathbf{A}}(\neg a) = \neg r_{\mathbf{A}}(a)$ because $F \in \neg r_{\mathbf{A}}(a)$ iff $F \notin (r_{\mathbf{A}}(a))^*$ iff $F^* \notin r_{\mathbf{A}}(a)$ iff $a \notin F^*$ iff $\neg a \in F$ iff $F \in r_{\mathbf{A}}(\neg a)$. And $r_{\mathbf{A}}(\nu a) = \nu_{R_\nu}(r_{\mathbf{A}}(a))$, because $F \in \nu_{R_\nu}(r_{\mathbf{A}}(a))$ iff $\forall G \in X_{\mathbf{A}}(FR_\nu G \Rightarrow G \in r_{\mathbf{A}}(a))$ iff $\forall G \in X_{\mathbf{A}}(\nu^{-1}(F) \subseteq G \Rightarrow a \in G)$ iff $a \in \nu^{-1}(G)$. This last step is by the prime filter theorem. One direction is obvious; but if $a \notin \nu^{-1}(F)$, then we can extend $\nu^{-1}(F)$ to a prime filter G with $a \notin G$. So we have an equivalence. But $a \in \nu^{-1}(F)$ iff $\nu a \in F$ iff $F \in r_{\mathbf{A}}(\nu a)$. The proof that $r_{\mathbf{A}}(\mu a) = \mu_{R'_\mu}(r_{\mathbf{A}}(a))$ is similar. So $r_{\mathbf{A}}$ is a homomorphism.

Now we show that $r_{\mathbf{A}}$ is injective. Suppose $a, b \in A$ and $a \neq b$ with $a \not\leq b$, say. Then by the prime filter theorem we have some $F \in X_{\mathbf{A}}$ with $a \in F$ and $b \notin F$. So $F \in r_{\mathbf{A}}(a)$ and $F \notin r_{\mathbf{A}}(b)$, so $r_{\mathbf{A}}(a) \neq r_{\mathbf{A}}(b)$.

Also $r_{\mathbf{A}}$ is surjective. The crucial point of the topology is to allow this part of the proof to go through. Let U be a clopen cone of $\mathbf{A}_+$ with $G \notin U$. Because

U is a cone, For any $F \in U$ we have $F \not\subseteq G$, and so for some $a_F \in A$ we have $a_F \in F$ and $a_F \notin G$. This means that $F \in r_{\mathbf{A}}(a_F)$ and $G \notin r_{\mathbf{A}}(a_F)$. Now $r_{\mathbf{A}}(a_F)$ is open, so $\{r_{\mathbf{A}}(a_F) \mid F \in U\}$ is an open cover of U. But U is closed, and therefore compact, so for some finite $J \subseteq U$, $\{r_{\mathbf{A}}(a_F) \mid F \in J\}$ covers U. Let $a_G = \bigvee\{a_F \mid F \in J\}$. Because G is prime, J is finite and for each a_F with $F \in J$ we have $a_F \notin G$, we may conclude that $a_G \notin G$; that is, $G \notin r_{\mathbf{A}}(a_G)$. But $U \subseteq r_{\mathbf{A}}(a_G)$: $\{r_{\mathbf{A}}(a_F) \mid F \in J\}$ covers U, and so if $a_F \in H \in U$, say, then $a_G \in H \in U$ because H is a filter. So $-U = \bigcup\{-r_{\mathbf{A}}(a_G) \mid G \in -U\}$. Because U is clopen, $-U$ is closed and therefore compact, which means that for some finite $K \subseteq -U$, $-U = \bigcup\{-r_{\mathbf{A}}(a_G) \mid G \in K\}$. So $U = \bigcap\{r_{\mathbf{A}}(a_G) \mid G \in K\}$ which is equal to $r_{\mathbf{A}}(\bigwedge\{a_G \mid G \in K\})$, because $r_{\mathbf{A}}$ is a homomorphism. Clearly $\bigwedge\{a_G \mid G \in K\} \in A$, so $r_{\mathbf{A}}$ is surjective. This completes the proof of

Theorem 21 $\mathbf{A} \cong (\mathbf{A}_+)^+$. □

Because $r_{\mathbf{A}}$ is surjective, if $U \in \mathbf{cl}(X)$ then $U = r_{\mathbf{A}}(a)$, say. But in showing that $r_{\mathbf{A}}$ is a homomorphism, we saw that $r_{\mathbf{A}}(\neg a) = \neg r_{\mathbf{A}}(a)$; so $-U^*$ is a clopen cone, being in the image of $r_{\mathbf{A}}$, so U^* is clopen. That $\nu_R(U)$ is clopen if U is also follows from the fact that $r_{\mathbf{A}}$ is a surjective homomorphism. So finally we have

Theorem 22 $\mathbf{A}_+$ *is a modal space.* □

A dual theorem can now be given which shows the equivalence of $\mathcal{C}$ and $(\mathcal{C}^+)_+$ for any modal space $\mathcal{C}$. First however, the notion of an *ordered frame morphism* and a *topological space morphism* must be defined. An ordered frame morphism is defined as a frame morphism except that the final clause for frame morphism is changed to

3. If $f(x)Sy$ then $\exists x' \in X$ with xRx' and $f(x') \leq y$;

and in addition we require that

$$x \leq y \Rightarrow f(x) \leq f(y).$$

This change to the definition of frame morphism is not in fact used in the following theorem, but it is the natural notion of ordered frame morphism, and we shall encounter it later.

A *continuous* mapping f between topologies is one such that if V is open then so is $f^{-1}(V)$. A *homeomorphism* is a injective and surjective continuous mapping f such that its inverse f^{-1} is also continuous: that is, if U is open then so is $f(U)$. This is a property that can be checked by considering only the subbases of the topology. A *modal space* morphism is then defined to be a continuous ordered frame morphism. So for $\mathcal{C} =< X, \leq, *, R, \mathcal{T} >$ and $x \in X$, let $r_{\mathcal{C}}(x) =_{def} \{Y \in \mathbf{cl}(X) \mid x \in Y\}$. Clearly $r_{\mathcal{C}}$ is a mapping because it is easily seen that $r_{\mathcal{C}}(x)$ is a prime filter of $\mathcal{C}^+$. Now the appropriate definition of isomorphism required here is that $r_{\mathcal{C}}$ is a homeomorphism, and both it and its inverse are ordered frame morphisms.

Theorem 23 *$\mathcal{C}$ is isomorphic to $(\mathcal{C}^+)_+$.*

Proof. The conditions to be shown are:

1. $r_{\mathcal{C}}$ is injective. This follows from the fact that $r_{\mathcal{C}}$ is injective as a partial ordering morphism. If $x \leq y$ and $x \in U \in \mathbf{cl}(X)$, then because U is a cone we have $y \in U$ and so $r_{\mathcal{C}}(x) \subseteq r_{\mathcal{C}}(y)$. If $x \not\leq y$ then total order-separation gives $U \in \mathbf{cl}(X)$ with $x \in U$ and $y \notin U$, and so $r_{\mathcal{C}}(x) \not\subseteq r_{\mathcal{C}}(y)$. So $x \leq y$ iff $r_{\mathcal{C}}(x) \subseteq r_{\mathcal{C}}(y)$. But then because $\leq$ is anti-symmetric if $x \neq y$ then $r_{\mathcal{C}}(x) \neq r_{\mathcal{C}}(y)$.

2. $r_{\mathcal{C}}$ is surjective. Again the key idea is to exploit the compactness of modal spaces. Let F be an element of $(\mathcal{C}^+)_+$, that is a prime filter of clopen cones of $\mathcal{C}$, and suppose that $r_{\mathcal{C}}$ is not surjective: that for all $x \in X$ we have $r_{\mathcal{C}}(x) \neq F$. This means that for all x we can find a clopen cone Y_x with either

$$(i)\ x \in Y_x \text{ and } Y_x \notin F \text{ or}$$

$$(ii)\ x \notin Y_x \text{ and } Y_x \in F.$$

Choosing such a Y_x for each $x \in X$, let

$$\Xi = \{Y_x \mid (i) \text{ holds }\} \bigcup \{-Y_x \mid (ii) \text{ holds }\}$$

Then Ξ is an open cover of X, and using the compactness of $\mathcal{C}$ it has a finite subcover, say

$$\{Y_{x_1}, \dots Y_{x_m}, -Y_{y_1}, \dots - Y_{y_n}\}$$

and so

$$\bigcap\{Y_{y_j} \mid 1 \leq j \leq n\} \subseteq \bigcup\{Y_{x_i} \mid 1 \leq i \leq m\}$$

But for each Y_{y_j}, we have $Y_{y_j} \in F$; and so because F is a filter, $\bigcap\{Y_{y_j} \mid 1 \leq j \leq n\} \in F$. And by the above inclusion, $\bigcup\{F_{x_i} \mid 1 \leq i \leq m\} \in F$. But F is prime, so for some $i \leq n$, $Y_{x_i} \in F$. But this contradicts the fact that Y_{x_i} satisfies (i). So $r_{\mathcal{C}}$ is an isomorphism of partial orderings.

3. $r_{\mathcal{C}}(x^*) = (r_{\mathcal{C}}(x))^*$. Let Y be a clopen cone in $\mathcal{C}$. Then $Y \in (r_{\mathcal{C}}(x))^*$ iff $\neg Y \notin r_{\mathcal{C}}(x)$ iff $-Y^* \notin r_{\mathcal{C}}(x)$ iff $x \notin -Y^*$ iff $x \in Y^*$ iff $x^* \in Y$ iff $Y \in r_{\mathcal{C}}(x^*)$.

4. Suppose xRy. Let Y be a clopen cone with $\nu_R(Y) \in r_{\mathcal{C}}(x)$: then $x \in \nu_R(Y)$ so $\forall z(xRz \Rightarrow z \in Y)$. In particular xRy so $y \in Y$. This shows that $Y \in r_{\mathcal{C}}(y)$; so $\nu_R^{-1}(r_{\mathcal{C}}(x)) \subseteq r_{\mathcal{C}}(y)$ which means that $r_{\mathcal{C}}(x) R_{\nu_R} r_{\mathcal{C}}(y)$.

5. Suppose not xRy. Then $y \notin Rx$ so $y \in -Rx$. But $-Rx$ is a unions of clopen cocones, so there is a clopen cocone Y with $y \in Y$ and $Rx \cap Y = \emptyset$. Consequently, $-Y \in \mathrm{cl}(X)$ and $Rx \subseteq -Y$; so $\forall y(xRy \Rightarrow y \in -Y)$, which means that $x \in \nu_R(-Y)$ and so $\nu_R(-Y) \in r_{\mathcal{C}}(x)$. But $y \in Y$ implies that $y \notin -Y$, and so $-Y \notin r_{\mathcal{C}}(y)$. So $-Y$ is a counterexample to $\nu_R^{-1}(r_{\mathcal{C}}(x)) \subseteq r_{\mathcal{C}}(y)$ and it has been shown that not $r_{\mathcal{C}}(x) R_{\nu_R} r_{\mathcal{C}}(y)$. So xRy iff $r_{\mathcal{C}}(x) R_{\nu_R} r_{\mathcal{C}}(y)$, which is sufficient to show that both $r_{\mathcal{C}}$ and its inverse satisfy the conditions imposed on the relations in an ordered frame morphism, so we have

an ordered frame isomorphism. The clause for R' may be shown similarly or may be deduced from the proofs above concerning R and $*$.

6. r_C is a homeomorphism. Let U be in the subbase of $(C^+)_+$: then recalling the definition of its subbase, U is either $r_{C^+}(Y)$ or $-r_{C^+}(Y)$ for some $Y \in \mathbf{cl}(X)$. Spelling out what this means, U is either $\{F \in X_{C^+} \mid Y \in F\}$ or its complement. Now we have shown earlier in this section that $C^+ \cong ((C^+)_+)^+$, so if U is of the form $\{F \in X_{C^+} \mid Y \in F\}$, then $r_C^{-1}(U) = Y$ which is open; and if U is of the form

$$-\{F \in X_{C^+} \mid Y \in F\} = \{F \in X_{C^+} \mid Y \notin F\}$$

then $r_C(x) \in U$ iff $Y \notin r_C(x)$ iff $x \notin Y$. So $r_C^{-1}(U) = -Y$. But $Y \in \mathbf{cl}(X)$, so $-Y$ is open. For the other direction use the fact that the clopen cones and their complements form a subbase for the the topology on C. Let U be a clopen cone in C: then

$$r_C(U) = \{r_C(x) \mid x \in U\} = \{r_C(x) \mid U \in r_C(x)\}.$$

Because r_C is surjective, every $F \in X_{C^+}$ is of the form $r_C(x)$ for some $x \in X$. So $r_C(U) = \{F \in X_{C^+} \mid U \in F\}$: the union of a set of elements belonging to the declared subbase of $(C^+)_+$, so $r_C(U)$ is open, being the union of open sets. Similarly, for U a clopen cone,

$$r_C(-U) = \{r_C(x) \mid x \notin U\} = \{r_C(x) \mid U \notin r_C(x)\} =$$

$$\{F \in X_{C^+} \mid U \notin F\} = -r_C(U),$$

which again is the union of a set of elements in the declared subbase, and so is open. So r_C is a homeomorphism. □

So any modal space can be represented as the prime filters of a modal algebra.

4.5 More About Duality

The operations $(\ )_+$ and $(\ )^+$ still make sense when only the underlying ordered frame or the underlying frame of a a modal space is considered. In defining $\mathbf{A}_+$ the topology, or both the topology and the ordering, are forgotten. And for $\mathbf{C}$ an ordered frame or a frame, $\mathbf{C}^+$ is defined by taking its universe to be $\mathbf{c}(X)$, the cones of the universe X of $\mathbf{C}$. If $\leq$ is discrete, and so $\mathbf{C}$ is a frame, then the universe turns out to be $\mathcal{P}(X)$, the powerset of the universe of $\mathbf{C}$. Clearly by the foregoing proofs, $\mathbf{c}(X)$ is closed under the algebraic operations defined for $\mathcal{C}^+$ with $\mathcal{C}$ a modal space, and so $\mathbf{C}^+$ is in each of these cases a modal algebra.

The definitions of $(\ )_+$ and $(\ )^+$ are now extended to algebraic homomorphisms and ordered frame morphisms respectively.

Theorem 24 *For ordered frames* $\mathbf{C}_1 =< X, \leq, *, R >$ *and* $\mathbf{C}_2 =< Y, \leq, *, S >$ *with* $\mathbf{C}_1 \xrightarrow{f} \mathbf{C}_2$ *and* $Z \in \mathbf{c}(Y)$, *define* $f^+(Z) = f^{-1}(Z) = \{x \in X \mid f(x) \in Z\}$. *Then* $\mathbf{C}_2^+ \xrightarrow{f^+} \mathbf{C}_1^+$ *is a homomorphism.*

Proof. It is well known that f^+ is a mapping and that it is a bounded distributive lattice homomorphism. So we need only check that f^+ preserves $\neg$ and ν_S. First $f^{-1}(\neg Z) = \neg f^{-1}(Z)$, for $x \in f^{-1}(\neg Z)$ iff $x \in f^{-1}(-Z^*)$ iff $f(x) \in -Z^*$ iff $f(x) \notin Z^*$ iff $f(x^*) = f(x)^* \notin Z$ iff $x^* \notin f^{-1}(Z)$ iff $x \notin (f^{-1}(Z))^*$ iff $x \in -(f^{-1}(Z))^* = \neg f^+(Z)$. And also $f^+(\nu_S Z) = \nu_R f^+(Z)$: for suppose $x \in f^{-1}(\nu_S Z)$ and xRy. Because f is an ordered frame morphism, $f(x)Sf(y)$; and $x \in f^{-1}(\nu_S Z)$ implies that $\forall z \in Y(f(x)Sy \Rightarrow z \in Z)$, so $f(y) \in Z$ and $y \in f^{-1}(Z)$. y was arbitrary, so $x \in \nu_R f^{-1}(Z)$. Conversely, let $x \in \nu_R f^{-1}(Z)$ and $f(x)Sy$. Then $\exists w$ with xRw and $f(w) \leq y$. But $x \in \nu_R f^{-1}(Z)$ implies that $w \in f^{-1}(Z)$, so $f(w) \in Z$ which is a cone so $y \in Z$. This show that $\forall y(f(x)Sy \Rightarrow y \in Z)$, so $f(x) \in \nu_S(Z)$ and $x \in f^{-1}(\nu_S Z)$. □

Theorem 25 *If f is surjective then f^+ is injective.*

Proof. Let $Z, Z' \in \mathbf{c}(Y)$ with, say, $y \in Z$, $y \notin Z'$. Because f is surjective there is some $x \in X$ with $f(x) = y$. So $x \in f^{-1}(Z)$ and $x \notin f^{-1}(Z')$. So $Z \neq Z'$ implies that $f^{+}(Z) \neq f^{+}(Z')$. $\square$

Now let $\mathbf{A}_1 \xrightarrow{h} \mathbf{A}_2$ be a homomorphism between modal algebras. For $F \in X_{\mathbf{A}_2}$ a prime filter of $\mathbf{A}_2$, define $h_+(F) = h^{-1}(F) = \{x \in A \mid h(x) \in F\}$. Then

Theorem 26 $\mathbf{A}_{2+} \xrightarrow{h_+} \mathbf{A}_{1+}$ *is an ordered frame morphism.*

Proof. We check:

1. h_+ is a mapping: $F \in X_{\mathbf{A}2}$ is prime and $h(a \vee b) = h(a) \vee h(b)$, so $a \vee b \in h^{-1}(F)$ iff $a \in h^{-1}(F)$ or $b \in h^{-1}(F)$. Also, $1 \in h^{-1}(1)$ and $1 \in F$, and $h(0) = 0$ and $0 \notin F$, so $h^{-1}(F)$ is a prime filter.

2. $F \subseteq G \Rightarrow h_+(F) \subseteq h_+(G)$ is obvious: if $h(a) \in F$ and $F \subseteq G$ then $h(a) \in G$.

3. $h_+(F^*) = (h_+(F))^*$. $a \in h^{-1}(F^*)$ iff $h(a) \in \{\neg b \mid b \notin F\}$ iff $\neg h(a) \notin F$ iff $h(\neg a) \notin F$ iff $a \in \{\neg a \mid h(a) \notin F\}$, which is $\{\neg a \mid a \notin h^{-1}(F)\} = (h^{-1}(F))^*$.

4. If $FR_{\nu_2}G$ then $h_+(F)R_{\nu_1}h_+(G)$. Let $\nu_2^{-1}(F) \subseteq G$ and suppose that $a \in \nu_1^{-1}(h^{-1}(F))$. Then $\nu_1 a \in h^{-1}(F)$ so $\nu_2 h(a) = h(\nu_1 a) \in F$. This means that $h(a) \in \nu_2^{-1}(F)$ so $h(a) \in G$ and $a \in h^{-1}(G)$. So we have shown that $\nu_1^{-1}(h_+(F)) \subseteq h_+(G)$, as required.

5. If $h_+(F)R_{\nu_1}H$ then there is a $G \in X_{\mathbf{A}_2}$ with $FR_{\nu_2}G$ and $h_+(G) \subseteq H$. Consider $h(-H)$. $h(-H)$ is closed under finite joins, because if $a, b \in -H$ then $a \vee b \in -H$ because H is prime, so $h(a \vee b) = h(a) \vee h(b) \in h(-H)$. Moreover $\nu_2^{-1}(F)$ is closed under finite meets because $\nu_2(a \wedge b) = \nu_2(a) \wedge \nu_2(b)$. Since $\nu_2(1) \in F$ it is in fact a filter. Let I and J be finite subsets of $\nu_2^{-1}(F)$ and $h(-H)$ respectively, and suppose that $\bigwedge I \leq \bigvee J$. Then

$\bigwedge I \in \nu_2^{-1}(F)$ and $\bigvee J \in h(-H)$; also $\bigvee J \in \nu_2^{-1}(F)$ because $\nu_2^{-1}(F)$ is a filter. Now since $h_+(F)R_{\nu_1}H$ this means that if $h(\nu_1 b) \in F$ then $b \in H$; so if $b \notin H$ then $h(\nu_1 b) = \nu_2 h(b) \notin F$. For some $b' \in -H$ we have $h(b') = \bigvee J$, and so by the above $\nu_2 \bigvee J \notin F$, that is $\bigvee J \notin \nu_2^{-1}(F)$, which is a contradiction. Given that there are no such finite subsets I and J, we can apply the prime filter theorem and extend $\nu_2^{-1}(F)$ to a prime filter G which is disjoint from $h(-H)$. Since $\nu_2^{-1}(F) \subseteq G$ we have $FR_{\nu_2}G$; and if $h(a) \in G$ then $h(a) \notin h(-H)$, so $a \notin -H$ and $a \in H$. This shows that $h^{-1}(G) = h_+(G) \subseteq H$ as required, completing the proof that h_+ is a frame morphism. □

Theorem 27 *If* $\mathbf{A}_1 \xrightarrow{h} \mathbf{A}_2$ *is injective then* $\mathbf{A}_{2+} \xrightarrow{h_+} \mathbf{A}_{1+}$ *is surjective.*

Proof. Let F be a prime filter of $\mathbf{A}_1$. Because h is injective $h(F) \cap h(-F) = \emptyset$. As in the previous proof it can be checked that the conditions of the prime filter theorem are satisfied for $h(F)$ and $h(-F)$, and so h(F) can be extended to a prime filter G of $\mathbf{A}_2$ disjoint from $h(-F)$. This disjointness show that $h_+(G) = F$. □

Theorem 28 *If* $\mathbf{A}_1 \xrightarrow{h} \mathbf{A}_2$ *is surjective then* $\mathbf{A}_{2+} \xrightarrow{h_+} \mathbf{A}_{1+}$ *is injective.*

Proof. If F and G are distinct prime filters of $\mathbf{A}_2$, say with $b \in F$, $b \notin G$, then because h is surjective we have some $a \in A$ with $h(a) = b$. So $h(a) \in F$ and $h(a) \notin G$; that is $h_+(F) \neq h_+(G)$. □

There is one obvious omission here: it has not been shown that if f is an injective ordered frame morphism then f^+ is surjective. This is in fact false, and a counterexample will be given in the next chapter[3]. A property of f which is stronger than injectivity, that $x \leq y$ iff $f(x) \leq f(y)$, ensures that f^+ is surjective, however. First, note that

[3]And so theorem 2.3.1(3) of Goldblatt 1989 is wrong. As another counterexample, let the frames $\mathbf{C}_i$ for $i \in \{1, 2\}$ have common universe $\{x, x^*\}$, empty relations and $\leq$

Theorem 29 *If* $\mathbf{A}_1 \xrightarrow{h} \mathbf{A}_2$ *is surjective, then* $x \leq y$ *iff* $h_+(x) \leq h_+(y)$.

Proof. With F and G prime filters of $\mathbf{A}_2$ and $F \not\subseteq G$, let $b \in F$ and $b \notin G$. Then because h is surjective there is some $a \in A_1$ with $h(a) = b$. So $h(a) \in F$ and $h(a) \notin G$, implying that $h^{-1}(F) \not\subseteq h^{-1}(G)$. The other direction is obvious: if $h^{-1}(F) \not\subseteq h^{-1}(G)$, because h is surjective we have some $a \in A_1$ with $h(a) \in F$ and $h(a) \notin G$, so $F \not\subseteq G$. $\square$

Theorem 30 *Let* $\mathbf{C}_1 \xrightarrow{f} \mathbf{C}_2$ *be such that* $x \leq y$ *iff* $f(x) \leq f(y)$. *Then* f^+ *is surjective.*

Proof. Let U be a cone of $\mathbf{C}_1$ and let V be the least cone of $\mathbf{C}_2$ containing $\{f(x) \mid x \in U\}$: that is, $y \in V$ iff $\exists x \in U\ f(x) \leq y$. Now $f^+(V) = \{x \in C_1 \mid f(x) \in V\}$: we show that $f^+(V) = U$. First suppose that $x \in U$. Then $f(x) \leq f(x)$, so $f(x) \in V$ and $x \in f^+(V)$. Next let $x \in f^+(V)$. So $f(x) \in V$ and there is some $x' \in U$ with $f(x') \leq f(x)$. But then by hypothesis $x' \leq x$, and U is a cone so $x \in U$. $\square$

So in particular this is true for (discrete) frames. As a corollary of these theorems, in the area of ordered frames we have

Corollary 31 *If* h *is a surjective (injective) homomorphism, then so is* $(h_+)^+$.
$\square$

Now if the topology is put back on $\mathbf{A}_{1+}$ and $\mathbf{A}_{2+}$ it can be shown that h_+ is a continuous ordered frame morphism. All that is left to check is the following:

---

discrete on $\mathbf{C}_1$ but with in addition $x \leq x^*$ on $\mathbf{C}_2$. Then where $\mathbf{C}_1 \xrightarrow{f} \mathbf{C}_2$ is the identity mapping on the universe, f is an injective ordered frame morphism, but $\{x\} \in \mathbf{C}_1^+$ is a cone not in the image of f^+.

Theorem 32 *If U is a subbasic open set in $\mathbf{A}_{1+}$, then $h_+^{-1}(U)$ is open in $\mathbf{A}_{2+}$.*

We know that U is of the form $r_{\mathbf{A}1}(a)$ or $-r_{\mathbf{A}1}(a)$, for some $a \in A$. In the first case

$$h_+^{-1}(r_{\mathbf{A}_1}(a)) = \{G \in X_{\mathbf{A}_2} \mid h_+(G) \in r_{\mathbf{A}_1}(a)\} = \{G \in X_{\mathbf{A}_2} \mid a \in h^{-1}(G)\}$$

but

$$\{G \in X_{\mathbf{A}_2} \mid a \in h^{-1}(G)\} = \{G \in X_{\mathbf{A}_2} \mid h(a) \in G\} = r_{\mathbf{A}_2}(h(a))$$

which belongs to the subbase of $\mathbf{A}_2$ and so is open. On the other hand,

$$h_+^{-1}(-r_{\mathbf{A}_1}(a)) = \{G \in X_{\mathbf{A}_2} \mid h_+(G) \notin r_{\mathbf{A}_1}(a)\} =$$

$$\{G \in X_{\mathbf{A}_2} \mid h(a) \notin G\} = -r_{\mathbf{A}_2}(h(a)),$$

which again is open, being in the subbase. So h_+ is a continuous mapping. □

The formal correspondence between modal algebras and modal spaces is expressed in the following, rather informal, way. A *category* is a set O of objects together with a set of arrows, each of which has a specified domain and codomain in O. These arrows are closed under composition, denoted $\circ$, and for each object $a \in O$ the identity arrow 1_a from a to itself exists. Composition and identity are required to have the usual properties.

A *functor* $\mathcal{F}$ from a category $\mathbf{C}$ to a category $\mathbf{D}$ maps objects to objects and arrows to arrows in such a way that for objects a and composible arrows $b \xrightarrow{f} c \xrightarrow{g} d$ we have $\mathcal{F}(1_a) = 1_{\mathcal{F}(a)}$ and $\mathcal{F}(g \circ f) = \mathcal{F}(g) \circ \mathcal{F}(f)$. In particular, the *identity functor* on a category maps all objects and arrows to themselves. Where $\mathbf{D}^{op}$ is the category with the same objects as $\mathbf{D}$ but with the directions of the arrows reversed, a *contravariant functor* from $\mathbf{C}$ to $\mathbf{D}$ is a functor from $\mathbf{C}$ to $\mathbf{D}^{op}$. It is obvious that $\mathcal{MS}$, modal spaces with continous ordered frame morphisms, and $\mathcal{MA}$, modal algebras with homomorphisms are categories, and it has been shown above that $(\ )^+$ and $(\ )_+$ are contravariant functors in their respective directions between them. It can now be seen that $((\,)_+)^+$ and $((\,)^+)_+$ are in a sense the same things as the identity functors on the categories $\mathcal{MA}$ and $\mathcal{MS}$. If, with MacLane[4], we regard a functor of giving a picture of one category in another, then a comparison of, say $1_{\mathcal{MA}}$ - the identity functor on $\mathcal{MA}$ - and $((\,)_+)^+$ reveals that they give the same picture of $\mathcal{MA}$ in $\mathcal{MA}$. We have seen that this is true of the objects of the two categories: for any $\mathbf{A} \in \mathcal{MA}$ and any $\mathcal{C} \in \mathcal{MS}$ we have isomorphisms $r_{\mathbf{A}} : (\mathbf{A}) \cong (\mathbf{A}_+)^+$ and $r_{\mathcal{C}} : (\mathcal{C}) \cong (\mathcal{C}^+)_+$. As for the morphisms, what is left to be shown technically for any g with $\mathbf{A} \xrightarrow{g} \mathbf{B}$ to give the corresponding notion of being to all intents and purposes 'the same', is that $r_{\mathbf{B}} \circ g = (g_+)^+ \circ r_{\mathbf{A}}$. But this has been demonstrated above:

$$(g_+)^+ \circ r_{\mathbf{A}}(a) = (g_+)^+(\{F \in X_{\mathbf{A}} \mid a \in F\}) = \{G \in X_{\mathbf{B}} \mid a \in g_+(G)\} =$$

$$\{G \in X_{\mathbf{B}} \mid g(a) \in G\} = r_{\mathbf{B}}(g(a)) = r_{\mathbf{B}} \circ g(a).$$

And for modal spaces it must be shown that if $\mathcal{C} \xrightarrow{f} \mathcal{D}$, then $r_{\mathcal{D}} \circ f = (f^+)_+ \circ r_{\mathcal{C}}$. And again, where X are the worlds of $\mathcal{C}$ and Y are the worlds of $\mathcal{D}$,

$$(f^+)_+ \circ r_{\mathcal{C}}(x) = (f^+)_+(\{U \in \mathbf{cl}(X) \mid x \in U\}) = \{V \in \mathbf{cl}(Y) \mid x \in f^+(V)\} =$$

$$\{V \in \mathbf{cl}(Y) \mid f(x) \in V\} = r_{\mathcal{D}}(f(x)) = r_{\mathcal{D}} \circ f(x).$$

At the level of frames, or even ordered frames, where there is no topology to guarantee that $r_{\mathbf{A}}$ and $r_{\mathcal{C}}$ are surjective, there is no such theorem. But $(\,)_+$ and $(\,)^+$ are very useful constructions. We concentrate on frames.

Theorem 33 *If* $\mathbf{A} \not\models \alpha$ *then* $\mathbf{A}_+ \not\models \alpha$.

Proof. If $\mathbf{A} \not\models t^\alpha = 1$ because for $a_1, \ldots a_n \in A$ and the assignment $p_i \mapsto a_i$, $p^\alpha(a_1, \ldots a_n) \neq 1$, then the frame valuation $v(p_i) = r_{\mathbf{A}}(a_i)$ on $\mathbf{A}_+$ is such that $v(\alpha) \neq X_{\mathbf{A}+}$. □

[4]MacLane 1971.

Theorem 34 *For* $\mathbf{C}$ *a frame,* $\mathbf{C} \models \alpha$ *iff* $\mathbf{C}^+ \models \alpha$.

Proof. If $\mathbf{C}, v, x \not\models \alpha$, then v may be regarded as an assignment of values to the formulae in the universe $\mathcal{P}(X)$ of $\mathbf{C}^+$. Given how the operations on $\mathbf{C}^+$ were defined, this assignment to the arguments of the polynomial p^α is not equal to 1. The converse is also obvious: $\mathbf{C}^+ \not\models \alpha$, then the counterexample assignment to the universe $\mathcal{P}(X)$ of $\mathbf{C}$ is a frame valuation v for $\mathbf{C}$ with $v(\alpha) \neq 1$. □

This shows that $\mathbf{C}$ and $\mathbf{C}^+$ have the same logic.

Corollary 35 *The class of all frames has the same logic as the class of all modal algebras, namely* $\vdash_K$.

Proof. By the above equivalence, or by noting that $L(\mathbf{A}) = L(S(\mathbf{A}))$ and every algebra is a subalgebra of a powerset algebra $\mathbf{C}^+$: let $\mathbf{C}$ be the unordered frame $\mathbf{A}_+$. The proof that $r_{\mathbf{A}}$ is an injective homomorphism, for $\mathbf{A}_+$ a modal space, does not depend on the topology or the ordering. □

Just as in classical modal logic there is a distinction between the more natural model of a Kripke frame and the notion of a *general frame*, which has a tighter connection to the logic, here a three stage gradation emerges. Modal spaces, of course correspond precisely to the algebraic models; and ordered frames have a looser correspondence. $((\ )_+)^+$ and $((\ )^+)_+$ no longer result in isomorphisms $r_{\mathbf{A}}$ and $r_{\mathcal{C}}$ on objects, but they are still contravariant functors. The ordering, however, serves no purpose other that a technical one: it is there because the prime filters of a modal algebra need not be discrete under the inclusion ordering, and this relation on prime filters was needed in some of the previous proofs. It has, however, little to do with the intended purpose of frames as model structures for the given modal language. For that reason I find it more natural to deal only with discretely ordered frames, or simply frames. This move naturally loses some of the foregoing theorems, but the main motivation is one of naturalness.

First of all, we maintain all of the results proved for ordered frames involving $(\ )^+$, since these did not involve the ordering: this is still a functor from frames to algebras. And $\mathbf{A}_+$ is still a frame. But $(\ )_+$ is not a functor, because for a homomorphism h, h_+ need not be a frame morphism: this requires that

$$\text{If } f(x)Sy \text{ then } \exists x' \in X \text{ with } xRx' \text{ and } f(x') = y$$

whereas we could only show that h_+ satisfied

$$\text{If } h_+(x)Sy \text{ then } \exists x' \in X \text{ with } xRx' \text{ and } h_+(x') \leq y.$$

In general, the prime filters of a modal algebra are not discretely ordered: consider, for example any modal algebra based on the three element chain $\underline{\mathbf{3}}$ with $0 < a < 1$ and $\neg a = a$: say, with $\nu a = 0$. But if they are discretely ordered on the codomain of h_+, for example for $\mathbf{C}^+ \xrightarrow{h} \mathbf{B}$ with C a discrete frame, then h_+ is indeed an ordered frame morphism for the two above requirements are then equivalent. This particular example works because let F, G be prime in $\mathbf{C}^+$ with $F \subset G$, $a \in G$ and $a \notin F$. Since this is a complemented algebra, $-a$ exists and $a \vee -a \in F$ which is prime and does not contain a, so $-a \in F \subset G$. But then $a \wedge -a = 0 \in G$, which is not the case. So the prime filters are discretely ordered. Another exception is the case where $\mathbf{A} \xrightarrow{h} \mathbf{B}$ is injective, when again h_+ is a frame morphism. Consider again part 5 of theorem 26: then because $h(H) \cap h(-H) = \emptyset$ it is readily seen that these two subsets of B satisfy the premise of the prime filter theorem - regard $\mathbf{A}$ as a subalgebra of $\mathbf{B}$, then this follows from the fact that H and $-H$ in $\mathbf{A}$ clearly satisfy this condition.

Both complemented algebras and powerset algebras $\mathbf{C}^+$ will play an important part in the study of four-valued modal logics. This is partly due to the fact that the properties of $(\ )_+$ and $(\ )^+$ that still hold for frames are sufficient to characterise various properties of classes of frames in terms of properties of varieties of modal algebras.

Chapter 5

Incomplete Logics

5.1 Preliminary Results And Definitions

The lattice of modal logics was seen to be anti-isomorphic to the lattice of varieties of modal algebras. The mappings used in showing this can be meaningfully transferred to classes of frames in order to see how the frame-based semantics fares in modelling logics. So for a frame $\mathbf{C}$ define the logic of $\mathbf{C}$ $L(\mathbf{C})$ to be $\{\alpha \mid \mathbf{C} \models \alpha\}$, and for a class of frames K define $L(K) = \bigcap\{L(\mathbf{C}) \mid \mathbf{C} \in K\}$. These are logics, given that all frames are models for $\vdash_K$ and logics are closed under intersection. Conversely, say that a frame $\mathbf{C}$ is a frame for the logic L if all theorems of L are valid in $\mathbf{C}$: then let $F(L)$ be the class of all such frames. In showing completeness for modal algebras it was seen that $L = L(V(L))$ for all logics L, and we have already seen that $\vdash_K = L(F(\vdash_K))$. When, for a modal logic L we have $L = L(F(L))$, then L is a *complete* logic; otherwise it is *incomplete*. Eventually we shall exhibit many incomplete non-classical modal logics. Incomplete logics are logics which are not determined by a class of frames: for if $L = L(F(L))$, then L is determined by the class $F(L)$. For the reverse implication, note that we always have $L \subseteq L(F(L))$; and if for a class of frames K we have $L = L(K)$ and $\alpha \notin L$, then for some $\mathbf{C} \in K$, $\mathbf{C} \not\models \alpha$. But clearly $K \subseteq F(L(K))$, so $\mathbf{C} \in F(L(K)) = F(L)$ and thus $\alpha \notin L(F(L))$. As an example the reader may wish to verify that the two distinct logics determined by the axioms $\alpha \vee \Box\neg\Box\alpha$ and $\alpha \vee \bigcirc\neg\bigcirc\alpha$ are both true in precisely the same, easily defined class of frames.

Now for a variety V of modal algebras, let $F(V)$ be the class of frames $\{\mathbf{C} \mid \mathbf{C}^+ \in V\}$. Then because $\mathbf{C}$ and $\mathbf{C}^+$ have the same logics, $F(V(L)) = F(L)$: that is, if $V = V(L)$ is the class of all L-algebras, then $F(V)$ is the class of all L-frames. Now there is an algebraic definition of completeness: a logic L is complete iff $L = L(F(L))$, but because $\mathbf{C}$ and $\mathbf{C}^+$ have the same logics, $L(F(L)) = L(F(L)^+)$, and so by Birkhoff's theorem $L(F(L)) = L(V(F(L)^+))$. So L is complete iff $L = L(V(F(L)^+))$. But because V is an anti-isomorphism from logics to varieties of algebras, this is equivalent to $V(L) = V(L(V(F(L)^+)))$: by algebraic completeness this equation becomes $V(L) = V(F(L)^+)$. So L is complete iff $F(L)^+$ generates the variety $V(L)$. So a *variety* can be defined to be complete if $V = V(F(V)^+)$: this is natural, because then we have L is complete iff $V(L)$ is complete.

On the way to another characterisation of completeness we can now collect some properties of $F(V)$, for V a variety, which hold in general for non-classical discrete frames.

Theorem 36 *Let $\{\mathbf{C}_i \mid i \in I\}$ be a class of frames. Then $(\sum_{i\in I} \mathbf{C}_i)^+ \cong \prod_{i\in I} \mathbf{C}_i^+$.*

Proof. Each $\mathbf{C}_j \xrightarrow{f_j} \sum_{i\in I} \mathbf{C}_i$ is injective, so each $(\sum_{i\in I} \mathbf{C}_i)^+ \xrightarrow{f_j^+} \mathbf{C}_j^+$ is surjective, with $f_j^+(Y) = \{x \mid (x,j) \in Y\}$. Now define $(\sum_{i\in I} \mathbf{C}_i)^+ \xrightarrow{f} \prod_{i\in I} \mathbf{C}_i^+$ by $f(Y)(j) = f_j^+(Y) = \{x \mid (x,j) \in Y\}$. Then f is surjective because if $W \in \prod_{i\in I} \mathbf{C}_i^+$ with $W(j) = W_j \in \mathbf{C}_j$, then $W = f(\bigcup_{i\in I} f_i^+(W_i))$: that is, index the elements of W then take their union - then f removes the indices. And f is injective because if $f(V) = f(W)$, then $\forall i \in I(f(V)(i) = f(W)(i))$, so $\forall i \in I(\{x \mid (x,i) \in V\} = \{x \mid (x,i) \in W\})$; and so $(x,i) \in V$ iff $(x,i) \in W$, so $V = W$. $\Box$

Given the appropriate definition of disjoint union for ordered frames, it is clear that f_j has the property $x \leq y$ iff $f_j(x) \leq f_j(y)$, and so this proof could be extended to ordered frames.

Theorem 37 *Let V be a variety. Then $F(V)$ is closed under subframes, p-morphic images, disjoint unions, and if $(\mathbf{C}^+)_+ \in F(V)$ then $\mathbf{C} \in F(V)$. In other words, we now restate the three preservation theorems given in the previous chapter and add the promised fourth one.*

Proof. In turn:

1. If $\mathbf{C}^+ \in V$ and $\mathbf{D} \xrightarrow{f} \mathbf{C}$ is an injective frame morphism, then $\mathbf{C}^+ \xrightarrow{f^+} \mathbf{D}^+$ is a surjective homomorphism. But V is a variety and so is closed under homomorphic images, so $\mathbf{D}^+ \in V$ and $\mathbf{D} \in F(V)$.

2. If $\mathbf{C}^+ \in V$ and $\mathbf{C} \xrightarrow{f} \mathbf{D}$ is an surjective frame morphism, then $\mathbf{D}^+ \xrightarrow{f^+} \mathbf{C}^+$ is a injective homomorphism. But V is a variety and so is closed under subalgebras, so $\mathbf{D}^+ \in V$ and $\mathbf{D} \in F(V)$.

3. If $\{\mathbf{C}_i^+ \mid i \in I\} \subseteq V$, then because V is closed under products, $(\sum_{i\in I} \mathbf{C}_i)^+ \cong \prod_{i\in I} \mathbf{C}_i^+ \in V$, so $\sum_{i\in I} \mathbf{C}_i \in F(V)$.

4. If $(\mathbf{C}^+)_+ \in F(V)$ then $((\mathbf{C}^+)_+)^+ \in V$. But $r_{\mathbf{C}^+} : \mathbf{C}^+ \to ((\mathbf{C}^+)_+)^+$ is injective and V is closed under subalgebras, so $\mathbf{C}^+ \in V$ and $\mathbf{C} \in F(V)$. □

Recall some definitions and theorems from universal algebra which will be useful in the exploration of modal logics. An algebra $\mathbf{A}$ is a *subdirect product* of the algebras $\{\mathbf{A}_i \mid i \in I\}$ if there is an injective $\mathbf{A} \xrightarrow{h} \prod_{i\in I} \mathbf{A}_i$ such that $\mathbf{A} \xrightarrow{\pi_i \circ h} \mathbf{A}_i$ is surjective for all $i \in I$, where $\prod_{i\in I} \mathbf{A}_i \xrightarrow{\pi_i} \mathbf{A}_i$ is the projection map with $\pi_i(a) = a(i)$. h is then called a *subdirect embedding.* An algebra $\mathbf{A}$ is *subdirectly irreducible* if for every subdirect embedding $\mathbf{A} \xrightarrow{h} \prod_{i\in I} \mathbf{A}_i$ there is some $j \in I$ with $\pi_j \circ h$ and isomorphism. Now it is known that every algebra is a subdirect product of subdirectly irreducible algebras, so these will be important in looking at varieties of modal algebras. For this reason we give a useful equivalent characterisation of subdirect irreducibility. A *congruence* on an algebra $\mathbf{A}$ is an equivalence relation θ on its universe on its universe A such that for any n-ary operator ω of $\mathbf{A}$, if $a_i\theta b_i$, for $1 \leq i \leq n$ and $a_i, b_i \in A$ then $\omega(a_1, \ldots a_n)\theta\omega(b_1, \ldots b_n)$. Under inclusion, the set $\mathbf{Con}(\mathbf{A})$ of all congruences on $\mathbf{A}$ form a lattice; indeed, if $\mathbf{A}$ is a modal algebra, then $\mathbf{Con}(\mathbf{A})$ is a distributive lattice. This follows from the fact that the lattice of congruences on a lattice is distributive, and modal algebra congruences are also lattice congruences. There are two distinguished congruences on any algebra: Δ is such that $a\Delta b$ iff $a = b$; and ∇ is such that for all $a, b \in A$, $a\nabla b$.

We now state the equivalent formulation of subdirect irreducibility.

Theorem 38 $\mathbf{A}$ *is subdirectly irreducible iff the set* $\mathbf{Con}(\mathbf{A})\backslash\{\Delta\}$ *has a least element.* □

If K is a class of modal algebras let K_{SI} denote the subdirectly irreducible algebras in K. Then because every algebra is a subdirect product of subdirectly irreducible algebras, every variety is generated by its subdirectly irreducible algebras: $V = V(V_{SI})$.

For $\{\mathbf{A}_i \mid i \in I\}$ a family of algebras, an *ultrafilter U on I* is a prime filter of $\mathcal{P}(I)$. Then U defines a congruence θ^U on $\prod_{i\in I} \mathbf{A}_i$ given by $a\theta^U b$ iff $\{i \in I \mid a(i) = b(i)\} \in U$. The quotient of $\prod_{i\in I} \mathbf{A}_i$ by θ^U, with elements the θ^U- equivalence classes is written $\prod_{i\in I} \mathbf{A}_i/U$ and is called an *ultraproduct* of $\{\mathbf{A}_i \mid i \in I\}$. If for each $i \in I$, $\mathbf{A}_i = \mathbf{A}$ we sometimes write instead $\mathbf{A}^I/U$ and call it an *ultrapower* of $\mathbf{A}$. For a class of algebras K, $P_U(K)$ is the class of all ultraproducts of algebras in K. A theorem of Loś of use here is that if a sentence in the first order language adequate to talk about algebras is true of all $\{\mathbf{A}_i \mid i \in I\}$ then is is true of any ultraproduct $\prod_{i\in I} \mathbf{A}_i/U$.

Next are presented without proof some useful theorems which apply to our modal algebras. It was noted previously that for a class of algebras K, $V(K) = HSP(K)$; now given that all modal algebras have distributive congruence lattices, we have

Theorem 39 (Jónsson) *If a modal algebra* $\mathbf{A}$ *is subdirectly irreducible in* $V(K)$*, then* $\mathbf{A} \in HSP_U(K)$. $\square$

But given that we also have

Theorem 40 *If* $\prod_{i \in I} \mathbf{A}_i/U$ *is an ultraproduct with* $\mathbf{A}_i \in \{\mathbf{B}_1, \ldots \mathbf{B}_n\}$ *for each* $i \in I$*, and for each* j*,* $1 \leq j \leq n$ *the universe of* $\mathbf{B}_j$ *finite, then for some* j*,* $1 \leq j \leq n$*,* $\prod_{i \in I} \mathbf{A}_i/U \cong \mathbf{B}_j$*,* $\square$

then

Corollary 41 *If* $\mathbf{A}$ *is subdirectly irreducible in* $V(\{\mathbf{A}_1, \ldots \mathbf{A}_n\})$ *with each* $\mathbf{A}_i$ *finite, then* $\mathbf{A} \in HS(\{\mathbf{A}_1, \ldots \mathbf{A}_n\})$. $\square$

Corollary 42 *For* $\mathbf{A}_1, \mathbf{A}_2$ *finite subdirectly irreducible algebras,* $V(\mathbf{A}_1) = V(\mathbf{A}_2)$ *iff* $\mathbf{A}_1 \cong \mathbf{A}_2$. $\square$

Corollary 43 *If* V *and* V' *are varieties of modal algebras and* $V \vee V'$ *is their sum in the lattice of varieties, then* $(V \vee V')_{SI} = V_{SI} \cup V'_{SI}$. $\square$

To illustrate some of these concepts, we now prove a claim of the previous chapter and give an example of an injective ordered frame morphism f such that f^+ is not surjective. In fact we show more: a morphism g is *epi* if for any h, h' with $h \circ g = h' \circ g$ we have $h = h'$. Then surjective homomorphisms are all epi. However, we have

Theorem 44 *It is possible that an ordered frame morphism* f *is injective but* f^+ *is not even epi.*

Proof. We use a frame $\mathbf{C}_1$ which does not admit a compact topology: let $\mathbf{C}_1$ be the frame with universe $C = \{\ldots -3, -2, -1, 1, 2, 3 \ldots\}$, $n^* = -n$, the empty relations and the discrete ordering - $\leq$ coincides with $=$. Let $\mathbf{C}_2$ differ from $\mathbf{C}_1$ only in that its ordering $\leq$ is the partial ordering given by the left-to-right presentation of the universe above. It should be obvious that these are indeed ordered frames. Define the ordered frame morphism $\mathbf{C}_1 \xrightarrow{f} \mathbf{C}_2$ to be the obvious identity mapping on the universe. Then f is an injective and surjective ordered frame morphism.

If $\mathbf{F}$ is the frame with two-element universe $\{x, x^*\}$, the empty relations and the discrete ordering, then $\mathbf{C}_1$ is a disjoint union of copies of $\mathbf{F}$; and where $\mathbf{N}$ are the natural numbers, we have $\mathbf{C}_1^+ \cong \prod_{n\in N} \mathbf{F}^+$. By considering the cones of $\mathbf{C}_2$, it is not difficult to see that $f^+(\mathbf{C}_2^+)$ is a chain in $\mathbf{C}_1^+$.

Consider now the filter G of cofinite subsets of $\mathbf{N}$ and the quotient $\prod_{n\in N} \mathbf{F}^+/G$ with for $X, Y \in \prod_{n\in N} \mathbf{F}^+$, $X\theta^G Y$ iff $\{n \mid X \cap \{n, n^*\} = Y \cap \{n, n^*\}\}$ is cofinite in $\mathbf{N}$. This is indeed a congruence, and its restriction to $f^+(\mathbf{C}_2^+)$ is a three-element algebra $\mathbf{3}^+$ with $0 < a = \neg a < 1$ and $\nu 0 = 1$. For if X, Y are cones of $\mathbf{C}_2$ other than $\emptyset$ and C, then they are of the form $\{m \mid n \leq m\}$, for some $n \in C$. Let $X = \{m \mid n \leq m\}$ and $Y = \{m \mid n' \leq m\}$; and let p be the $\leq$-greatest of the set $\{\pm n, \pm n'\}$. Then X and Y have the same intersection with the set $\{p', -p'\}$ for any $p' \geq p$ - that is, at a cofinite number of indices. So X and Y are congruent. From here, it is easy to show that this restriction of the congruence is indeed isomorphic to $\mathbf{3}^+$.

Now $\mathbf{3}^+$ has no non-isomorphic non-trivial homomorphic images, so to show that f^+ is not epi it is enough to find congruences $\psi \neq \psi'$ on $\prod_{n\in N} \mathbf{F}^+$ with $\theta^G < \psi < \nabla$, $\theta^G < \psi' < \nabla$ and $\prod_{n\in N} \mathbf{F}^+/\psi \cong \prod_{n\in N} \mathbf{F}^+/\psi'$. This will then provide distinct homomorphisms from $\mathbf{C}_1^+$ to the quotient which agree on $f^+(\mathbf{C}_2^+)$, thus showing that f^+ is not epi.

But this is easily done: $\mathbf{F}^+$ is finite and so isomorphic to any of its ultrapowers, so we need only find distinct ultrafilters over $\mathbf{N}$ containing G, thus providing distinct quotient morphisms to the same algebra $\mathbf{F}^+$. Let $e, o \subseteq \mathbf{N}$ be the even and odd numbers respectively. Clearly $\{e\}$ is separated from G, so G can be extended to an ultrafilter U over $\mathbf{N}$ not containing e and so containing its complement o. Similarly, let U' be an ultrafilter extending G which contains e and not o. Then to complete the proof it remains to be shown that U and U' define different congruences θ^U and $\theta^{U'}$ on $\prod_{n\in N}\mathbf{F}^+$. But considering $\mathbf{N}$ as an element of this algebra, we have $e\theta^U\emptyset$ and $e\theta^{U'}\mathbf{N}$ because, respectively, $o \in U$ and $e \in U'$; however $\emptyset\theta^U\mathbf{N}$ fails, and therefore so does $e\theta^U\mathbf{N}$. So θ^U and $\theta^{U'}$ are distinct congruences. □

The same example illustrates another negative result:

Theorem 45 *If f is an injective ordered frame morphism, then $(f^+)_+$ need not be injective.*

Proof. Consider the prime filter $\mathbf{C}_2^+\backslash\{\emptyset\}$ of $\mathbf{C}_2^+$ in the example above, consisting of all the non-empty cones of $\mathbf{C}_2$, as well as $f^+(\mathbf{C}_2^+\backslash\{\emptyset\})$ in $\mathbf{C}_1^+$ which is easily seen to be closed under finite meets $\mathbf{C}_2^+\backslash\{\emptyset\}$ is easily seen to be the greatest prime filter of $\mathbf{C}_2^+$, so any prime filter U of $\mathbf{C}_1^+$ containing $f^+(\mathbf{C}_2^+\backslash\{\emptyset\})$ will be mapped by $(f^+)_+$ to $\mathbf{C}_2^+\backslash\{\emptyset\}$. So if two such prime filters can be found, then $(f^+)_+$ is not injective. So let $e' = \{n, -n \in C \mid n \text{ is even }\}$ and $o' = \{n, -n \in C \mid n \text{ is odd }\}$. Then because every element of $f^+(\mathbf{C}_2^+\backslash\{\emptyset\})$ contains both even and odd numbers, and so for $x \in f^+(\mathbf{C}_2^+\backslash\{\emptyset\})$, $x \not\leq e'$ and $x \not\leq o'$, the prime filter theorem may be applied to this set in connection with both $\{e'\}$ and $\{o'\}$. As in the previous proof, it can be seen that the two resulting prime filters must be distinct. □

The notion of subdirect irreducibility can be related to frames. If $\mathbf{C} =< X, *, R >$ is a frame and $x \in X$, let $\mathbf{C}_x =< X_x, *, R_x >$ be the smallest subframe of $\mathbf{C}$ containing x. Then we have the injective frame morphism $\mathbf{C}_x \xrightarrow{f_x} \mathbf{C}$, and so the surjective homomorphism $\mathbf{C}^+ \xrightarrow{f_x^+} \mathbf{C}_x^+$, for each $x \in X$. Now we show that $\mathbf{C}^+ \xrightarrow{h} \prod_{x\in X}\mathbf{C}_x^+$, given by $h(Z)(x) = f_x^+(Z)$ is injective, for any $x \in X$ and $Z \subseteq X$. Now $f_x^+(Z) = Z \cap X_x$; so if $Y \neq Z$ with $w \in Y$ and $w \notin Z$, say, then $w \in Y \cap X_w$ but $w \notin Z \cap X_w$, so $f_w^+(Y) \neq f_w^+(Z)$. This means that $h(Y)(w) \neq h(Z)(w)$, so $h(Y) \neq h(Z)$. So h is a subdirect embedding.

To show that h represents $\mathbf{C}^+$ as a subdirect product of subdirectly irreducible algebras, it must be shown that $\mathbf{C}_x$ is subdirectly irreducible. First, simply for the purpose of illumination of certain important algebras, let x be such that $\forall y \in X(xRy \text{ or } xR'y \Rightarrow y = x \text{ or } y = x^*)$. Then $X_x = \{x, x^*\}$. Either $x = x^*$ or $x \neq x^*$. If $x = x^*$ then the underlying non-modal algebra of $\mathbf{C}_x^+$ is the Boolean algebra $\mathbf{2}$ with $X_x = \{0, 1\}$. If $\nu 0 = 1$, then this algebra is known as $\underline{\mathbf{2}}^+$, and if $\nu 0 = 0$ then it is called $\underline{\mathbf{2}}$. Clearly both are subdirectly irreducible, the only congruences being Δ and ∇. Now suppose that $x \neq x^*$. Then the underlying non-modal algebra of $\mathbf{C}_x^+$ is $\mathbf{4}$, with universe $\{a, b, 0, 1\}$ and $a \wedge b = 0$, $a \vee b = 1$, $\neg a = a$ and $\neg b = b$. $\mathbf{4}$ is characteristic in the study of the non-modal part of the logic in the same way as $\mathbf{2}$ is characteristic in the study of classical propositional logic. As will be seen later, there are 10 non-isomorphic modal algebras based on $\mathbf{4}$, and all are subdirectly irreducible because Δ and ∇ are the only congruences on $\mathbf{4}$: if $a\theta b$ then $(a \wedge b)\theta(b \wedge b) = b$, and $(a \vee b)\theta(b \vee b) = b$. So $0\theta b$ and $b\theta 1$ so $0\theta 1$ - that is $\theta = \nabla$. And if $a\theta 1$, say, then $\neg a\theta\neg 1$ so $a\theta 0$ and again $\theta = \nabla$. So ∇ is the only congruence other that Δ. But for the general case, let $\theta \neq \Delta$ be a congruence on $\mathbf{C}^+$. Because $\mathbf{C}^+$ is a powerset algebra it is complemented, and congruences on complemented algebras have the following properties. Recall that we denote the complement of Y by $-Y$ and further that $Y^* = \{y^* \mid y \in Y\}$. Also let $Y + Z =_{def} (-Y \cup Z) \cap (Y \cup -Z)$. None of these are operations on modal algebras, but if an algebra is complemented, then

Lemma 46 *If $Y\theta Z$ then $-Y\theta - Z$.*

Proof. Let $Y\theta Z$. Then $-Z\theta - Z$ so $(Y \cap -Z)\theta(Z \cap -Z) = \emptyset$. Also $-Y\theta - Y$, so $(-Y \cup (Y \cap -Z))\theta(-Y \cup \emptyset) = -Y$, and using distributivity and the fact that $Y \cup -Y = 1$, we have $(-Y \cup -Z)\theta - Y$. Interchanging Y and Z in the above gives $(-Y \cup -Z)\theta - Z$, so $(-Y \cup -Z)\theta(-Y \cap -Z)$. And in distributive lattices this is the case precisely whenever $-Y\theta - Z$. □

Lemma 47 *If $Y\theta Z$ then $Y^*\theta Z^*$.*

Proof. If $Y\theta Z$ then $\neg Y\theta\neg Z$ and $-\neg Y\theta-\neg Z$. But $-\neg Z = --Z^* = Z^*$, and similarly for Y, so $Y^*\theta Z^*$. □

Lemma 48 *$Y\theta Z$ iff $(Y+Z)\theta 1$.*

Proof. $Y\theta Z$ iff $(Y\cap Z)\theta(Y\cup Z)$, so $((Y+Z)\cup(Y\cap Z))\theta((Y+Z)\cup(Y\cup Z))$. But for a,b complemented, it is straightforward to work out that $(a+b)\wedge(a\vee b) = a\wedge b$ and $(a+b)\vee(a\vee b) = 1$: $a+b$ is the relative complement of $a\vee b$ in the interval $[a\wedge b, 1]$. This shows that $(Y+Z)\theta 1$. If $(Y+Z)\theta 1$ then $((Y\cup Z)\cap(Y+Z))\theta((Y\cup Z)\cap 1)$, that is $(Y\cap Z)\theta(Y\cup Z)$ so $Y\theta Z$. □

Theorem 49 *$\mathbf{C}_x^+$ is subdirectly irreducible.*

Proof. Let $\theta \neq \Delta$ be a congruence on $\mathbf{C}_x^+$ with $Y\theta Z$ and $Y \neq Z$. By the above lemma and the fact that $\theta \neq \Delta$ we have $(Y+Z)\theta 1$ and $Y+Z \neq 1$, that is $Y+Z \subset X_x$. Let κ_i vary over $\{\nu_R, \mu_{R'}, *\}$; then for any n we have $(\kappa_1\cdots\kappa_n(Y+Z))\theta 1$. The cases of ν_R and $\mu_{R'}$ are clear, and that of $*$ follows from the above lemma and the fact that $1^* = 1$ because X_x is closed under $*$. But on the other hand for some n we must have $\kappa_1\cdots\kappa_n(Y+Z) \leq X_x\backslash\{x\}$. For if not, then where S_i is the frame relation defining κ_i, for any path $x = y_1S_1\ldots S_ny_n = y$ of length n in $\mathbf{C}_x$ we have $y \in Y+Z$. But since $Y+Z \subset X_x$ this contradicts the construction of $\mathbf{C}_x$. So let $(\kappa_1\cdots\kappa_n(Y+Z))\theta 1$ and $\kappa_1\cdots\kappa_n(Y+Z) \leq X_x\backslash\{x\}$. Then

$$X_x\backslash\{x\} = (\kappa_1\cdots\kappa_n(Y+Z)\vee X_x\backslash\{x\})\theta(1\vee X_x\backslash\{x\}) = 1$$

So letting $\theta_{(X_x\backslash\{x\},1)}$ be the least congruence θ such that $X_x\backslash\{x\}\theta 1$, it has been shown that any congruence other than Δ is contained in $\theta_{(X_x\backslash\{x\},1)}$. So the set $\mathrm{Con}(\mathbf{C}_x^+)/\{\Delta\}$ has a minimum element and so $\mathbf{C}_x^+$ is subdirectly irreducible. □

Corollary 50 *For any variety V, if $\mathbf{C}^+ \in V$ is a powerset algebra then there are subdirectly irreducible $\{\mathbf{C}_i^+ \mid i \in I\} \subseteq V$ such that $\mathbf{C}^+ \in V$ is a subdirect product of $\{\mathbf{C}_i^+ \mid i \in I\}$.* □

It was seen earlier that a variety V is complete iff it is generated by $F(V)^+ = \{\mathbf{C}^+ \mid \mathbf{C}^+ \in V\}$. Now for any class of algebras K we have $SPSP(K) = SP(K)$ and it has just been shown that $F(V)^+ \subseteq SP(F(V)^+_{SI})$ so for a complete variety V,

$$V = HSP(F(V)^+) \subseteq HSPSP(F(V)^+_{SI}) = V(F(V)^+_{SI}).$$

And obviously $F(V)^+_{SI} \subseteq V$ so $V(F(V)^+_{SI}) \subseteq V$, and $V(F(V)^+_{SI}) = V$. In other words,

Theorem 51 *A variety is complete iff it is generated by its subdirectly irreducible powerset algebras.* □

5.2 A Variety With $2^{\aleph_0}$ Non-Classical Covers

An example of a complete variety is that defined by the axiom $\vdash \alpha \vee \neg\alpha$. Its algebras satisfy the equation $a \vee \neg a = 1$, and the two modal operators turn out to be identical. For modal algebras satisfy $\nu a \vee \neg\mu a = 1$ and $\nu a \wedge \neg\mu a = 0$, and adding $a \vee \neg a = 1$ gives $\nu a \vee \neg\nu a = 1$ and $\nu a \wedge \neg\nu a = 0$. But complements are unique in distributive lattices, so $\neg\nu a = \neg\mu a$ and so $\nu a = \mu a$. This is the variety of classical modal algebras, and the frames determined by this axiom are precisely those such that for all worlds x, $x = x^*$. This change forces the collapse $R = R'$, since xRy iff $x = x^*R'y$. Subvarieties of this variety are those that we are *not* in particular interested in, since these have been studied in depth elsewhere. Some of that study has been heuristically useful, however, in looking at the rest of the lattice of varieties, and our example is placed at the edge of this sublattice of varieties.

Thus, we will adapt an example of Blok[1] to show that there are $2^{\aleph_0}$ incomplete non-classical modal logics: more work is required here though, because many of the proof techniques used there fail in the weaker non-classical semantics, so this serves as an illustration of those methods of proof required. First, however, the example will be put to a different use. Here the reader is advised that, for much of the following, understanding requires some intimacy with the frames.

Let $\mathbf{N}$ be the natural numbers and for $M \subseteq \mathbf{N}$ define $\mathbf{A}_{\underline{M}}$ to be the algebra of finite and cofinite subsets of the universe of the frame $\mathbf{D}_{\underline{M}} =< \mathbf{N}, *, R >$ defined by

$$mRn \text{ iff } \begin{cases} m-1=n, \\ m+1 \leq n \\ \text{or } m=n \text{ and } m \notin M; \end{cases}$$

and $\forall n \in \mathbf{N}$, $n^* = n$. It is easy to check that $\mathbf{A}_{\underline{M}}$ is a modal algebra with, for $Z \in A_{\underline{M}}$, νZ being $\emptyset$ if Z is finite and being cofinite if Z is cofinite. Then define the variety $K_0 = V(\{\mathbf{A}_{\underline{M}} \mid M \subseteq \mathbf{N}\})$. It can be seen that K_0 determines a classical modal logic.

Now for $a, a^* \notin \mathbf{N}$, $a \neq a^*$ and $M \subseteq \mathbf{N}\backslash\{1,2,3\}$, define $\mathbf{B}_{\underline{M}}$ to be the algebra of finite and cofinite subsets of the frame $\mathbf{C}_{\underline{M}} =< \mathbf{N} \cup \{a, a^*\}, *, R >$ with

$$mRn \text{ iff } \begin{cases} m-1=n, \\ m+1 \leq n, \\ m=n \text{ and } m \notin M, \\ m=1 \text{ and } n=a \\ \text{or } m=a \text{ and } n=3; \end{cases}$$

and $\forall n \in \mathbf{N}$, $n^* = n$. As usual, $xR'y$ iff x^*Ry: for example, we have $1R'a$ and $a^*R'3$. $\mathbf{B}_{\underline{M}}$ is indeed a modal algebra: the finite and cofinite subsets of $B_{\underline{M}} = \mathbf{N} \cup \{a, a^*\}$ are clearly closed under $\cup$ and $\cap$, and for $Y \in B_{\underline{M}}$, Y^* is finite

---

[1] Blok 1980.

iff Y is finite and cofinite iff Y is cofinite; so if Y is finite then $\neg Y = -Y^*$ is cofinite, and if Y is cofinite then $\neg Y$ is finite. If $Y \subseteq \mathbf{N}$ is finite then it is easily seen that $\nu(Y \cup \{a, a^*\}) \subseteq \{a, a^*\}$ and $\mu(Y \cup \{a, a^*\}) \subseteq \{a, a^*\}$ and so are finite. This shows $\nu Z, \mu Z$ to be finite for any finite $Z \in B_{\underline{M}}$. And if $Y \in B_{\underline{M}}$ is cofinite, then for some n, $Y' = [n, \infty) \subseteq Y$. But $[n+1, \infty) \subseteq \nu Y' \subseteq \nu Y$, and $[n+1, \infty)$ is cofinite so νY is cofinite. The same can be shown for μY, so $\mathbf{B}_{\underline{M}}$ is a modal algebra.

Theorem 52 *$\mathbf{B}_{\underline{M}}$ has no proper subalgebras.*

Proof. We show that $\mathbf{B}_{\underline{M}}$ is the least subalgebra of $\mathbf{B}_{\underline{M}}$ containing 0: in other words that $\mathbf{B}_{\underline{M}}$ is *0-generated.* First, we have $\neg 0 = 1$, $\nu 0 = \{a^*\}$, $\mu 0 = \{a\}$ and $\neg\nu\neg\nu 0 = \{1\}$. Now, letting

$$\nu^n(a) =_{def} \overbrace{\nu \cdots \nu}^{\text{n times}}(a)$$

and defining μ^n similarly, for any $n \geq 1$ it can be seen that

$$\neg\nu 0 \wedge \neg\mu 0 \wedge \nu^n \neg\nu 0 = \mathbf{N} \backslash \{1, \ldots n\}.$$

So given that the complement of $\nu^n\neg\nu 0$ is $\neg\mu\nu^{n-1}\neg\nu 0$, it is not difficult to see that for any $n \geq 2$

$$\{n\} = \neg\nu 0 \wedge \neg\mu 0 \wedge \nu^{n-1}\neg\nu 0 \wedge \neg\mu\nu^{n-1}\neg\nu 0.$$

For any $x \in \mathbf{N} \cup \{a, a^*\}$ let $\underline{x}$ be the element $\{x\}$ in $B_{\underline{M}}$. Then for any finite $Z \in B_{\underline{M}}$, $Z = \bigvee\{\underline{x_i} \mid i \in I\}$ for some finite I, and so is in the least subalgebra of $\mathbf{B}_{\underline{M}}$. And if Z is cofinite then it can be expressed as the complement of a finite element: because $\neg\nu a$ is the complement of μa and $\neg\mu a$ is the complement of νa, it can be seen, for example, that $\neg\nu a \wedge \mu b$ is the complement of $\mu a \vee \neg\nu b$, and $\neg\nu a \vee \mu b$ is the complement of $\mu a \wedge \neg\nu b$. Continuing this process on the term which is equal to the complement of Z gives a term for Z, so all cofinite elements are also in the least subalgebra of $\mathbf{B}_{\underline{M}}$. So $\mathbf{B}_{\underline{M}}$ has no proper subalgebras. $\square$

Let $n \in \mathbf{N}\backslash\{1,2,3\}$. Then $\neg\underline{n} = -\{n\}^* = -\{n^*\} = -\{n\}$. So $\neg\underline{n}$ is the complement of $\underline{n}$. It can be shown that $n \in \nu\neg\{n\}$ iff n is R-irreflexive, so $\neg\underline{n} \vee \nu\neg\underline{n} = 1$ iff n is irreflexive: that is, if $n \in M$. So $\mathbf{B}_{\underline{M}} \models \neg\underline{n} \vee \nu\neg\underline{n} = 1$ iff $n \in M$, and also $\mathbf{B}_{\underline{M}} \models \neg\underline{n} \vee \neg\nu\neg\underline{n} = 1$ iff $n \notin M$. So if $M \neq N$, then there is an equation valid in $\mathbf{B}_{\underline{M}}$ and not in $\mathbf{B}_{\underline{N}}$, and vice versa, so $V(\mathbf{B}_{\underline{M}}) \neq V(\mathbf{B}_{\underline{N}})$, and $\mathbf{B}_{\underline{M}} \notin V(\mathbf{B}_{\underline{N}})$, $\mathbf{B}_{\underline{N}} \notin V(\mathbf{B}_{\underline{M}})$.

Theorem 53 $\mathbf{B}_{\underline{M}}$ *is subdirectly irreducible, for any* $M \subseteq \mathbf{N}\backslash\{1,2,3\}$.

Proof. Because $\mathbf{B}_{\underline{M}}$ is complemented, If $\theta \neq \Delta$ we may assume that for some $a \in B_{\underline{M}}$, $a\theta 1$ and $a < 1$. So $-a\theta 0$ and $0 < -a$. Because if $0 \leq b \leq c$ and $0\theta c$ then $0\theta b$, we may assume that either for some n, $\underline{n}\theta 0$; or $\underline{a}\theta 0$ or $\underline{a}^*\theta 0$. Take each of the possibilities in turn:

1. $\underline{n}\theta 0$. $\{n\}\theta 0$ so $\neg\{n\}\theta 1$ and $\nu\neg\{n\}\theta 1$, and given that $-\nu\neg\underline{n} = \neg\mu\neg\underline{n}$ we have $\neg\mu\neg\underline{n}\theta 0$. But it can be checked that for any n, the frame element 1 is not in $\nu\neg\{n\}$, so $1 \in -\nu\neg\{n\} = \neg\mu\neg\{n\}$ and so $0 < \underline{1} \leq \neg\mu\neg\underline{n}$. From which we have $\underline{1}\theta 0$.

2. $\underline{a}\theta 0$. That is to say $\mu 0\theta 0$. Given that $-\mu 0 = \neg\nu 0$, we have $\neg\nu 0\theta 1$ and so $\nu 0\theta 0$: in other words, $\underline{a}^*\theta 0$. So

3. $\underline{a}^*\theta 0$. In other words $\nu 0\theta 0$ and so $\neg\nu 0\theta 1$ and $\mu\neg\nu 0\theta 1$. So $\underline{1} = \neg\nu\neg\nu 0 = -\mu\neg\nu\theta 0$, and $\underline{1}\theta 0$.

So $\theta_{(\underline{1},0)}$ is contained in any congruence $\theta \neq \Delta$, showing that $\mathbf{B}_{\underline{M}}$ is subdirectly irreducible. □

A *cover* of an element a is a lattice is an element b with $a < b$ and for any c with $a \leq c \leq b$ then $c = a$ or $c = b$. This is written as $a \prec b$. It will now be shown that for any $M \subseteq \mathbf{N}\backslash\{1,2,3\}$, $K_0 \prec K_0 \vee V(\mathbf{B}_{\underline{M}})$, and that K_0 has $2^{\aleph_0}$ covers. It should be clear that $K_0 < K_0 \vee V(\mathbf{B}_{\underline{M}})$, because K_0 satisfies $\nu 0 = 0$, whereas $V(\mathbf{B}_{\underline{M}})$ does not, and so $K_0 \vee V(\mathbf{B}_{\underline{M}})$ does not. The tactic of the proof is to show that if $\mathbf{A} \in (K_0 \vee V(\mathbf{B}_{\underline{M}}))_{SI}$, then $\mathbf{A} \in K_0$ or $\mathbf{B}_{\underline{M}} \in S(\mathbf{A})$. For then if $K_0 < V < K_0 \vee \mathbf{B}_{\underline{M}}$ there is some $\mathbf{A} \in V_{SI}$ with $\mathbf{A} \notin K_0$; then $\mathbf{B}_{\underline{M}} \in S(\mathbf{A})$, so $\mathbf{B}_{\underline{M}} \in V_{SI}$ and thus $V(\mathbf{B}_{\underline{M}}) \leq V$ and $K_0 \vee V(\mathbf{B}_{\underline{M}}) \leq V$, contrary to assumption.

So let $\mathbf{A}$ be a subdirectly irreducible algebra in $V(\mathbf{B}_{\underline{M}})$. $\mathbf{A} \in (K_0)_{SI} \cup V(\mathbf{B}_{\underline{M}})_{SI}$, so if $\mathbf{A} \in (K_0)_{SI}$, then the above already is satisfied. Assume then that $\mathbf{A} \in HSP_U(V(\mathbf{B}_{\underline{M}}))$: then we must have some $\mathbf{A}_1$, $\mathbf{A}_2$, I and U with $\mathbf{A}_1 = \mathbf{B}_{\underline{M}}^I/U$, $\mathbf{A}_2 \xrightarrow{k} \mathbf{A}_1$ injective and $\mathbf{A}_2 \xrightarrow{h} \mathbf{A}$ surjective. Then either h is injective or it is not. First suppose that it is. Then $\mathbf{A}$ is isomorphic to a subalgebra of $\mathbf{A}_1$. Now consider the least subalgebra of $\mathbf{A}_1$, that generated by the element 0. But mapping each $x \in \mathbf{B}_{\underline{M}}$ to $\overline{x}/U$ in $\mathbf{A}_1$, where $\forall i \in I(\overline{x}(i) = x)$, is an embedding: mapping x to $\overline{x}$ in $\mathbf{B}_{\underline{M}}^I$ is clearly an injective homomorphism, but because $\mathbf{B}_{\underline{M}} \not\models \nu\neg\nu 0 = 1$, by Loś's theorem this also fails in $\mathbf{A}_1$. So $\overline{B_{\underline{M}}\backslash\{1\}}/U \neq \overline{1}/U$, and taking complements, $\overline{\{1\}}/U \neq \overline{O}/U$. By the previous theorem, this shows the homomorphism to be injective. Because $\mathbf{B}_{\underline{M}}$ is generated by 0, this mapping makes $\mathbf{B}_{\underline{M}}$ isomorphic to the least subalgebra of $\mathbf{A}_1$, which is also the least subalgebra of $\mathbf{A} \cong \mathbf{A}_2$. this shows that $\mathbf{B}_{\underline{M}} \in S(\mathbf{A})$.

Otherwise, h is not injective. The plan for this stage of the proof as follows. First we establish some first-order properties of $\mathbf{B}_{\underline{M}}$. Because of Loś's theorem on ultraproducts, all of these will be true of $\mathbf{A}_1$; and because a universal sentence true of an algebra is also true of all of its subalgebras, the universal sentences will be true of $\mathbf{A}_2$. This is an argument which will be used frequently throughout this chapter. Then it is shown that the fact that h is not injective means that $\mathbf{A}$ is a classical modal algebra, and is in fact a subalgebra of a homomorphic image $\mathbf{A}_1/\theta$ of $\mathbf{A}_1$. These facts will allow us to show that $\mathbf{A}_1/\theta$ is also a homomorphic image of $\mathbf{A}_{\underline{M}}{}^I/U$ and so that $\mathbf{A} \in K_0$. First of all the first-order properties, which are not difficult to check by inspecting $\mathbf{B}_{\underline{M}}$ and the frame by means of which it was defined.

Theorem 54 *The following properties hold true of* $\mathbf{B}_{\underline{M}}$*:*

1. *Apart from the top element* 1, $\{a\}$ *is the only* $x \in B_{\underline{M}}$ *such that* $\mu x = x$. *So let* $A(x)$ *abbreviate* $x \neq 1 \wedge \mu x = x$. *Then*

$$\mathbf{B}_{\underline{M}} \models \exists! x A(x).$$

2. *Apart from* 1, $\{a^*\}$ *is the only* $x \in B_{\underline{M}}$ *such that* $\nu x = x$. *Letting* $A^*(x)$ *abbreviate* $x \neq 1 \wedge \nu x = x$, *we have*

$$\mathbf{B}_{\underline{M}} \models \exists! x A^*(x).$$

Asserting that complements exist is in general a $\forall\exists$*-property, but for any* $x \in B_{\underline{M}}$ *we can determine its complement given its relation to the elements* $\{a\}$ *and* $\{a^*\}$.

3. *If* $\{a, a^*\} \leq z$, *then* $\neg z$ *is the complement of* z*:*

$$\mathbf{B}_{\underline{M}} \models \forall x \forall y \forall z (A(x) \wedge A^*(y) \wedge (x \leq z) \wedge (y \leq z) \Rightarrow (z \wedge \neg z = 0) \wedge (z \vee \neg z = 1)).$$

4. *If* $\{a, a^*\} \wedge z = 0$, *then* $\neg z$ *is the complement of* z*:*

$$\mathbf{B}_{\underline{M}} \models \forall x \forall y \forall z (A(x) \wedge A^*(y) \wedge (x \not\leq z) \wedge (y \not\leq z) \Rightarrow (z \wedge \neg z = 0) \wedge (z \vee \neg z = 1)).$$

In the two remaining cases we have either:

5. *the complement of* z *is* $\neg\{a^*\} \wedge (\{a^*\} \vee \neg z)$*:*

$$\mathbf{B}_{\underline{M}} \models \forall x \forall y \forall z (A(x) \wedge A^*(y) \wedge (x \leq z) \wedge (y \not\leq z) \Rightarrow$$
$$(z \wedge (\neg y \wedge (y \vee \neg z)) = 0) \wedge (z \vee (\neg y \wedge (y \vee \neg z)) = 1)).$$

6. *or the complement of* z *is* $\neg\{a\} \wedge (\{a\} \vee \neg z)$*:*

$$\mathbf{B}_{\underline{M}} \models \forall x \forall y \forall z (A(x) \wedge A^*(y) \wedge (x \not\leq z) \wedge (y \leq z) \Rightarrow$$
$$(z \wedge (\neg x \wedge (x \vee \neg z)) = 0) \wedge (z \vee (\neg x \wedge (x \vee \neg z)) = 1)).$$

Proof. These can be seen by inspection. □

So $\mathbf{A}_1$ satifies these properties, and so is complemented; also, $\mathbf{A}_2$ satisfies the universal properties among them. But now we have

Theorem 55 $\mathbf{A}_2$ *is complemented.*

Proof. Rewrite the first two sentences above, decomposing each into an existential sentence expressing existence and a universal sentence expressing uniqueness. Then the universal sentences hold in $\mathbf{A}_2$ because it is true of $\mathbf{A}_1$. But it was seen earlier that $\mathbf{B}_{\underline{M}}$ is isomorphic to the least subalgebra of $\mathbf{A}_1$ and so of $\mathbf{A}_2$. So the existential sentences hold in $\mathbf{A}_2$, given that they hold in one of its subalgebras. This shows that $\mathbf{A}_2$ has unique elements u^a and u^{a^*} such that $A(u^a)$ and $A^*(u^{a^*})$. Then each element of $\mathbf{A}_2$ matches precisely one of the four mutually exclusive antecedent of the last four universal sentences, and this gives its complement. □

Where convenient, u^a and u^{a^*} will also name the corresponding elements in $\mathbf{A}_1$ and $\mathbf{B}_{\underline{M}}$, which should not cause confusion given the subalgebra connection between them. Now it is shown that since h is not injective, it collapses the distinctions between ν and μ, and between negation and complementation.

Theorem 56 *For all* $z \in \mathbf{A}_2$, $h(\nu z) = h(\mu z)$, *and* $h(\neg z) = h(-z)$.

Proof. It is easy to see that in $\mathbf{B}_{\underline{M}}$, νz and μz differ, if at all, only with respect to containing $\{a\}$ and $\{a^*\}$, and that each must contain at least one of these. This justifies

$$\mathbf{B}_{\underline{M}} \models \forall x \forall y \forall z (A(x) \wedge A^*(y) \Rightarrow (y \vee \mu z) = (x \vee \nu z))$$

So for all $z \in \mathbf{A}_2$ we have $u^{a^*} \vee \mu z = u^a \vee \nu z$, and consequently if $h(u^a) = h(u^{a^*}) = 0$ then $h(\nu z) = h(\mu z)$. Also, by considering the first-order sentences above giving complements, it can be seen that if $h(u^a) = h(u^{a^*}) = 0$, then for any $z \in \mathbf{A}_2$ we have $h(\neg z) = h(-z)$. Another property of $\mathbf{B}_{\underline{M}}$ is that apart from 1, $B_{\underline{M}} \backslash \{a^*\} = \neg u^a$ is the only element x with $\nu x = \mu x = 1$; and excluding these two elements, for any other x we have $\nu x \leq B_{\underline{M}} \backslash \{1\} = \nu \neg u^{a^*}$. So

$$\mathbf{B}_{\underline{M}} \models \forall x \forall y \forall z (A(x) \wedge A^*(y) \wedge (z \neq \neg x) \wedge (z \neq 1) \Rightarrow (\nu z \leq \nu \neg y)).$$

Clearly $\mathbf{A}_1$ and $\mathbf{A}_2$ also satisfy these properties. Now because $\mathbf{A}_2$ is complemented and h is surjective but not injective there is some $z \in A_2$ with $z \neq 1$ and $h(z) = 1$. Suppose $z \neq \neg u^a$. Then $\nu z \leq \nu\neg u^{a^*}$ by the above, and given that $h(\nu z) = \nu h(z) = 1$, then $h(\nu\neg u^{a^*}) = 1$. Otherwise $z = \neg u^a$ and so $h(u^a) = 0$. But u^a and $\neg u^{a^*}$ are complements in $\mathbf{A}_2$, and given this together with the fact that $h(u^a) = 0$ it follows that $h(\neg u^{a^*}) = 1$ and so again that $h(\nu\neg u^{a^*}) = 1$. Now regarding $\nu\neg u^{a^*} \in \mathbf{A}_2$ as an element of $\mathbf{B}_{\underline{M}}$, it is $\nu\neg\nu 0 = B_{\underline{M}}\backslash\{1\}$, And from this it can be seen that $\mathbf{B}_{\underline{M}}$, and so $\mathbf{A}_1$ and $\mathbf{A}_2$, satisfies the inequality

$$\nu^3(\nu\neg u^{a^*}) \wedge \mu\nu^2(\nu\neg u^{a^*}) \leq \neg u^a \wedge \neg u^{a^*}.$$

This may be rewritten as an equation in the modal language, which then is true in $V(\mathbf{B}_{\underline{M}})$. But given that $h(\nu\neg u^{a^*}) = 1$, we have $h(\nu^3(\nu\neg u^{a^*}) \wedge \mu\nu^2(\nu\neg u^{a^*}) = 1$ and so $h(\neg u^a \wedge \neg u^{a^*}) = 1$; thus $h(u^a) = h(u^{a^*}) = 0$. So h eliminates any distinction between complement and negation and between νz and μz: in other words, $h(\nu z) = h(\mu z)$. □

In general in algebra we have $SH(\mathbf{A}) \subseteq HS(\mathbf{A})$, and the reverse inclusion also holds for classical modal algebras $\mathbf{A}$. This reverse inclusion $HS(\mathbf{A}) \subseteq SH(\mathbf{A})$ is known as the *congruence extension property*, and in general it fails for our modal algebras. At this stage of the proof we can, however, extend the congruence on $\mathbf{A}_2$ given by h to a congruence on $\mathbf{A}_1$ by means of the next theorem.

Let F be a filter in a *complemented* modal algebra $\mathbf{A}$. If for any $a \in F$ then $-\neg a \in F$, then F is called *strongly $\neg$-consistent*; and if for any $a \in F$ we have $\nu a \in F$ then F is called *ν-open.* If *μ-open* is defined similarly, it can be seen that any strongly negation consistent ν-open filter is also μ-open, given that $-\neg\nu a = \mu a$; so we need only talk about strongly $\neg$-consistent filters being *open*.

Theorem 57 *Let F be a strongly $\neg$-consistent open filter in a complemented modal algebra $\mathbf{A}$, and define θ^F by $a\theta^F b$ iff $a + b \in F$. Then θ^F is a congruence on $\mathbf{A}$.*

Proof. Some cases are familiar from the theory of Boolean algebras. Of these, only the fact that θ^F is an equivalence relation will be given as illustration.

1. θ^F is an equivalence relation. It is easy to see that θ^F is reflexive and symmetric, given that for all $a \in A$ $a + a = 1 \in F$, and that $a + b = b + a$, so if $a + b \in F$ then $b + a \in F$. To show transitivity it is shown that $a + c = (a + b) + (b + c)$. For then if $a + b, b + c \in F$, then $(a + b) \wedge (b + c)$ is also in F because F is a filter; and so is $(a + b) + (b + c)$, because $(a + b) \wedge (b + c) \leq (a + b) + (b + c)$. Now,

$$\begin{aligned} & -(a+b) \vee (b+c) \\ = \; & (a \wedge -b) \vee (-a \wedge b) \vee ((-b \vee c) \wedge (b \vee -c)) \\ = \; & ((a \wedge -b) \vee (-a \wedge b) \vee -b \vee c) \wedge \\ & ((a \wedge -b) \vee (-a \wedge b) \vee b \vee -c) \\ = \; & ((a \wedge -b) \vee -a \vee -b \vee c) \wedge \\ & ((-a \wedge b) \vee a \vee b \vee -c) \\ = \; & (-a \vee -b \vee c) \wedge (a \vee b \vee -c) \end{aligned}$$

And similarly we have

$$(a+b) \vee -(b+c) = (a \vee -b \vee -c) \wedge (-a \vee b \vee c).$$

So

$$\begin{aligned} & (a+b) + (b+c) \\ = \; & (-a \vee -b \vee c) \wedge (a \vee b \vee -c) \wedge \\ & (a \vee -b \vee -c) \wedge (-a \vee b \vee c) \\ = \; & (b \vee (a \vee -c)) \wedge (-b \vee (a \vee -c)) \wedge \\ & (b \vee (-a \vee c)) \wedge (b \vee (-a \vee c)) \\ = \; & (a \vee -c) \wedge (-a \vee c) \\ = \; & a + c. \end{aligned}$$

2. If $a\theta^F b$ then $\neg a \theta^F \neg b$. Suppose $a + b \in F$; then because F is strongly $\neg$-consistent $-\neg(a + b) \in F$. But it can be seen that $-\neg(a + b) = \neg a + \neg b$:

this could be proved in the mechanical fashion of the previous proof, or as follows. We know that $(a+b)\wedge(a\vee b)=a\wedge b$ and $(a+b)\vee(a\vee b)=1$; and given that $(\neg a+\neg b)\wedge(\neg a\vee\neg b)=\neg a\wedge\neg b$ and $(\neg a+\neg b)\vee(\neg a\vee\neg b)=1$, by negating we also have $\neg(\neg a+\neg b)\wedge(a\wedge b)=0$ and $\neg(\neg a+\neg b)\vee(a\wedge b)=a\vee b$. So

$$\begin{aligned} & \neg(\neg a+\neg b)\wedge a+b \\ = & (\neg(\neg a+\neg b)\wedge(a\vee b))\wedge(a+b) \\ = & \neg(\neg a+\neg b)\wedge((a\vee b)\wedge(a+b)) \\ = & \neg(\neg a+\neg b)\wedge(a\wedge b) \\ = & 0 \end{aligned}$$

It can be similarly shown that $\neg(\neg a+\neg b)\vee(a+b)=1$. By negating these two conclusions we then have the desired result that $-\neg(a+b)=\neg a+\neg b$.

3. If $a\theta^F b$ then $\nu a\theta^F\nu b$. If $a+b\in F$ then because F is open $\nu(a+b)\in F$. Now by definition of $+$: $(a\wedge b)\leq(a+b)$ and so $(\nu a\wedge\nu b)=\nu(a\wedge b)\leq\nu(a+b)$; and also $(\nu a\wedge\nu b)\leq(\nu a+\nu b)$ and $(\nu a+\nu b)\vee(\nu a\vee\nu b)=1$. It follows that $(\nu a\wedge\nu b)\leq((\nu a+\nu b)\wedge(\nu(a+b)))$. Now

$$\begin{aligned} & \nu(a+b)\wedge(\nu a\vee\nu b) \\ = & (\nu(a+b)\wedge\nu a)\vee(\nu(a+b)\wedge\nu b) \\ = & \nu((a+b)\wedge a)\vee\nu((a+b)\wedge b) \end{aligned}$$

But $((a+b)\wedge a)=((a+b)\wedge b)=(a\wedge b)$, so $\nu(a+b)\wedge(\nu a\vee\nu b)=(\nu(a\wedge b))=(\nu a\wedge\nu b)$. Using these facts we then have

$$\begin{aligned} & \nu(a+b) \\ = & \nu(a+b)\wedge 1 \\ = & \nu(a+b)\wedge((\nu a+\nu b)\vee(\nu a\vee\nu b)) \\ = & (\nu(a+b)\wedge(\nu a+\nu b))\vee(\nu(a+b)\wedge(\nu a\vee\nu b)) \\ = & (\nu(a+b)\wedge(\nu a+\nu b))\vee(\nu a\wedge\nu b) \\ = & \nu(a+b)\wedge(\nu a+\nu b). \end{aligned}$$

So $\nu(a+b)\leq(\nu a+\nu b)$, and since $\nu(a+b)\in F$ it follows that $\nu a+\nu b\in F$.

The case for μ is similar, and the cases for $\wedge$ and $\vee$ are standard. So θ^F is a congruence on $\mathbf{A}$. $\square$

For θ a congruence on a complemented algebra $\mathbf{A}$ it is straightforward to check that $1/\theta$, the congruence class of 1, is a strongly $\neg$-consistent and open filter of $\mathbf{A}$. So we have

Theorem 58 *If $\mathbf{A}$ is a complemented algebra, then $\theta \mapsto 1/\theta$ is a one-one map from congruences on $\mathbf{A}$ to strongly $\neg$-consistent and open filters of $\mathbf{A}$.*

Proof. It has been shown that the map is onto. But we saw earlier that $a\theta b$ iff $a + b\theta 1$ in complemented algebras. This shows that $1/\theta$ determines θ. $\square$

We now return to the main proof. Regarding $\mathbf{A}_2$ as a subalgebra of $\mathbf{A}_1$, and letting θ be the congruence determined by h, θ extends to $\mathbf{A}_1$ by

$$a\theta b \text{ iff } \exists c \in A_2(h(c) = 1 \ \wedge \ c \leq a + b)$$

θ is indeed a congruence since it is easily checked that

$$\{a \in A_1 \mid \exists c \in A_2(h(c) = 1 \ \wedge \ c \leq a)\}$$

is strongly $\neg$-consistent and open: if $c \leq a$ then $\neg a \leq \neg c$ so $-\neg c \leq -\neg a$. And if $h(c) = 1$ then $h(\neg c) = 0$ so $h(-\neg c) = 1$. Thus if $c \leq a$ with $h(c) = 1$ we also have $-\neg c \leq -\neg a$ with $h(-\neg c) = 1$. As for being open, if $h(c) = 1$ and $c \leq a$, then $h(\nu a) = 1$ and $\nu a \leq \nu c$. Finally it is easy to show that the restriction of θ on $\mathbf{A}_2$ to $\mathbf{A}_1$ is in fact the congruence determined by h, and so that mapping $h(a) \in \mathbf{A}$ to $a/\theta \in \mathbf{A}_2/\theta$ is an injective homomorphism. Consequently $\mathbf{A} \in SH(\mathbf{A}_2)$.

Let h' be the homomorphism extending h which maps elements of $\mathbf{A}_1$ to their θ-congruence class. Regarding u^a and u^{a^*} as elements of $\mathbf{A}_1$ we have $h(u^a) = h(u^{a^*}) = 0$, so by previous considerations $\mathbf{A}_1$ is a classical modal algebra. Now recall that $\mathbf{A}_1 = \mathbf{B}^I_{\underline{M}}/U$ and let $\mathbf{A}' = \mathbf{A}_{\underline{\mathbf{M}}}{}^I/U$, where U is the same ultrafilter on the same index I. Then each element in $\mathbf{A}'$ may in a natural way be regarded as an element of $\mathbf{A}_1$, and using the same name for them these let $\mathbf{A}' \xrightarrow{g} \mathbf{A}_1/\theta$ be defined by $g(x) = h'(x)$: then

Theorem 59 *g is a surjective homomorphism.*

Proof. This will be seen to follow essentially from the one fact.

1. It should be clear that the clauses for 0, $\wedge$ and $\vee$ hold, since $0 \in \mathbf{A}_1$ is an element of $\mathbf{A}'$, and if $x, y \in \mathbf{A}_1$ are also elements of $\mathbf{A}'$, then so are $x \wedge y$ and $x \vee y$.

2. $g(1) = 1$. Now $1 \in \mathbf{A}'$ corresponds to the U-equivalence class of the constant function $\overline{\mathbf{B}_{\underline{M}} \backslash \{a, a^*\}} = \neg u^a \wedge \neg u^{a^*}$ in $\mathbf{A}_1$. But we have already seen that $h'(u^a) = h'(u^{a*}) = 0$ and so $h'(\neg u^a) = h'(\neg u^{a*}) = 1$. This shows that h' maps the U-equivalence class of $\overline{\mathbf{B}_{\underline{M}} \backslash \{a, a^*\}}$ to 1 in $\mathbf{A}_1/\theta$, so $g(1) = 1$.

3. $g(\nu x) = \nu g(x)$. This follows from the fact that for $Z \subseteq A_{\underline{M}}$, νZ in $\mathbf{A}_{\underline{\mathbf{M}}}$ and νZ in $\mathbf{B}_{\underline{M}}$ differ only in that νZ in $\mathbf{B}_{\underline{M}}$ may contain a or a^*. The fact that h' eliminates the distinction at the level of the ultrapowers by mapping $\overline{B_{\underline{M}} \backslash \{a, a^*\}}/U$ to 1 allows this clause to go through.

4. $g(\neg x) = \neg g(x)$. As this algebraic operation on the ultrapowers is defined in terms of its working on the algebras $\mathbf{A}_{\underline{\mathbf{M}}}$ and $\mathbf{B}_{\underline{\mathbf{M}}}$, and h' has eliminated any distinction between negation in $\mathbf{A}_{\underline{\mathbf{M}}}$, which was complementation over its universe $\mathbf{N} = \mathbf{B}_{\underline{M}} \backslash \{a, a^*\}$ and negation in $\mathbf{B}_{\underline{M}}$, this is true.

5. g is surjective. Take $z \in \mathbf{A}_1/\theta$, say $z = y/\theta$. Recall that $\neg u^a \wedge \neg u^{a^*}$ is the element of $\mathbf{B}_{\underline{M}}$ which at all indices is the element $\{\mathbf{N}\}$, then

$$h'(z \wedge (\neg u^a \wedge \neg u^{a^*}/U)) = h'(z) \wedge h'(\neg u^a \wedge \neg u^{a^*}/U) = h'(z) \wedge 1 = h'(z).$$

   But clearly $z \wedge (\neg u^a \wedge \neg u^{a^*}/U)$ is an element of $\mathbf{A}'$, and g then maps it to y/θ. □

So if h is not injective, then $\mathbf{A} \in K_0$ because since $\mathbf{A}' \in P_U(\mathbf{A}_{\underline{M}})$ we have $\mathbf{A}' \in K_0$, and also $\mathbf{A} \in SH(\mathbf{A}') \subseteq HS(\mathbf{A}')$. So $\mathbf{A} \in V(\mathbf{A}_{\underline{M}}) \subseteq K_0$. This completes the proof of

Theorem 60 $K_0 \prec K_0 \vee V(\mathbf{B}_{\underline{M}})$. □

And since we have also shown that if $M \neq N$ then $\mathbf{B}_{\underline{N}} \notin V(\mathbf{B}_{\underline{M}})$ and $\mathbf{B}_{\underline{M}} \notin V(\mathbf{B}_{\underline{N}})$, with each $\mathbf{B}_{\underline{N}}$ being subdirectly irreducible, it follows that

Theorem 61 K_0 *has* $2^{\aleph_0}$ *covers.* □

5.3 Examples Of Incomplete Varieties

$V(\mathbf{B}_{\underline{M}})$ will be shown to be incomplete, to which end some more properties of $\mathbf{B}_{\underline{M}}$ are established. First let ∇x abbreviate $x \wedge \nu x \wedge \mu x$. ∇ acts like a modal operator, with $\nabla 1 = 1$ and $\nabla(x \wedge y) = \nabla x \wedge \nabla y$. Also, let the definition be extended by letting it be more fully defined as $\nabla^0 x = x$; and for $n \geq 1$

$$\nabla^n x = \nabla^{n-1} x \wedge \nu \nabla^{n-1} x \wedge \mu \nabla^{n-1} x.$$

Then clearly

Theorem 62 $\mathbf{B}_{\underline{M}} \models \nabla x \leq x$. □

And in $\mathbf{B}_{\underline{M}}$ we also have

Theorem 63 $\nabla^2 x = 0$ *iff* x *is finite.*

Proof. The if part of the proof follows from these easily established cases, where x is finite:

If $\{a, a^*\} \not\subseteq x$ then $\nabla x = 0$;
If $\{a, a^*\} = x$ then $\nabla x = 0$;
If $\{a, a^*\} \subseteq x$ and $3 \in x$ then $\nabla x = \{a, a^*\}$;
If $\{a, a^*\} \subseteq x$ and $3 \notin x$ then $\nabla x = 0$.

For the other direction suppose x is not finite: then it is cofinite. For the cases where $\mathbf{N} \subseteq x$, we have $\nabla\{\mathbf{N}\} = \mathbf{N}\backslash\{1\}$ and $\nabla^2\{\mathbf{N}\} = \mathbf{N}\backslash\{1,2\}$, so $\nabla^2 x$ is cofinite. Otherwise, for some n we have $n \notin x$ and $[n+1,\infty) \subseteq x$. But $\nabla^2[n+1,\infty) = [n+3,\infty)$, so again $\nabla^2 x$ is cofinite. □

Together with the fact that ∇x is cofinite if x is, this justifies

Corollary 64 $\mathbf{B}_{\underline{M}} \models \forall x(\nabla^2 x > 0 \Rightarrow \neg\nabla^2\neg\nabla^2 x = 1)$. □

Below we shall use the instance

$$\mathbf{B}_{\underline{M}} \models \forall x(\nabla^4 x > 0 \Rightarrow \neg\nabla^2\neg\nabla^4 x = 1).$$

Now if $x \subseteq \mathbf{N}$ is coinitial in $\mathbf{N}$, say $x = [n,\infty)$, then $\nabla x = [n+1,\infty)$ is also coinitial in $\mathbf{N}$ and $\{\nabla^m x \mid m < \infty\}$ is an infinite descending chain. In fact we have

Theorem 65 *$\{\nabla^4 x \mid x \in \mathbf{B}_{\underline{M}}\}$ is linearly ordered.*

Proof. We show that $\nabla^4 x$ is either $1 = B_{\underline{M}}$, $B_{\underline{M}}\backslash\{a^*\}$, coinitial in $\mathbf{N}$, or 0: this is clearly a linearly ordered set. If x is finite then $\nabla^4 x = \nabla^2 x = 0$, as was seen above. If $x = B_{\underline{M}}\backslash\{a^*\}$, then $\nabla x = x$ so $\nabla^4 x = x$. Now suppose that $x = B_{\underline{M}}\backslash\{n\}$. If $n > 3$ then ∇x is either $\{a, a^*\} \cup \{n\} \cup [n+2,\infty)$ or $\{a, a^*\} \cup [n+2,\infty)$, depending on whether or not $n \in M$; but because $aR3$, $a^*R'3$, $nRn+1$, and $3, n+1 \notin \nabla x$, in either case we have $\nabla^2 x = [n+3,\infty)$, and so $\nabla^4 x = [n+5,\infty)$ is coinitial in $\mathbf{N}$. For $n \leq 3$, it is not too difficult to see that $\nabla^3(B_{\underline{M}}\backslash\{1\}) = \nabla^2(B_{\underline{M}}\backslash\{2\}) = \nabla(B_{\underline{M}}\backslash\{3\}) = [5,\infty)$. Finally, if $x = B_{\underline{M}}\backslash\{a\}$, then $\nabla x = B_{\underline{M}}\backslash\{1\}$, so $\nabla^4 x = [5,\infty)$. Now, every cofinite subset of $B_{\underline{M}}$ is a finite meet of these cases and the linearly ordered set described above is obviously closed under meet. This, together with the fact that $\nabla^4(x \wedge y) = (\nabla^4 x \wedge \nabla^4 y)$, completes the proof. □

It can also be seen that

Theorem 66 $\mathbf{B}_{\underline{M}} \models \neg\nabla^2\neg\nabla^3 x \wedge \nabla^2(\neg\nabla^3 x \vee \nabla^4 x) \leq \nabla^6 x$.

Proof. If x is finite then $\neg\nabla^2\neg\nabla^3 x = 0$, and if $x = 1$ then $\nabla^6 x = 1$, so the inequation holds for these cases. Otherwise, suppose $x \neq B_{\underline{M}}\backslash\{a\}$ is cofinite and $\nabla^3 x = [n+1, \infty)$. This covers the cases of $x = \mathbf{N}$ and of those x such that for some n, $n \notin x$. Then consideration of the previous proof shows that $n > 3$, and we have $\nabla^6 x = [n+4, \infty)$ and $\neg\nabla^3 x \vee \nabla^4 x = B_{\underline{M}}\backslash\{n+1\}$. So $\nabla^2(\neg\nabla^3 x \vee \nabla^4 x) = [n+4, \infty)$, and again the inequation holds. As for $x = B_{\underline{M}}\backslash\{a\}$, it turns out that $\neg\nabla^3 x \vee \nabla^4 x = B_{\underline{M}}\backslash\{a^*, 4\}$ and that $\nabla^2(\neg\nabla^3 x \vee \nabla^4 x) = \nabla^6 x = [7, \infty)$. If $x = B_{\underline{M}}\backslash\{a^*\}$, then $\nabla x = x$ and $\neg x < x$ show that it goes through. □

Let $\mathbf{A} \cong \mathbf{C}^+ \in V(\mathbf{B}_{\underline{M}})_{SI}$ with $\mathbf{A}_1 \in P_U(\mathbf{B}_{\underline{M}})$, $\mathbf{A}_2 \xrightarrow{k} \mathbf{A}_1$ injective and $\mathbf{A}_2 \xrightarrow{h} \mathbf{A}$ surjective. Now $\underline{\mathbf{2}} \in V(\mathbf{B}_{\underline{M}})$ since mapping cofinite subsets of $B_{\underline{M}}$ to 1 and finite subsets to 0 is a surjective homomorphism, and $\underline{\mathbf{2}}^+ \notin V(\mathbf{B}_{\underline{M}})$ since $\mathbf{B}_{\underline{M}} \models \nu 0 \wedge \mu 0 = 0$ and $\underline{\mathbf{2}}^+ \not\models \nu 0 \wedge \mu 0 = 0$. The next step is to show that there is no other such $\mathbf{A}$.

So suppose $\mathbf{A}$ is as above and $\mathbf{A} \ncong \underline{\mathbf{2}}$. Now the universe of $\mathbf{A}$ must have more that two elements, so it may be assumed that there is some $x \in A$ with $0 < x < 1$. But there is more that we know about $\mathbf{A}$. Because $\{\nabla^4 x \mid x \in \mathbf{B}_{\underline{M}}\}$ is linearly ordered in $\mathbf{B}_{\underline{M}}$, for by now familiar reasons so are $\{\nabla^4 x \mid x \in \mathbf{A}_1\}$ and $\{\nabla^4 x \mid x \in \mathbf{A}_2\}$. But this is also the case for $\{\nabla^4 x \mid x \in \mathbf{A}\}$, since $\{\nabla^4 x \mid x \in \mathbf{A}\} = \{h(\nabla^4 x) \mid x \in \mathbf{A}_2\}$. Next,

$$\mathbf{B}_{\underline{M}} \models \neg\nabla^2\neg\nabla^4 x \vee \neg\nabla^2\neg\nabla^4\neg x = 1$$

for if x is cofinite then $\neg\nabla^4 x$ is finite; so $\nabla^2\neg\nabla^4 x = 0$ and $\neg\nabla^2\neg\nabla^4 x = 1$. Otherwise x is finite so $\neg x$ is cofinite and $\neg\nabla^2\neg\nabla^4\neg x = 1$. So this equation must also be true in $\mathbf{A}$. This means that for $x \in A$ with $0 < x < 1$ either $0 < \nabla^4 x$ or $0 < \nabla^4\neg x$; for otherwise $\neg\nabla^2\neg\nabla^4 x = \neg\nabla^2\neg\nabla^4\neg x = 0$, violating the equation. So we may assume without loss of generality that $0 < \nabla^4 x$. Also, we may assume that the complement of $\nabla^4 x$, which exists since $\mathbf{A} \cong \mathbf{C}^+$, is in fact $\neg\nabla^4 x$. For suppose this does not hold of x: considering again the set $\{\nabla^4 x \mid x \in \mathbf{B}_{\underline{M}}\}$, this fails to be true only of $\mathbf{B}_{\underline{M}} \backslash \{a^*\}$, so

$$\mathbf{B}_{\underline{M}} \models \forall x(\nabla^4 x \vee \neg\nabla^4 x \neq 1 \Rightarrow x = \neg\mu 0).$$

So this holds in $\mathbf{A}_1$ and $\mathbf{A}_2$. And for $h(v) = x$, we have $h(\nabla^4 v) = \nabla^4 x$ and $h(\nabla^4 v \vee \neg\nabla^4 v) = \nabla^4 x \vee \neg\nabla^4 x \neq 1$. So $\nabla^4 v \vee \neg\nabla^4 v \neq 1$ and $v = \neg\mu 0$. But then consider $-\neg x$ in $\mathbf{A}$. It is easy to see that $0 < -\neg x < 1$: if not, then $x = 1$ or $x = 0$. Also, $h(\mu 0) = \neg x$, and $\neg\nu 0$ is the complement of $\mu 0$, so $h(\neg\nu 0) = -\neg x$. It can then easily be shown that $\nabla^4\neg\nu 0 \vee \neg\nabla^4\neg\nu 0$ is valid in $\mathbf{B}_{\underline{M}}$ and so in $V(\mathbf{B}_{\underline{M}})$, proving that the negation of $\nabla^4 - \neg x$ is its complement. So we assume in what follows that we have chosen x, $0 < x < 1$, with this property.

We know that $0 < \nabla^4 x \leq x < 1$ and $\nabla^4 x \vee \neg\nabla^4 x = 1$. Since $\mathbf{B}_{\underline{M}} \models \forall x(\nabla^4 x > 0 \Rightarrow \neg\nabla^2\neg\nabla^4 x = 1)$, this also holds in the ultraproduct $\mathbf{A}_1$, and, since it is a universal sentence, in its subalgebra $\mathbf{A}_2$. Let $v, w \in A_2$ be such that $h(v) = x$ and $w = \nabla^4 v$; so $h(w) = \nabla^4 x$. Then $h(w) \neq 0$, so $w \neq 0$ and by this universal sentence $\neg\nabla^2\neg w = 1$, so $h(\neg\nabla^2\neg w) = 1$. Now we show that

Theorem 67 $\nabla^5 x < \nabla^4 x$.

Proof. Clearly we have $\nabla^5 x \leq \nabla^4 x$, so suppose for contradiction that $\nabla^5 x = \nabla^4 x$. Then because $h(\neg\nabla^2\neg w) = 1$, it follows that

$$h(\neg\nabla^2\neg w \wedge \nabla^2(\neg w \vee \nabla w)) = h(\nabla^2(\neg w \vee \nabla w)).$$

And if $\nabla^5 x = \nabla^4 x$, then $h(\nabla w) = h(w)$, and since $\nabla^4 x \vee \neg\nabla^4 x = 1$ we have $\neg h(w) \vee h(\nabla w) = 1$ and consequently $\nabla^2(\neg h(w) \vee h(\nabla w)) = 1$. But

$$h(\nabla^2(\neg w \vee \nabla w)) = \nabla^2(\neg h(w) \vee h(\nabla w)) = 1,$$

and as an instance of a valid theorem of the variety we have

$$\neg\nabla^2\neg\nabla^4 v \wedge \nabla^2(\neg\nabla^4 v \vee \nabla^5 v) \leq \nabla^7 v,$$

so

$$\neg\nabla^2\neg w \wedge \nabla^2(\neg w \vee \nabla w) \leq \nabla^3 w.$$

And since $h(\nabla^2(\neg w \vee \nabla w)) = 1$ and $h(\neg\nabla^2\neg w) = 1$, we must have $h(\nabla^3 w) = 1$. But $\nabla^3 w \leq w$ so $h(w) = 1$, contrary to the assumption that $h(w) = \nabla^4 x < 1$. This shows that $\nabla^5 x < \nabla^4 x$. □

From this it follows that $\nabla^8 x < \nabla^4 x$. Now it is shown that

Theorem 68 $\nabla^8 x \neq 0$.

Proof. We know that

$$\neg\nabla^2\neg\nabla^8 x \vee \neg\nabla^2\neg\nabla^4\neg\nabla^4 x = 1,$$

and that $\{\nabla^4 x \mid x \in \mathbf{A}\}$ is linearly ordered. If $\nabla^8 x = 0$, then $\neg\nabla^2\neg\nabla^8 x = 0$, so to satisfy the above equation we must have $\nabla^4\neg\nabla^4 x \neq 0$. And by the linear ordering we have

$$\nabla^4\neg\nabla^4 x \leq \nabla^4 x \text{ or } \nabla^4 x \leq \nabla^4\neg\nabla^4 x.$$

We have seen that $\nabla^4 x \vee \neg\nabla^4 x = 1$, and so by negating, $\nabla^4 x \wedge \neg\nabla^4 x = 0$; also we have $\nabla^4\neg\nabla^4 x \leq \neg\nabla^4 x$. Suppose the first disjunct is true, that is $\nabla^4\neg\nabla^4 x \leq \nabla^4 x$. Then $0 \neq \nabla^4\neg\nabla^4 x \leq \nabla^4 x \wedge \neg\nabla^4 x$, contradicting $\nabla^4 x \wedge \neg\nabla^4 x = 0$. So suppose that the second disjunct holds. But then

$$0 < \nabla^4 x \leq \nabla^4\neg\nabla^4 x \leq \neg\nabla^4 x,$$

which again contradicts $\nabla^4 x \wedge \neg\nabla^4 x = 0$. □

This shows that $\nabla^8 x \neq 0$, and so that $\{\nabla^{4n} x \mid n < \infty\}$ is an infinite descending chain. It is a subset of $\{\nabla^4 x \mid x \in \mathbf{A}\}\backslash\{0\}$, of which it contains a co-initial subset, according to this downward ordering. Also negation coincides with complementation for every $\nabla^{4n} x$: if $\neg x = -x$ then $\neg\nabla x = -\nabla x$, for

$$\neg\nabla x = (\neg x \vee \neg\nu x \vee \neg\mu x) = (-x \vee -\mu x \vee -\nu x) = -\nabla x.$$

Let C be this anti-chain $\{\nabla^{4n}x \mid n < \infty\}$ in $\mathbf{A} \cong \mathbf{C}^+$. Also let λ be a limit ordinal indexing a coinitial subset of C: that is, we have $\{a_\tau \mid \tau < \lambda\}$ coinitial in C with $a_{\tau 2} < a_{\tau 1}$ iff $\tau_1 < \tau_2 < \lambda$. Because $\mathbf{C}^+$ is a powerset algebra,

$$z = \bigvee\{a_\tau \wedge \neg a_{\tau+1} \mid \tau < \lambda \text{ and } \tau \text{ is even}\}$$

exists; it can also be seen that $a_\tau \wedge \neg a_{\tau+1} \neq 0$ since $-a_\tau = \neg a_\tau < \neg a_{\tau+1}$.

Theorem 69 *Let σ be odd and τ be even. Then $a_\sigma \wedge \neg a_{\sigma+1} \wedge a_\tau \wedge \neg a_{\tau+1} = 0$.*

Proof. For suppose $\sigma < \tau$: then $\sigma < \sigma+1 \leq \tau < \tau+1$, $a_{\tau+1} < a_\tau \leq a_{\sigma+1} < a_\sigma$ and $\neg a_\sigma < \neg a_{\sigma+1} \leq \neg a_\tau < \neg a_{\tau+1}$. Because $a_\tau \wedge \neg a_\tau = 0$, this means that $a_\tau \wedge \neg a_{\sigma+1} = 0$. For $\tau < \sigma$ the proof is similar. □

Because this is a powerset algebra this shows that for σ odd $a_\sigma \wedge \neg a_{\sigma+1} \not\leq z$, since

$$a_\sigma \wedge \neg a_{\sigma+1} \wedge \bigvee\{a_\tau \wedge a_{\tau+1} \mid \tau < \lambda, \tau \text{ even}\} =$$

$$\bigvee\{a_\sigma \wedge \neg a_{\sigma+1} \wedge a_\tau \wedge a_{\tau+1} \mid \tau < \lambda, \tau \text{ even}\} = 0.$$

Also, $\neg z = -z$. Because $\mathbf{A}$ is a powerset algebra and for each $\tau < \lambda$, $\neg a_\tau = -a_\tau$, this means that $-(a_\tau)^* = -a_\tau$ and so $a_\tau^* = a_\tau$. Similarly we have $(\neg a_\tau)^* = \neg a_\tau$, and because the meet of any two complemented elements is also complemented, in particular we have $(a_\tau \wedge \neg a_{\tau+1})^* = a_\tau \wedge \neg a_{\tau+1}$. So

$$\begin{aligned} & \neg \bigvee\{a_\tau \wedge \neg a_{\tau+1} \mid \tau < \lambda \text{ and } \tau \text{ is even}\} \\ = \; & -(\bigvee\{a_\tau \wedge \neg a_{\tau+1} \mid \tau < \lambda \text{ and } \tau \text{ is even}\})^* \\ = \; & -(\bigvee\{(a_\tau \wedge \neg a_{\tau+1})^* \mid \tau < \lambda \text{ and } \tau \text{ is even}\}) \\ = \; & -(\bigvee\{a_\tau \wedge \neg a_{\tau+1} \mid \tau < \lambda \text{ and } \tau \text{ is even}\}) \end{aligned}$$

Suppose $\nabla^4 z \neq 0$: obviously we have $\nabla^4 z \in \{\nabla^4 x \mid x \in \mathbf{A}\}$, so let σ be odd with $a_\sigma < \nabla^4 z$. But then

$$a_\sigma \wedge \neg a_{\sigma+1} < \nabla^4 z \leq z,$$

which we saw to be false, so $\nabla^4 z = 0$. Similarly it is shown that $\nabla^4 \neg z = 0$: otherwise, take *even* σ with $a_\sigma < \nabla^4 \neg z \in \mathcal{C}$. Then

$$0 \neq a_\sigma \wedge \neg a_{\sigma+1} < \nabla^4 \neg z \leq \neg z,$$

which is again a contradiction, since $a_\sigma \wedge \neg a_{\sigma+1} \leq z = -\neg z$. So $\nabla^4 z = \nabla^4 \neg z = 0$, contradicting the valid equation $\neg\nabla^2\neg\nabla^4 z \vee \neg\nabla^2\neg\nabla^4\neg z = 1$.

So the only algebra of the form $\mathbf{C}^+$ in $V(\mathbf{B}_{\underline{M}})$ is $\underline{\mathbf{2}}$. Because, for example, $\underline{\mathbf{2}} \models \nu 0 = 0$ unlike $\mathbf{B}_{\underline{M}}$, $V(\underline{\mathbf{2}}) \neq V(\mathbf{B}_{\underline{M}})$, thus $V(\mathbf{B}_{\underline{M}})$ is not generated by its powerset algebras, and

Theorem 70 *$V(\mathbf{B}_{\underline{M}})$ is incomplete.* □

Corollary 71 *There are $2^{\aleph_0}$ distinct logics, all valid in exactly the same frames.*

Proof. These are the logics $L(V(\mathbf{B}_{\underline{M}}))$, for each $M \subseteq \mathbf{N}\backslash\{1,2,3\}$. □

As a further example we have the following corollary.

Corollary 72 *If a logic L contains $\neg\mu\alpha \vee \alpha$, then there are $2^{\aleph_0}$ logics satisfied by the same class of frames as L.*

Proof. First, if V is non-trivial and $V \models \neg\mu\alpha \vee \alpha$ it can be seen that $\underline{\mathbf{2}} \in V$: for any $\mathbf{A} \in V$ we have $\mathbf{A} \models \mu 0 = 0$ - and therefore also $\mathbf{A} \models \nu 0 = 0$ - so $\underline{\mathbf{2}}$ is the least subalgebra of $\mathbf{A}$, and so is in V. Also, for any $M \subseteq \mathbf{N}\backslash\{1,2,3\}$, $\mathbf{B}_{\underline{M}} \notin V$ since $\mathbf{B}_{\underline{M}} \not\models \neg\mu\alpha \vee \alpha$. This shows that $\{V \vee V(\mathbf{B}_{\underline{M}}) \mid M \subseteq \mathbf{N}\backslash\{1,2,3\}\}$ are $2^{\aleph_0}$ distinct varieties. The only subdirectly irreducible powerset algebra in $V(\mathbf{B}_{\underline{M}})$ is $\underline{\mathbf{2}}$, which is also in V, so all the subdirectly irreducible powerset algebras of $V \vee V(\mathbf{B}_{\underline{M}})$ are in V. Since any $\mathbf{C}^+$ is a subdirect product of some $\{\mathbf{C}_i^+ \mid i \in I\}$, this shows that $F(V \vee V(\mathbf{B}_{\underline{M}})) = F(V)$ and so $F(L(V \vee V(\mathbf{B}_{\underline{M}}))) = F(L(V))$, completing the proof. □

Chapter 6

The Lattice Of Modal Logics

6.1 Logics Based On 4

In the previous chapter it was seen how a particular feature of the lattice of classical modal logics could be generalised to the whole lattice of four-valued modal logics, and that fewer logical techniques were available in the proof because of the relative weakness of the corresponding algebras. Now, in concentrating on the top of the lattice of four-valued modal logics, we show some important differences between the two lattices and at the same time illustrate conditions which enable the return of these powerful, rather atypical, logical properties as an example of their usefulness. More tangible examples will be given than in the previous chapter.

The Boolean algebra **2** has many familiar logical and algebraic properties, amongst which is the fact that it is characteristic for classical logic: a formula is classically valid just in case it is valid in **2** - or equivalently, the variety of Boolean algebras is generated by **2**. The two possible modal algebras based on **2**, which we called $\underline{\mathbf{2}}$ and $\underline{\mathbf{2}}^+$, also have a privileged position in classical modal logic: any consistent classical modal logic is contained in the logic of one of these two algebras. Algebraically, this means that any non-trivial classical variety contains $V(\underline{\mathbf{2}})$ or $V(\underline{\mathbf{2}}^+)$. For if $\mathbf{A}$ is a classical modal algebra and $\nu 0 = 0$, then its least subalgebra is $\underline{\mathbf{2}}$, so $V(\underline{\mathbf{2}}) \leq V(\mathbf{A})$. Otherwise $\mu 0 = \nu 0 > 0$: then it is not difficult to see that $[\nu 0)$, is a strongly $\neg$-consistent and open filter defining a congruence on $\mathbf{A}$ the quotient of which is isomorphic to $\underline{\mathbf{2}}^+$. This shows that $\underline{\mathbf{2}}^+ \in V(\mathbf{A})$ and $V(\underline{\mathbf{2}}^+) \leq V(\mathbf{A})$. Consideration of what this says about $\mathbf{A}_+$ illustrates a similar result concerning the frames corresponding to $\underline{\mathbf{2}}$ and $\underline{\mathbf{2}}^+$: the R-reflexive one point frame and the R-irreflexive one point frame respectively. If $\mathbf{C}$ is a classical frame such that for all $x \in C$ there is some $y \in C$ with xRy, the one point reflexive frame can be seen to be a p-morphic image of $\mathbf{C}$. Otherwise, there is some $x \in C$ with no y such that xRy. Then because $\mathbf{C}$ is classical we have $x = x^*$ and so $\mathbf{C}_x$, the subframe generated by x, can be seen to have only x in its universe. Clearly $\mathbf{C}_x$ is irreflexive. Now since if $\mathbf{A}$ is finite we have that $V(\mathbf{A})_{SI} \subseteq HS(\mathbf{A})$, and also $\underline{\mathbf{2}} \not\cong \underline{\mathbf{2}}^+$, it is clear that - ignoring trivial algebras - $V(\underline{\mathbf{2}})_{SI} = \{\underline{\mathbf{2}}\}$ and $V(\underline{\mathbf{2}}^+)_{SI} = \{\underline{\mathbf{2}}^+\}$, and so $V(\underline{\mathbf{2}}) \neq V(\underline{\mathbf{2}}^+)$. Together, this shows that $\underline{\mathbf{2}}$ and $\underline{\mathbf{2}}^+$ define distinct *Post complete* logics: that is, logics which are maximally consistent. Equivalently, the varieties they generate are atoms in the lattice of varieties - they cover the bottom element. And because every classical variety contains either $\underline{\mathbf{2}}$ or $\underline{\mathbf{2}}^+$, these are the *only* classical Post complete logics. This situation prompts several natural questions about four-valued modal logic in general.

It was mentioned earlier that the four element algebra $\mathbf{4}$ is characteristic for the propositional part of four-valued modal logic: $V(\mathbf{4})$ is the class of algebraic models for this logic. So it would be interesting to look at the modal algebras based on $\mathbf{4}$, not least to see whether any play a role similar to that played by the modal algebras based on $\mathbf{2}$ in classical modal logic. Recall that $\mathbf{4}$ is the algebra with universe $\{a, b, 0, 1\}$ satisfying the equations $a \wedge b = 0$, $a \vee b = 1$, $\neg a = a$ and $\neg b = b$, and $\mathbf{3}$ is the three-element subalgebra of $\mathbf{4}$ containing, say, a. There are in fact, up to isomorphism, ten possible modal algebras based on $\mathbf{4}$, which we name as follows.

$$
\begin{array}{lll}
\nu 0 = 1 & & S_1 \\
\nu 0 = b & \left\{ \begin{array}{ll} \nu a = 1; & \nu b = b \\ \nu a = b; & \nu b = 1 \\ \nu a = b; & \nu b = b \end{array} \right. & \begin{array}{l} S_2 \\ S_3 \\ S_4 \end{array} \\
\nu 0 = 0 & \left\{ \begin{array}{ll} \nu a = a; & \nu b = b \\ \nu a = b; & \nu b = a \\ \nu a = 1; & \nu b = 0 \\ \nu a = b; & \nu b = 0 \\ \nu a = a; & \nu b = 0 \\ \nu a = 0; & \nu b = 0 \end{array} \right. & \begin{array}{l} S_5 \\ S_6 \\ S_7 \\ S_8 \\ S_9 \\ S_{10} \end{array}
\end{array}
$$

Each algebra S_i is isomorphic to F_i^+ for a frame F_i with a two element universe $\{x, x^*\}$ and $x \neq x^*$. With x and x^* as *names* and not variables, such frames may be given by taking the following as a full specification of the R-relation on this universe:

$$
\begin{array}{lll}
xRx & \text{in} & F_2, F_4, F_6, F_7, F_8, F_9, F_{10}; \\
xRx^* & \text{in} & F_3, F_4, F_5, F_8, F_{10}; \\
x^*Rx & \text{in} & F_5, F_7, F_8, F_9, F_{10}; \\
x^*Rx^* & \text{in} & F_6, F_9, F_{10}.
\end{array}
$$

Thus, for example, F_1 has R and so also R' empty. Any other two-element frame with $x \neq x^*$ is p-morphically equivalent to one of these ten frames. It is straightforward but tedious to check that these propositions are true, and that these algebras are the only modal algebras based on **4**. It is easier to see that there are three algebras based on **3**: let them be named as follows. $\underline{\mathbf{3}}^+$ has $\nu 0 = 1$; otherwise $\nu 0 = 0$, and in this case $\underline{\mathbf{3}}_1$ is the algebra with $\nu a = 0$, and $\underline{\mathbf{3}}_2$ is the algebra with $\nu a = 1$. Finally, for convenience the variety of trivial algebras, those satisfying $0 = 1$, is denoted $\underline{\mathbf{T}}$.

It has already been noted that the algebras S_i are simple, that is have only two congruences, and so are subdirectly irreducible: and using the fact that for finite $\mathbf{A}$ we have $V(\mathbf{A})_{SI} \subseteq HS(\mathbf{A})$, it can be shown that the following covering relations hold between the varieties they generate.

$$\begin{aligned}
&\underline{\mathbf{T}} \prec V(\underline{\mathbf{2}}^+) \prec V(\underline{\mathbf{3}}^+) \prec V(S_1) \\
&\underline{\mathbf{T}} \prec \{V(S_2),\ V(S_3),\ V(S_4)\} \\
&\underline{\mathbf{T}} \prec V(\underline{\mathbf{2}}) \prec \{V(S_5),\ V(S_6),\ V(\underline{\mathbf{3}}_1),\ V(\underline{\mathbf{3}}_2)\} \\
&V(\underline{\mathbf{3}}_1) \prec \{V(S_8),\ V(S_9),\ V(S_{10})\} \\
&V(\underline{\mathbf{3}}_2) \vee V(\underline{\mathbf{3}}_2) \prec V(S_7)
\end{aligned}$$

For example, S_2 determines a Post complete four-valued modal logic because it only has one subalgebra, namely itself, and one non-trivial homomorphic image, given by the identity mapping onto itself, so discounting trivial algebras $V(S_2)_{SI} = \{S_2\}$. The other covering relations are determined similarly: since all these two-, three- and four-element algebras are simple - none has a non-trivial homomorphic image not isomorphic to itself - so only the subalgebra relation need be considered. Then it is not difficult to see that the above covering relations between varieties correspond to the subalgebra relations on their generating algebras.

It is of great practical and theoretical use to have a characterisation of the congruences of algebras in a given class, a description of $\mathbf{Con}(\mathbf{A})$ in terms of some internal characteristics of $\mathbf{A}$. Thus for Boolean algebras the relation between congruences and filters is well known, and for classical modal algebras we have seen that open filters play this role. This followed from our characterisation of congruences on *complemented* modal algebras, the utility of which was illustrated in the previous chapter. So there arises the question of whether such a useful characterisation is forthcoming for four-valued modal algebras in general. The answer, as far as we have been able to discern, is in the negative, but probing towards a characterisation illuminates the algebras in other ways.

One obvious approach is through the following theorem concerning $V(\mathbf{4})$, which defines the underlying non-modal four-valued logic. For F a prime filter of $\mathbf{A} \in V(\mathbf{4})$, define $\sigma(F)$ to be $\{a \in A \mid \neg a \notin F\}$ - this is easily seen to be a prime filter too. Then

Theorem 73 *Any set* Π *of prime filters of* $\mathbf{A} \in V(\mathbf{4})$ *closed under* σ *defines a congruence* θ *on* $\mathbf{A}$*, given by* $a\theta b$ *iff* $\forall F \in \Pi$*,* $a \in F$ *iff* $b \in F$*.*

Proof. This is obviously an equivalence relation, so suppose that $F \in \Pi$, $a\theta b$ and $c\theta d$. Then $a \wedge c \in F$ iff $a \in F$ and $c \in F$ iff $b \in F$ and $d \in F$ iff $b \wedge d \in F$. Also, using the fact that F is prime we have $a \vee c \in F$ iff $a \in F$ or $c \in F$ iff $b \in F$ or $d \in F$ iff $b \vee d \in F$. Finally, $\neg a \in F$ iff $a \notin \sigma(F)$ iff $b \notin \sigma(F)$ iff $\neg b \in F$. □

The idea then might be to mesh this theorem in some way with the approach to modal algebras by adding clauses to the definition of Π to account for the modal operators. But the obvious options involving μ- and ν-openness only seem to work for the algebras of sublogics where axioms are imposed to fit the definition. The modal algebras based on **4**, taking their two prime filters $\{a, 1\}$ and $\{b, 1\}$, provide counterexamples to most of the attempts. As an example of a localised characterisation of congruences, Loureiro[1] has shown that in $V(S_{10})$ any congruence θ may be characterised by the prime filters containing $\{a \mid a\theta 1\}$, and that any family of prime filters the intersection of which is an open filter determines a congruence. To see why this is possible, observe that the crucial validities of $V(S_{10})$ are $\mu x = \nu x$, $x \vee \neg\nu x = 1$ and $x \wedge \neg\nu x = x \wedge \neg x$. These allow the inference from $a \in F$ iff $b \in F$ to $\nu a \in F$ iff $\nu b \in F$ as follows. Suppose the former and let $\nu a \in F$; then $\neg\nu a \notin \sigma(F)$, and because $\sigma(F)$ is prime and $a \vee \neg\nu a \in \sigma(F)$ we have $a \in \sigma(F)$ and so $b \in \sigma(F)$. Because $\neg\nu a \notin \sigma(F)$, also $a \wedge \neg\nu a = a \wedge \neg a \notin \sigma(F)$, giving $\neg a \notin \sigma(F)$ and so $\neg b \notin \sigma(F)$. But then $b \wedge \neg b = b \wedge \neg\nu b \notin \sigma(F)$ and $b \in \sigma(F)$; so $\neg\nu b \notin \sigma(F)$ and $\nu b \in F$. Therefore it would seem that such a set of axioms proves to be essential in extending this use of such a family of prime filters to a system of four-valued modal logic.

---

[1]Loureiro 1985.

So the algebras appear too weak to support any general internal characterisation; if the congruence extension property were applicable, for example, there might be a fruitful approach to the problem using the fact that for any algebra $\mathbf{A}$, $(\mathbf{A}_+)^+$ is complemented and so its congruences are characterisable. In the next chapter we look at something along those lines, but we begin this chapter by showing that *principal* congruences can be defined for the varieties of nearly all of the above algebras based on $\mathbf{4}$, congruences of the form $\theta_{(a,b)}$. This, it should be recalled, is the least congruence ψ such that $a\psi b$.

The following generalises a proof by Loureiro which covered only $V(S_{10})$, algebras which she called *tetravalent modal algebras.* The greatest difference is the fact that $V(S_{10})$ has in effect only one modal operator whereas, of course, we have in general two. S_1 is the only algebra based on $\mathbf{4}$ to be omitted from the following proof, which therefore pertains to any algebra in $V(\{S_2, \ldots S_{10}\})$. The result follows from the fact that this is a so-called *discriminator* variety.

Definition 74 *A discriminator term t for an algebra $\mathbf{A}$ is a term representing the ternary function*

$$t(a,b,c) = \begin{cases} a & \text{if } a \neq b \\ c & \text{if } a = b \end{cases}$$

If a set K of algebras has a common discriminator term, then $V(K)$ is a discriminator variety. □

This is an extremely powerful notion, entailing many more properties than shall be examined here, but first it must be shown that $S_2, \ldots S_{10}$ have a common discriminator term.

First define $d(x) = (\nu x \wedge \neg\mu\neg x) \vee (\mu x \wedge \neg\nu\neg x)$. Also let $\triangle^0 = x$, and for $n \geq 1$, $\triangle^n x = \nu\triangle^{n-1}x \wedge \mu\triangle^{n-1}x$. $\triangle^1 x$ will be written as $\triangle x$. Then we have

Theorem 75 *For $x \in S_i$, $2 \leq i \leq 10$,*

$$d(x) = \begin{cases} 1 & \text{iff } x = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The four possible cases are as follows.

1. $x = 1$. Then because $\mu 1 = \nu 1 = 1$, we have $d(1) = \neg\mu 0 \vee \neg\nu 0 = \neg\Delta 0$. But since in $S_2, \ldots S_{10}$ it can be seen that $\Delta 0 = 0$, we have $d(1) = 1$.

2. $x = 0$. $\neg\mu\neg 0 = \neg\nu\neg 0 = 0$, so $d(0) = 0$.

3. $x = a$. There are the following subcases.

   - If $\nu a = 1$, then $\nu a = \mu a = \nu\neg a = \mu\neg a = 1$. So $d(a) = (1\wedge 0)\vee(1\wedge 0) = 0$.
   - If $\nu a = a$, then $\nu a = \nu\neg a = a$ and $\mu a = \mu\neg a = b$, so $d(a) = (a \wedge \neg b) \vee (a \wedge \neg b) = 0 \vee 0 = 0$.
   - If $\nu a = b$, then $\nu a = \nu\neg a = b$ and $\mu a = \mu\neg a = a$, so $d(a) = (b \wedge \neg a) \vee (b \wedge \neg a) = 0 \vee 0 = 0$.
   - If $\nu a = 0$, then $\nu a = \nu\neg a = \mu a = \mu\neg a = 0$, so $d(a) = (0\wedge 1)\vee(0\wedge 1) = 0$.

4. The case for $x = b$ is exactly the same as for $x = a$. □

Now define $x \dagger y$ to be the term $(d(x\wedge y)\vee\neg d(x\vee y))\wedge(d(\neg x\wedge\neg y)\vee\neg d(\neg x\vee\neg y))$. Then we can show

Theorem 76 *Let x, $y \in S_i$, $2 \leq i \leq 10$. Then*

$$x \dagger y = \begin{cases} 1 & \textit{if } x = y \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. $x \dagger x = (d(x) \vee \neg d(x)) \wedge (d(\neg x) \vee \neg d(\neg x))$. For $x = 1$ this is $(1 \vee 0) \wedge (0 \vee 1) = 1$; if $x = 0$ this is $(0 \vee 1) \wedge (1 \vee 0) = 1$; and if $x = a$ or $x = b$ then this is $(0 \vee 1) \wedge (0 \vee 1) = 1$. On the other hand suppose that $x \neq y$. Because, as is easily seen, $x \dagger y = \neg x \dagger \neg y$, the only cases that need to be checked are $0 \dagger 1$, $a \dagger 1$, $b \dagger 1$ and $a \dagger b$. Covering the first three cases, we have

$x \dagger 1 = d(x) \wedge \neg d(\neg x)$ following straightforwardly from the definition; but if $x \neq 1$ we know that $d(x) = 0$, and so $d(x) \wedge \neg d(\neg x) = 0$, proving these three cases. Finally, to show that $a \dagger b = 0$ first observe that $a \wedge b = \neg a \wedge \neg b = 0$ and $d(0) = 0$. So $a \dagger b = \neg d(a \vee b) \wedge \neg d(\neg a \vee \neg b) = \neg d(1) = 0$, and the last case holds. $\square$

Finally define $t(x, y, z) = ((x \dagger y) \wedge z) \vee (\neg(x \dagger y) \wedge x)$.

Theorem 77 *t is a common discriminator term for the algebras $S_2, \ldots S_{10}$.*

Proof. Let $x = y$. Then $t(x, y, z) = (1 \wedge z) \vee (0 \wedge x) = z$. Otherwise $x \neq y$, and so $t(x, y, z) = (0 \wedge z) \vee (1 \wedge x) = x$. $\square$.

A result about discriminator varieties concerning principal congruences which is of interest here is as follows. Any $\mathbf{A} \in V(\{S_2, \ldots S_{10}\})$ is a subdirect product of subdirectly irreducible algebras. $V(\{S_2, \ldots S_{10}\})_{SI}$ consists of $S_2, \ldots S_{10}$ themselves and their subalgebras based on **2** and **3** which were given above. The discriminator term t is clearly also a discriminator term for these subalgebras, so we may regard the subdirect representation of $\mathbf{A}$ as in fact presenting $\mathbf{A}$ as a *subalgebra* of a product of algebras for which t is a discriminator term. With this in mind it can be shown that

Theorem 78 *Let $\mathbf{A} \in V(\{S_2, \ldots S_{10}\})$ be given as a subdirect product of $\{B_i \mid i \in I\}$ where each B_i is a subalgebra of some S_j, $2 \leq j \leq 10$, and let $a, b, c, d \in A$. Then*

$$\theta_{(a,b)} = \{< c, d >\mid \forall i \in I\ (a(i) = b(i) \Rightarrow c(i) = d(i))\}.$$

Proof. It is straightforward to check that the righthand side is a congruence containing $< a, b >$. So suppose that $\forall i \in I(a(i) = b(i) \Rightarrow c(i) = d(i))$. Also, define $s(a, b, c, d) = t(t(a, b, c), t(a, b, d), d)$. s is known as a *switching* term, and it is not difficult to see that in B_i

$$s(a, b, c, d) = \begin{cases} c & \text{if } a = b \\ d & \text{if } a \neq b. \end{cases}$$

Now obviously we have $< s(a,a,c,d), s(a,b,c,d) >\in \theta_{(a,b)}$, so we need only show that $s(a,a,c,d) = c$ and $s(a,b,c,d) = d$. The first is obvious, given the definition of s and the fact that t is a discriminator term for B_i: for any i, $s(a,a,c,d)(i) = s(a(i),a(i),c(i),d(i)) = c(i)$. For the second, at any i either $a(i) \neq b(i)$ or $a(i) = b(i)$. If $a(i) \neq b(i)$ then by definition of s and the fact that t is a discriminator term for B_i we have $s(a,b,c,d)(i) = s(a(i),b(i),c(i),d(i)) = d(i)$, as required. Otherwise $a(i) = b(i)$ and so by hypothesis $c(i) = d(i)$; so $s(a,b,c,d)(i) = s(a(i),b(i),c(i),d(i)) = c(i) = d(i)$. This shows that $< c,d >\in \theta_{(a,b)}$, completing the proof. □

This leads to another general characterisation of principal congruences in discriminator varieties.

Theorem 79 *Let* $\mathbf{A}, a, b, c, d$ *be as above. Then* $< c,d >\in \theta_{(a,b)}$ *iff* $t(a,b,c) = t(a,b,d)$.

Proof. The 'if' part of the proof is obvious: assuming that $t(a,b,c) = t(a,b,d)$, if $a = b$ then $c = d$ since t is a discriminator term, and so $< c,d >\in \theta_{(a,b)}$. Otherwise suppose that $t(a,b,c) = t(a,b,d)$, $a \neq b$ and $a\theta b$. Then

$$c = t(a,a,c)\theta t(a,b,c) = t(a,b,d)\theta t(a,a,d) = d.$$

For the other direction suppose for contradiction that $< c,d >\in \theta_{(a,b)}$ and $t(a,b,c) \neq t(a,b,d)$ because, say, $t(a,b,c)(i) \neq t(a,b,d)(i)$. So t is a discriminator term for B_i and $t(a(i),b(i),c(i)) \neq t(a(i),b(i),d(i))$. Now, either $a(i) \neq b(i)$ or $a(i) = b(i)$. But if $a(i) \neq b(i)$ then $t(a(i),b(i),c(i)) = t(a(i),b(i),d(i)) = a(i)$, contradicting the hypothesis; and if $a(i) = b(i)$ we have $t(a(i),b(i),c(i)) = c(i)$ and $t(a(i),b(i),d(i)) = d(i)$, and so $c(i) \neq d(i)$. But by hypothesis $< c,d >\in \theta_{(a,b)}$ and so $\forall i(a(i) = b(i) \Rightarrow c(i) = d(i))$; but this implies that $a(i) \neq b(i)$, again contradicting the hypothesis. □

By considering the generating algebras and previous proofs it is immediate that in $V(\{S_2, \ldots S_{10}\})$ $a \dagger b$ and $\neg(a \dagger b)$ are complements, and the equation

$\triangle(a \dagger b) = a \dagger b$ is valid. This means that for $\mathbf{A} \in V(\{S_2, \ldots S_{10}\})$ and $a \dagger b \in A$, $[a \dagger b)$ is a strongly $\neg$-consistent and open filter if $\mathbf{A}$ is complemented. Even if $\mathbf{A}$ is not complemented, such filters can be seen to characterise principal congruences.

Theorem 80 *Let* $\mathbf{A} \in V(\{S_2, \ldots S_{10}\})$ *and* $a, b \in A$. *Then* $\theta_{(a,b)} = \theta_{(a \dagger b, 1)}$.

Proof.

1. $< a \dagger b, 1 > \in \theta_{(a,b)}$. We show that $t(a, b, a \dagger b) = t(a, b, 1)$.

   $t(a, b, a \dagger b) = (a \dagger b \wedge a \dagger b) \vee (\neg(a \dagger b) \wedge a) = (a \dagger b \wedge 1) \vee (\neg(a \dagger b) \wedge a) = t(a, b, 1)$.

2. $< a, b > \in \theta_{(a \dagger b, 1)}$. Giving $\mathbf{A}$ its subdirect representation by subalgebras $\{B_i \mid i \in I\}$ of $S_j, 2 \leq j \leq 10$, this follows from the fact that if $(a \dagger b)(i) = 1$ then $a(i) = b(i)$. For suppose that $a(i) \neq b(i)$: then $(a \dagger b)(i) = a(i) \dagger b(i) = 0 \neq 1$. □

This shows that any principal congruence may be characterised by an explicit filter of the form $[a \dagger b)$. Since $V(\{S_2, \ldots S_{10}\})$ is a discriminator variety it is endowed with many strong properties, amongst which are the useful ones possessed by classical modal algebras such as the congruence extension property and *congruence-permutability*:

$$\{< a, b > \mid \exists c \; a\theta c \text{ and } c\psi b\} \text{ is the join of the congruences } \theta \text{ and } \psi.$$

A variety in which every algebra is congruence-distributive and congruence-permutable is known as an *arithmetical* variety, and this is one of the consequences the subdirectly irreducible algebras possessing a common discriminator term.

6.2 Post Complete Logics

Now we further illustrate the fact that logics based on **4** and its subalgebras do not play the same important role played by the logics based on **2** in classical modal logic by showing that there are varieties which do not contain any of these algebras. This follows from the fact that there are at least $\aleph_0$ Post complete logics, which are now explicitly described. For $n > 2$, define a frame F_n as follows.

Definition 81 *F_n is the frame with universe $\{1^*, 1, 2, \ldots n\}$ with $1 \neq 1^*$, and with $*$ defined by $(1^*)^* = 1$ and for $m \notin \{1, 1^*\}$, $m^* = m$. The relations are given by letting R be the set $\{< m, m+1 >, < m+1, m >| \; 1 \leq m < n\}$.* □

Note that nothing is R-accessible from 1^*. It is now shown that $V(F_n^+)$ is an atom in the lattice of varieties, and so that $L(F_n^+)$ is a Post complete logic.

Theorem 82 *F_n^+ has no proper subalgebras.*

Proof. This is shown in the by now familiar fashion: we construct a 0-ary term $\underline{m}$ for each $\{m\} \in F_n^+$. Let $\underline{1} = \mu 0$ and $\underline{1^*} = \nu 0$. It is easily seen that these satisfy the requirements. Also, let $\underline{2} = \neg\nu\neg\nu 0$: then $\underline{2} = \{2\}$ can be checked to be true, because 2 is the only element of the frame from which 1 is R-accessible and $\neg\nu 0 = -\{1\}$. For $m \geq 2$ we have $\underline{m+1} = \neg\nu\neg\underline{m} \wedge \neg\underline{m-1}$: in the typical case $\neg\nu\neg\underline{m}$ is $\{m-1, m+1\}$ and $\neg\underline{m-1}$ is $-\{m-1\}$; and for the atypical case of $m = 2$ the equation can also be seen to hold, for then $\neg\underline{1} = -\{1^*\}$ and $\neg\nu\neg\underline{2} = \{1^*, 3\}$. This shows that F_n^+ is 0-generated, and so has no proper subalgebras. □

Theorem 83 *F_n^+ is simple.*

Proof. Suppose otherwise, say because $\Delta < \theta < \nabla$. Then surely for some co-atom x of F_n^+ we have $x\theta 1$. If $x = \neg\mu 0$ then $\mu 0\theta 0$ and $-\mu 0 = \neg\nu 0$ so $\neg\nu 0\theta 1$. So it may be assumed that x is of the form $F_n^+\backslash\{m\}$, for some $m \leq n$ and $m \neq 1^*$. And if $x\theta 1$, then we have $\Delta^n x\theta 1$. But for all $j \leq n$ there is an R-path in fewer that n steps from j to m; and 1^* has an n-step path with an initial R'-step followed by $n-1$ R-steps. So $\Delta^n x = 0$ and $0\theta 1$ so $\theta = \nabla$. This shows that F_n^+ is simple. □

Corollary 84 *Apart from the one-element algebra,* $V(F_n^+)_{SI} = \{F_n\}$. □

And so

Corollary 85 $V(F_n^+)$ *is an atom in the lattice of modal varieties.* □

If an axiomatic expression of the differences between between the varieties $V(F_n^+)$ for $n < \omega$ is wanted, first introduce a falsity constant $\perp =_{def.} \Box p_i \wedge \neg \bigcirc p_i$ to the language and let $\underline{m}$ equally denote the sentence whose translation is this algebraic term. Then by considering the proof that F_n^+ is simple it is not too difficult to see that $F_n^+ \models \neg\underline{m}$ iff $n < m$ and $F_n^+ \models \neg\underline{m} \vee \neg\nu\neg\underline{m+1}$ if $n > m$. So $F_n^+ \not\models \neg\underline{m} \vee \neg\nu\neg\underline{m+1}$ iff $m \neq n$ providing, for each of $F_m^+ \not\cong F_n^+$, a theorem valid there but not in the other. Familiarity with classical modal logic might initially have led one to believe that the frames must collapse onto some other frames such as the one point reflexive frame, but the fact that $1 \neq 1^*$ and that nothing is R-accessible from 1^* prevents this happening. It only seems odd because if a classical frame contains more than one element and *every* element in the frame generates the whole frame, then it has the one point reflexive frame as a p-morphic image. This need not be the case for frames for four-valued modal logic. Such familiarity should not, however, have tempted one to think that after introducing this constant for falsity we can obtain the essentially classical result that for any variable free formula α, either α or $\neg\alpha$ is a theorem of any Post complete logic: negation plays a very different role here.

A trivial consequence of the above is the following theorem.

Theorem 86 *For any natural number n there are $\aleph_0$ logics with precisely n Post complete extensions.*

Proof. There are $\aleph_0$ ways of choosing n of the varieties $\{V(F_n^+) \mid n < \omega\}$. For each distinct choice take the join of those varieties; then the fact that $(V \vee V')_{SI} = V_{SI} \cup V'_{SI}$ shows that the joins of distinct choices are themselves distinct. □

We now show that any finite join of elements of $\{V(F_n^+) \mid n < \omega\}$ is a discriminator variety and so, given the concomitant properties, we describe an extremely well behaved sublattice of the lattice of modal logics. The method of construction of the discriminator term is of necessity different to that used in the previous section. This time the key is to show that complementation is expressible as a term, and to exploit the fact that the complement of a set of frame elements differs from its negation, if at all, only with regard to their containing 1 or 1^*. Define

$$d_m^\mu(x) = \mu\Delta^m\neg x \wedge \neg\mu\Delta^m\neg x; \quad d_m^\nu(x) = \nu\Delta^m\neg x \wedge \neg\nu\Delta^m\neg x.$$

Then for $x \in F_n^+$ with $n \leq m$, as was seen in the proof that F_n^+ is simple,

$$\Delta^m x = \begin{cases} 0 & \text{if } x \neq 1 \\ 1 & \text{otherwise.} \end{cases}$$

So we have

Theorem 87 *Let $x \in F_n^+$ and $n \leq m$. Then*

$$d_m^\mu(x) = \begin{cases} 0 & \textit{iff } x = 0 \\ \mu 0 & \textit{otherwise.} \end{cases}$$

$$d_m^\nu(x) = \begin{cases} 0 & \textit{iff } x = 0 \\ \nu 0 & \textit{otherwise.} \end{cases}$$

Proof. This follows from the above, together with the facts that $x \neq 0$ iff $\neg x \neq 1$ and that in F_n^+ we have $\mu 0 \leq \neg\mu 0$ and $\nu 0 \leq \neg\nu 0$. □

Corollary 88 *With x, m and n as above,*

$$d_m^\mu(x \wedge \nu 0) = \begin{cases} 0 & \textit{iff } \nu 0 \not\leq x \\ \mu 0 & \textit{otherwise.} \end{cases}$$

$$d_m^\nu(x \wedge \mu 0) = \begin{cases} 0 & \textit{iff } \mu 0 \not\leq x \\ \nu 0 & \textit{otherwise.} \end{cases}$$

□

This allows us to find an expression for the complement of x. Consideration of the frame F_n show that intuitively this may be given as removing $\{1, 1^*\}$ from $\neg x$; then adding to this 1 if $1 \notin x$; and 1^* if $1^* \notin x$. And because $1 \notin x$ iff $\nu 0 \leq \neg x$ and $1^* \notin x$ iff $\mu 0 \leq \neg x$, we have

Theorem 89 *For $x \in F_n^+$ and $n \leq m$, the complement of x is*

$$\neg(x \vee \nu x \vee \mu x) \vee d_m^\mu(\neg x \wedge \nu 0) \vee d_m^\nu(\neg x \wedge \mu 0).$$

□

And extending to a number of algebras:

Theorem 90 *For any finite set $\{F_{k1}^+, \ldots F_{kn}^+\}$ there is a common term expressing complement.*

Proof. Use d_m^μ and d_m^ν, where $m = \bigvee\{k_i \mid 1 \leq i \leq n\}$. □

This means that the complementation operation $-$ is definable in the variety $V(\{F_{k1}^+, \ldots F_{kn}^+\})$; and consequently so is $+$. Now $+$ is associative - $(x+y)+z = x+(y+z)$; and also we have $x + x = 1$ and $x + 1 = x$. So the ternary term $m(x,y,z) =_{def.} x + y + z$ is what is known as a *Mal'cev term* for the variety

$V(\{F_{k1}^+, \dots F_{kn}^+\})$. This simply means that the variety satisfies the equations $m(x,x,y) = y$ and $m(x,y,y) = x$; but it is known from universal algebra that a variety has a Mal'cev term iff it is congruence-permutable. And since all varieties of modal algebras are congruence-distributive, it follows that

Theorem 91 *Any finite join of the varieties* $\{V(F_n^+) \mid n < \omega\}$ *is arithmetical.*
□

Theoretically, it is known that if **A** is finite and simple and $V(\mathbf{A})$ is arithmetical, then $V(\mathbf{A})$ is a discriminator variety; so this may be applied to $V(F_n^+)$. But the explicit construction of a discriminator term can be used to show that finite joins of such varieties are themselves discriminator varieties. As with the construction of a complementation term, we use $m = \bigvee\{k_i \mid 1 \le i \le n\}$ for the variety $V(\{F_{k1}^+, \dots F_{kn}^+\})$. As we have seen, for $x \in F_{ki}^+$ $\Delta^m x = 0$ iff $x \ne 1$, $\Delta^m x = 1$ otherwise; and also $x + y = 1$ iff $x = y$. This shows that

Theorem 92

$$t(x,y,z) =_{def.} (\Delta^m(x+y) \wedge z) \vee (\neg\Delta^m(x+y) \wedge x)$$

is a common discriminator term for $\{F_{k1}^+, \dots, F_{kn}^+\}$. □

So finite joins of $\{F_n^+ \mid n < \omega\}$ are also discriminator varieties. It is readily seen that in any algebra **A** in this variety $< c, d > \in \theta_{(a,b)}$, iff $\Delta^m(a+b) \le c + d$. But we have already defined a term for complement, so **A** is complemented and this use of the discriminator term to characterise principal congruences lacks the import of that in the previous section; it could have been deduced rather straightforwardly using the characterisation of congruences in complemented algebras. This does however differ from the general case because here we have a *term* for complement, and so an explicit description of the filter.

6.3 Splitting The Lattice

Continuing to concentrate on describing the bottom part of the lattice of modal logics, in this section we describe and axiomatise an infinite chain of logics which are said to *split* the lattice, and use the examples to illustrate more general properties of logics. A pair $< a, b >$ of elements of a lattice L are known as a *splitting pair* if $a \not\leq b$, and for any $c \in L$ either $a \leq c$ or $c \leq b$. Any splitting pair is uniquely determined by either of its components, so we may call an algebra $\mathbf{A}$ a *splitting algebra* if there is a variety V' such that $< V(\mathbf{A}), V' >$ is a splitting pair for the lattice of varieties of modal logics. For example, at the beginning of the chapter we in effect saw that $\underline{\mathbf{2}}^+$ is a splitting algebra for the subvariety of classical modal logics: if $\underline{\mathbf{2}}^+ \notin V(\mathbf{B})$, then $\underline{\mathbf{2}} \in V(\mathbf{B})$. But there is a least classical logic with with $L(\underline{\mathbf{2}})$ as its only Post complete extension, which is defined algebraically by the equations $\neg\nu 0 = 1$ and $x \vee \neg x = 1$. Any classical variety not satisfying $\neg\nu 0 = 1$ contains $\underline{\mathbf{2}}^+$, from which the result readily follows. $V(\underline{\mathbf{2}}^+)$ will also be seen to split the lattice of varieties of modal algebras in general, but the second component of the splitting pair is different to that in the classical case.

For $n \geq 1$ define a frame G_n in the following way.

Definition 93 *G_n is the frame with universe $\{1, \ldots n, 1^*, \ldots (n-1)^*\}$: so for $m < n$ we have $(m^*)^* = m$ and $m^* \neq m$. However, we stipulate that $n^* = n$. The relations are defined by having $mRm+1$, for $m < n$.* □

So for any frame element x other than n, x either has a unique R-successor or has a unique R'-successor, but not both. In the usual fashion, it can be shown that

Theorem 94 *G_n^+ is 0-generated.*

Proof. $\mu 0 = \{1, \ldots n-1, n\}$ and $\nu 0 = \{1^*, \ldots (n-1)^*, n\}$, so $\triangle 0 = \{n\}$. It is straightforward to show that

$$\triangle^m 0 \wedge \neg \triangle^{m-1} 0 \wedge \nu 0 = \{((n+1)-m)^*\} \text{ and } \triangle^m 0 \wedge \neg \triangle^{m-1} 0 \wedge \mu 0 = \{(n+1)-m\},$$

completing the proof. □

And furthermore

Theorem 95 *G_n^+ is subdirectly irreducible.*

Proof. Take any $\theta \neq \Delta$ on G_n^+: then for some atom $\{m\}$ with $m \leq n$ we surely have $\{m\}\theta 0$ - because $\{m^*\}\theta 0$ iff $-\neg\{m\} = \{m\}\theta 0$, this assumption is justified. So $-\{m\}\theta 1$ and $\nu^{m-1}(-\{m\})\theta 1$. But it is easy to see that $\nu^{m-1}(-\{m\}) = -\{1\}$, and so $\theta_{(\{1\},0)} \leq \theta$. □

This proof in fact shows that if $\{m\}\theta 0$ in G_n^+, then for any $j \leq m$ we also have $\{j\}\theta 0$. Also it can be shown that for $n > 1$, $G_n^+/\theta_{(\{1\},0)} \cong G_{n-1}^+$, so

Corollary 96 $\mathbf{Con}(G_n^+)$ *is an $n+1$ element chain: the quotients of congruences other than Δ and ∇ are isomorphic to the algebras $\{G_m^+ \mid 1 \leq m < n\}$. Thus because G_n^+ has no proper subalgebras, $V(G_n^+)_{SI} = \{G_m^+ \mid 1 \leq m \leq n\}$ and $L(G_1^+) = L(\underline{2}^+)$ is the only Post complete extension of $L(G_n^+)$. Indeed, the only consistent extensions of $L(G_n^+)$ are the logics $\{L(G_m^+) \mid m \leq n\}$.* □

Now because for varieties V and $\{W_i \mid i \in I\}$ we have

$$(V \vee \bigwedge_{i \in I} W_i)_{SI} = V_{SI} \cup \bigcap_{i \in I} (W_i)_{SI}$$

the distributivity of this set representation can be used to show that in general

$$V \vee \bigwedge_{i \in I} W_i = \bigwedge_{i \in I} (V \vee W_i).$$

The dual theorem with meet and join interchanged need not hold, however, and $\bigvee_{n<\omega} V(G_n^+)$ provides a good counterexample: its logic has more than one Post complete extension. To see this first consider $\prod_{n<\omega} G_n^+$, which clearly belongs to this variety. Let $\triangle 0$ in $\prod_{n<\omega} G_n^+$ be the element which is constantly $\triangle 0$ at each n, and $\theta_{(\triangle 0,0)}$ be the least congruence such that $\triangle 0 \theta 0$; the reader may check that $\theta_{(\triangle 0,0)} \neq \nabla$. The result follows from

Theorem 97

$$S_2 \in S(\prod_{n<\omega} G_n^+/\theta_{(\triangle 0,0)}).$$

Proof. First observe that in each G_n^+ the following equations can be verified to hold:

1. $\nu 0 \vee \mu 0 = 1$. This is because $\nu 0 = \{1^*, \ldots (n-1)^*, n\}$ and $\mu 0 = \{1, \ldots n\}$.

2. $\nu^2 0 = \neg \nu^2 0 \vee \triangle^2 0$. This is clear if $\nu^2 0 = 1$, for then also $\triangle^2 0 = 1$. Otherwise $\nu^2 0 = \{1^*, \ldots (n-1)^*, n\} \cup \{n-1\}$, $\neg \nu^2 0 = \{1^*, \ldots (n-2)^*\}$ and $\triangle^2 0 = \{(n-1)^*, n, n-1\}$, from which the equation can be seen to hold.

3. $\mu\mu 0 = \nu\mu 0 = \mu\nu 0 = 1$. These equations are immediate.

So they are also valid in $\prod_{n<\omega} G_n^+/\theta_{(\triangle 0,0)}$. But here we also have $\mu 0 \wedge \nu 0 = 0$, and so, given 1 above, $\mu 0$ and $\nu 0$ are Boolean complements, from which it follows that $\neg \nu 0 = \nu 0$ and $\neg \mu 0 = \mu 0$. Note that because the algebra is non-trivial, this means that these two elements must be distinct from 0 and 1 and so also distinct from one another. Now consider $\nu^2 0 = \neg \nu^2 0 \vee \triangle^2 0$. Because $\triangle 0 = 0$ we also have $\triangle^2 0 = 0$, so $\nu^2 0 = \neg \nu^2 0$. But in any algebra we have $\nu 0 \leq \nu^2 0$ and $\neg \nu^2 0 \leq \neg \nu 0$; so here $\nu 0 \leq \nu^2 0 \leq \neg \nu 0$. But $\neg \nu 0 = \nu 0$ so $\nu^2 0 = \nu 0$. This shows that the four elements $\{0, \nu 0, \mu 0, 1\}$ are closed under all the operations: but inspection shows that this least subalgebra is isomorphic to S_2, completing the proof. $\square$

This shows that $S_2 \in \bigvee_{n<\omega} V(G_n^+)$, and so $V(S_2) \leq \bigvee_{n<\omega} V(G_n^+)$; while for each n, $V(S_2) \wedge V(G_n^+) = \underline{\mathbf{T}}$, and so $\bigvee_{n<\omega}(V(S_2) \wedge V(G_n^+)) = \underline{\mathbf{T}}$.

We now assume familiarity with the notion of a free algebra, and for convenience introduce a constant $\perp$ for falsity into the language - say, $\perp =_{def.} \Box p_i \wedge \neg \bigcirc p_i$. Its algebraic interpretation is obviously the bottom element 0. There can be constructed what is known as a *finite presentation* of G_n^+ in 0 variables in the variety of modal algebras. Firstly, because G_n^+ is finite and $V(G_n^+)$ is congruence-distributive, it is known that there is guaranteed to be amongst the valid variable-free equations of G_n^+ a finite subset of these from which all the others may be derived. And because each element of the algebra is complemented and there is a constant for each element, we may presume that each such equation $x = y$ is rewritten as $(x+y) \wedge (\neg x + \neg y) = 1$. Since there are only a finite number of such terms $(x+y) \wedge (\neg x + \neg y)$, their conjunction is also a term: let $\delta_{G_n^+}$ be the *sentence* of the modal language which corresponds in the natural way to this term. Now, $\delta_{G_n^+}$ defines a variety in the obvious way - the algebras in which it is valid - and G_n^+ is isomorphic to the free algebra in 0 variables in this variety: there are no variable-free validities of G_n^+ which are not derivable from $\delta_{G_n^+}$, using the rules of the basic logic $\vdash_K$. So G_n^+ can in effect be regarded as the Lindenbaum algebra in the variable-free language for the logic given by extending $\vdash_K$ by the axiom $\delta_{G_n^+}$. Or algebraically, if $\delta_{\mathbf{G_n^+}}$ denotes the 0-ary term that is the translation of $\delta_{G_n^+}$, and if $\mathbf{F}_{\vdash_K}(\emptyset)$ is the free four-valued modal algebra on 0 generators, then

$$G_n^+ \cong \mathbf{F}_{\vdash_K}(\emptyset)/\theta_{(\delta_{\mathbf{G_n^+}},1)}.$$

Because we are presenting G_n^+ in a complete variety - the variety of all modal algebras - there is an obvious equivalence between what is said of $\delta_{\mathbf{G_n^+}}$ and what is said of $\delta_{G_n^+}$.

Recalling the definition of ∇x, it is not too difficult to see that it could equally well be defined as $\nabla^0 x = x$, and $\nabla^n x = \nabla^{n-1} x \wedge \triangle^n x$. We shall be free in using the notation $\triangle$ and ∇ to abbreviate both algebraic operators and the corresponding language connectives. Now in any algebra we have

Theorem 98 $-\neg\nabla^n \delta_{\mathbf{G_n^+}} = \nabla^n \delta_{\mathbf{G_n^+}}$.

Proof. First, $-\neg\delta_{\mathbf{G}_n^+} = \delta_{\mathbf{G}_n^+}$. Because $\neg\delta_{\mathbf{G}_n^+}$ is a 0-ary or variable-free term it has a complement; and using the facts that $-\neg(a \wedge b) = -\neg a \wedge -\neg b$, $-\neg(a+b) = \neg a + \neg b$ and that $\neg a + \neg b$ was stipulated to be a conjunct of $\delta_{\mathbf{G}_n^+}$ if $a+b$ was one, this result then follows. In general, for $n \geq 1$ we have $-\neg\triangle^n x = \triangle^n x$: this is because $-\neg\nu z = \mu z$ and $-\neg\mu z = \nu z$, and νz is a conjunct of $\triangle^n x$ just in case μz is. The result then follows from the fact that

$$\nabla^n\delta_{\mathbf{G}_n^+} = \delta_{\mathbf{G}_n^+} \wedge \triangle^1\delta_{\mathbf{G}_n^+} \wedge \ldots \wedge \triangle^n\delta_{\mathbf{G}_n^+}.$$

□

Corollary 99 $\vdash_K \neg\nabla^n\delta_{G_n^+} \vee \nabla^n\delta_{G_n^+}$.

Proof. This is the corresponding sequent, and the logic is complete. □

The same property obviously holds of $\triangle^n 0$ - that is, $-\neg\triangle^n 0 = \triangle^n 0$ - and so of the meet $\gamma_{\mathbf{G}_n^+} =_{def.} \triangle^n 0 \wedge \nabla^n\delta_{\mathbf{G}_n^+}$. Also define $\gamma_{G_n^+}$ similarly as the conjunction $\triangle^n\bot \wedge \nabla^n\delta_{G_n^+}$.

This shows half of

Theorem 100 *In any 0-generated modal algebra* $\mathbf{A}$, $[\gamma_{\mathbf{G}_n^+})$ *is a strongly* $\neg$-*consistent and open filter.*

Proof. Clearly all complements exist in $\mathbf{A}$. $[\gamma_{\mathbf{G}_n^+})$ is strongly $\neg$-consistent because if $\gamma_{\mathbf{G}_n^+} \leq a$ then $\gamma_{\mathbf{G}_n^+} = -\neg\gamma_{\mathbf{G}_n^+} \leq -\neg a$. To show that $[\gamma_{\mathbf{G}_n^+})$ is also open it suffices to show that $\gamma_{\mathbf{G}_n^+} \leq \triangle\gamma_{\mathbf{G}_n^+}$. Now,

$$\triangle\gamma_{\mathbf{G}_n^+} = \triangle^{n+1}0 \wedge \triangle\delta_{\mathbf{G}_n^+} \wedge \ldots \wedge \triangle^n\delta_{\mathbf{G}_n^+} \wedge \triangle^{n+1}\delta_{\mathbf{G}_n^+},$$

and because $0 \leq \delta_{\mathbf{G}_n^+}$ implies that $\triangle^{n+1}0 \leq \triangle^{n+1}\delta_{\mathbf{G}_n^+}$, it follows that

$$\triangle\gamma_{\mathbf{G}_n^+} \geq \triangle^{n+1}0 \wedge \triangle\delta_{\mathbf{G}_n^+} \wedge \ldots \wedge \triangle^n\delta_{\mathbf{G}_n^+};$$

and so

$$\triangle\gamma_{\mathbf{G}_n^+} \geq \triangle^{n+1}0 \wedge \delta_{\mathbf{G}_n^+} \wedge \triangle\delta_{\mathbf{G}_n^+} \wedge \ldots \wedge \triangle^n\delta_{\mathbf{G}_n^+}.$$

And finally, because $\triangle^{n+1}0 \geq \triangle^n 0$ we have

$$\triangle\gamma_{\mathbf{G}_n^+} \geq \triangle^n 0 \wedge \delta_{\mathbf{G}_n^+} \wedge \triangle\delta_{\mathbf{G}_n^+} \wedge \ldots \wedge \triangle^n \delta_{\mathbf{G}_n^+} = \gamma_{\mathbf{G}_n^+}.$$

□

Now let $\varepsilon_{\mathbf{G}_n^+}$ denote the equation corresponding to $\gamma_{\mathbf{G}_n^+} \leq \triangle^{n-1}0$; again let $\varepsilon_{G_n^+}$ be the corresponding sequent $\gamma_{G_n^+} \vdash \triangle^{n-1}\bot$. If we wished, this could have been rewritten as an equivalent axiom. $\varepsilon_{G_n^+}$ is not valid in G_n^+ because $G_n^+ \models \gamma_{G_n^+}$ but $G_n^+ \not\models \triangle^{n-1}\bot$. Indeed, by considering again the proof that G_n^+ is subdirectly irreducible it can be seen that the smallest non-identity congruence is in fact $\theta_{(\triangle^{n-1}0,1)}$. Now recall that $V(\varepsilon_{G_n^+})$ is the variety of algebras in which $\varepsilon_{G_n^+}$ is valid. We now show

Theorem 101 $< V(G_n^+), V(\varepsilon_{G_n^+}) >$ *splits the lattice of varieties.*

Proof. Let $\mathbf{A}$ be any algebra such that $\mathbf{A} \not\models \varepsilon_{G_n^+}$, and let $\mathbf{A}_0$ be its least subalgebra: given that $\varepsilon_{G_n^+}$ contains no variables we also have $\mathbf{A}_0 \not\models \varepsilon_{G_n^+}$, and so $\gamma_{\mathbf{G}_n^+} \not\leq \triangle^{n-1}0$ in $\mathbf{A}_0$. Because $\mathbf{A}_0$ is 0-generated it is complemented; and because $[\gamma_{\mathbf{G}_n^+})$ is a strongly $\neg$-consistent and open filter of $\mathbf{A}_0$ this defines a congruence ψ on $\mathbf{A}_0$. So for $a \in \mathbf{A}_0$ we have $a\psi 1$ iff $\gamma_{\mathbf{G}_n^+} \leq a$: this means that $\mathbf{A}_0/\psi \models \gamma_{G_n^+}$ but $\mathbf{A}_0/\psi \not\models \triangle^{n-1}\bot$. Now given the definition of $\gamma_{G_n^+}$ we have $\mathbf{A}_0/\psi \models \delta_{G_n^+}$; and $\mathbf{A}_0/\psi$ is 0-generated and G_n^+ is the free 0-generated algebra in the variety determined by $\delta_{G_n^+}$; so we have a homomorphism $G_n^+ \xrightarrow{f} \mathbf{A}_0/\psi$. Now if f were *not* injective, then because the least non-identity congruence on G_n^+ is $\theta_{(\triangle^{n-1}0,1)}$ we would have $f(\triangle^{n-1}0) = 1$. But this implies that $\mathbf{A}_0/\psi \models \triangle^{n-1}\bot$, which is not the case; so f is injective. This shows that $G_n^+ \in SHS(\mathbf{A}) \subseteq HSS(\mathbf{A}) = HS(\mathbf{A})$. In fact, G_n^+ is a homomorphic image of the least subalgebra of $\mathbf{A}$.

Now suppose for a family of varieties $\{V_i \mid i \in I\}$ we have $V(G_n^+) \leq \bigvee_{i \in I} V_i$. Because G_n^+ is subdirectly irreducible there are algebras $\{\mathbf{B}_j \mid j \in J\} \subseteq \bigcup\{V_i \mid i \in I\}$ such that $G_n^+ \in HSP_U(\{\mathbf{B}_j \mid j \in J\})$: say $\mathbf{B} \in P_U(\{\mathbf{B}_j \mid j \in J\})$ and $G_n^+ \in HS(\mathbf{B})$. Then because subalgebras and surjective homomorphisms preserve the validity of formulae and because $G_n^+ \not\models \varepsilon_{G_n^+}$, we may conclude that $\mathbf{B} \not\models \varepsilon_{G_n^+}$; and so for some $j \in J$, with, say, $k \in I$ and $\mathbf{B}_j \in V_k$, $\mathbf{B}_j \not\models \varepsilon_{G_n^+}$. But then $G_n^+ \in HS(\mathbf{B}_j)$, and so $V(G_n^+) \leq V(\mathbf{B}_j) \leq V_k$. This shows that

$$V(G_n^+) \not\leq \bigvee\{V \mid V \text{ is a variety and } V(G_n^+) \not\leq V\}.$$

And if a variety V' is such that $V(G_n^+) \not\leq V'$, then clearly $V' \leq \bigvee\{V \mid V(G_n^+) \not\leq V\}$; so $< V(G_n^+), \bigvee\{V \mid V(G_n^+) \not\leq V\} >$ is a splitting pair. Now if $\mathbf{A} \in \bigvee\{V \mid V(G_n^+) \not\leq V\}$ then $V(\mathbf{A}) \leq \bigvee\{V \mid V(G_n^+) \not\leq V\}$: suppose for contradiction that also $\mathbf{A} \not\models \varepsilon_{Gn^+}$. Then $G_n^+ \in HS(\mathbf{A})$ so $V(G_n^+) \leq V(\mathbf{A})$; but this implies that $V(G_n^+) \leq \bigvee\{V \mid V(G_n^+) \not\leq V\}$, which is false. So $\mathbf{A} \models \varepsilon_{Gn^+}$ iff $\mathbf{A} \in \bigvee\{V \mid V(G_n^+) \not\leq V\}$, and $< V(G_n^+), V(\varepsilon_{G_n^+}) >$ is a splitting pair. □

So the varieties $\{V(G_n^+) \mid 1 \leq n < \omega\}$ form an infinite chain of splitting elements for the lattice of modal logics. Note that there is no constraint that if $\mathbf{A} \models \varepsilon_{G_1^+}$ then $\mathbf{A}$ is classical, so the second component of the splitting determined by G_1^+ is different to that for the classical sublattice, even though G_1^+ is itself classical: it is simply determined by the equation $\triangle 0 = 0$.

These examples will now be used to illustrate another feature that varieties may possess, namely *canonicity*: a variety V is said to be canonical if it is generated by complemented algebras, and if $\mathbf{A} \in V$ then $(\mathbf{A}_+)^+ \in V$, where $\mathbf{A}_+$ is the *ordered* frame. It may in fact be shown that this is equivalent to requiring the second condition to hold only of complemented frames in V and so avoiding any reference to ordered frames. This follows fairly quickly from the fact that if $\mathbf{A} \in HS(\mathbf{B})$ then $(\mathbf{A}_+)^+ \in HS((\mathbf{B}_+)^+)$. The converse of second part of the definition has already been seen to hold in general, but this direction is not in general true. It will be shown by axiomatising the logics determined by these varieties, and showing that this axiomatisation corresponds to certain first-order conditions on the relations of a frame, conditions which pick out precisely the frames for the logic. The result then follows in a standard way.

For $L(G_n) = L(G_n^+)$ consider the axioms

1. $\Box\bot \vee \bigcirc\bot$;

2. $\neg\triangle\bot \vee \alpha \vee \neg\alpha$;

3. $\neg \bigcirc (\alpha \vee \beta) \vee \Box\alpha \vee \Box\beta$;

4. $\triangle^n\bot$.

The proposal is that these four sentences axiomatise the logic $L(G_n^+)$. As with $n = 1$, there may of course be some redundancy. So letting L be the logic determined by the axioms for an arbitrary n, we want to show that $L = L(G_n^+)$.

Note that for all frames $\mathbf{C}$, valuations v and worlds $x \in \mathbf{C}$ we have $\mathbf{C}, v, x \not\models \bot$, and further consider the following frame conditions, where variables range over worlds.

1. $\forall x \neg(\exists y\ xRy \ \wedge\ \exists z\ xR'z)$;

2. $\forall x\ (\neg\exists y\ xRy \ \wedge\ \neg\exists z\ xR'z \ \Rightarrow\ x = x^*)$;

3. $\forall x \forall y \forall z\ (xRy \ \wedge\ xRz \ \Rightarrow\ y = z)$;

4.
$$\begin{aligned}\forall y_1 \ldots \forall y_n \ \ &((y_1 R y_2 \ldots y_{n-1} R y_n \\ &\vee \ \ \vdots \\ &\vee\ y_1 R' y_2 \ldots y_{n-1} R' y_n) \ \Rightarrow\ \neg\exists z\ y_n R z \ \wedge \neg\exists z'\ y_n R' z').\end{aligned}$$

The antecedent of 4 is intended to be the disjunction of the 2^{n-1} possible $n-1$-step paths in the two relations R and R'. Let K be the class of frames $\mathbf{C}$ for which all four of these conditions are satisfied for the same n as was chosen for the proposed axiomatisation.

Theorem 102 *For any frame* $\mathbf{C}$, $\mathbf{C} \in K$ *iff* $\mathbf{C} \models L$.

Proof. We show that a frame satisfies one of these numbered conditions just in case it satisfies the frame condition of the same number, from which the result follows. For the first axiom observe first that for any frame **C** and world $x \in C$, $\mathbf{C}, x \models \Box\bot$ iff $\neg\exists y\ xRy$ and $\mathbf{C}, x \models \bigcirc\bot$ iff $\neg\exists z\ xR'z$ - there is no need to take account of valuations. So $\mathbf{C}, x \models \Box\bot \vee \bigcirc\bot$ iff $\neg\exists y\ xRy$ or $\neg\exists z\ xR'z$. This shows that $\mathbf{C} \models \Box\bot \vee \bigcirc\bot$ iff $\forall x\, \neg(\exists y\ xRy\ \wedge\ \exists z\ xR'z)$.

For axiom 2, it can be seen along lines similar to the above that $\mathbf{C} \not\models \neg\triangle\bot$ iff $\exists x \in C$ such that $\neg\exists y\ xRy$ and $\neg\exists z\ xR'z$. So $\mathbf{C} \models \neg\triangle\bot \vee \alpha \vee \neg\alpha$ just in case for all valuations v and worlds x, if $\neg\exists y\ xRy$ and $\neg\exists z\ xR'z$ then $\mathbf{C}, v, x \models \alpha \vee \neg\alpha$. But the consequent is true of x just in case x is classical - that is, $x = x^*$ - otherwise a valuation v with $v(\alpha) = \{x^*\}$ would provide a counterexample. And this is just what the frame condition states.

$\mathbf{C}, v, x \models \neg\bigcirc(\alpha \vee \beta)$ iff $\mathbf{C}, v, x \not\models \Box(\alpha \vee \beta)$, so axiom 3 is valid in F just in case for any such valuation v and world x $\mathbf{C}, v, x \models \Box(\alpha \vee \beta)$ implies that $\mathbf{C}, v, x \models \Box\alpha$ or $\mathbf{C}, v, x \models \Box\beta$. If this is true of **C** then there cannot be $x, y, z \in C$ with xRy, xRz and $y \neq z$: for then the valuation v with $v(\alpha) = \{y\}$ and $v(\beta) = \{z\}$ would provide a counterexample at x. So the corresponding frame condition then holds. For the other direction suppose, assuming some valuation v, that $x \models \Box(\alpha \vee \beta)$ but $x \not\models \Box\alpha$ and $x \not\models \Box\beta$, with these failing because xRy, xRz and $y \not\models \alpha$ and $z \not\models \beta$. But since $x \models \Box(\alpha \vee \beta)$ we have $y \models \alpha \vee \beta$ and so $y \models \beta$; but $z \not\models \beta$ so $y \neq z$, contradicting the frame condition.

Let Ω^n range over n-element sequents of connectives from the set $\{\Box, \bigcirc\}$: then axiom 4 is valid on a frame **C** just in case $\mathbf{C} \models \Omega^n\bot$ for all Ω^n. Now associate with Ω^n an n-tuple $\Re^n$ of relations by substituting $\Box$ for R and $\bigcirc$ for R'. $\mathbf{C} \not\models \Omega^n\bot$ iff for some Ω^n and some $x \in C$, $x, \not\models \Omega^n\bot$. Letting $\Re^n$ be the corresponding n-tuple $< S_1, \dots S_n >$ with each $S_i \in \{R, R'\}$, this can easily be seen to be so just in case there are elements $x = y_1, \dots y_n, z \in C$ and a path $x = y_1 S_1 y_2 \dots y_n S_n z$. But the frame condition 4 fails just in case there is some such path with final step R or R', and clearly S_n is one of these relations; so the axiom does define the frame condition. □

Now we show that

Theorem 103 $L(G_n^+) = L$.

Proof. First we show that $L \subseteq L(G_n^+)$. From the above result this can be seen to follow if $G_n \in K$. It has already been shown that the first axiom holds in G_n^+ and therefore in G_n; so G_n satisfies the first frame condition. For the second condition, the only world x in G_n of which the antecedent is true is n, and as required n is classical. The third condition is obviously fulfilled, because for $m < n$ we only have $mRm + 1$. And the fourth condition may be confirmed by observing that the longest paths in G_n are the two unique ones between 1 and n, and between 1^* and n, and these are the only $n-1$-step paths. In both cases substituting the element n for the variable y_n in the consequent of this fourth sentence shows it to be satisfied; so the frame G_n satisfies the condition.

For the other direction let $\mathbf{C} \in K$ be arbitrary and consider the subdirect representation by means of point-generated subframes $\mathbf{C}^+ \to \prod_{x \in C} \mathbf{C}_x^+$. Then if it can be shown that for an arbitrary $x \in C$, $\mathbf{C}_x^+ \in V(G_n^+)$, the result will follow. This is because then $L(G_n^+) \subseteq L(\mathbf{C}_x^+)$ for each $x \in X$, and so $L(G_n^+) \subseteq L(\mathbf{C})$ and $L(G_n^+) \subseteq \bigcap_{\mathbf{C} \in K} L(\mathbf{C}) = L$.

So it need only be shown that if $\mathbf{C} \in K$ is generated by a single frame element $x \in C$, then $\mathbf{C}^+ \in V(G_n^+)$. Now the fact that such a frame $\mathbf{C}$ satisfies the fourth frame condition means that for some $m \leq n$ and $S_i \in \{R, R'\}$ there is an $m-1$-step path $x = y_1 S_1 \ldots S_{m-1} y_m = z$ originating in x, and that there are no such paths of greater length. But we also know that there are no paths in the frame longer than this one: if there were, it could not originate in some y_i. If it originated in y_i^* with first step $y_i^* R u_i$ or $y_i^* R' u_i$, then have $y_i R' u_i$ or $y_i R u_i$, so this too cannot be the case. And if it originates in u_i^* with, say,

$y_i^* R' u_i$, and first step $u_i^* R u_j$, then we have $y_i R u_i R' u_j$. Continuing this process, we cover all elements of the point-generated frame **C**. Now it is readily seen that frame condition 3 also holds if R is replaced by R', so together they show that this path does not bifurcate; thus it is the only $m-1$-step path originating in x. Because **C** is point-generated condition 3 also then implies that for each $y \in C$ there is some $i \leq m$ with either $y = y_i$ or $y^* = y_i$. Now 2 tells us that the endpoint z is classical, but it is also the unique classical world in **C**. For if there were another one $z' = z'^*$ then we would have $z' = y_i$, for some $i < m$ - but this contradicts condition 1, which implies that classical worlds have no R-successors and no R'-successors. So for $i \neq m$ we have $y_i \neq y_i^*$, and apart from the classical world y_m these are the only worlds in C. So define f to be the p-morphism from **C** to G_m induced by

$$f(y_m) = m; \quad f(y_i) = \begin{cases} i & \text{if } y_i R y_{i+1} \\ i^* & \text{if } y_i R' y_{i+1}. \end{cases}$$

Both this and its inverse are readily seen to be p-morphisms, so G_m^+ is isomorphic to $\mathbf{C}^+$ by the mapping f^+. But $m \leq n$, so $G_m^+ \in V(G_n^+)$ and $\mathbf{C}^+ \in V(G_n^+)$. □

Because the logic L of G_n^+ is also the logic of a class of frames defined by a set of first-order sentences, it follows that $\mathbf{A} \in V(G_n^+)$ iff $(\mathbf{A}_+)^+ \in V(G_n^+)$. Only the left to right implication need be shown. From model theory it is known that if **C** is a frame for this logic, then so is any elementary extension of **C**; and $(\mathbf{C}^+)_+$ can be shown to be a p-morphic image of a suitably saturated elementary extension of **C**: we give a quick outline of the proof.

Theorem 104 *For any frame* **C**, *there is an elementary extension* $\mathbf{C}^\ddagger$ *of* **C** *with a surjective p-morphism* $\mathbf{C}^\ddagger \xrightarrow{\sigma} (\mathbf{C}^+)_+$.

Proof. Let $\mathbf{C}'$ be the first-order structure just as **C** except that in addition it has one-place relational constants $\underline{Y}$ for each subset Y of the universe X of **C**. Then by model theory it is known that there exists an ω-saturated elementary extension of $\mathbf{C}^{\ddagger}$ of $\mathbf{C}'$, which will be sufficient for our purposes. To show what this means, let L be the first-order language with constants for the relations of the frame as well as $\{\underline{Y} \mid Y \subseteq X\}$; furthermore for each $x_i \in X$ let $\underline{x_i}$ be an individual constant denoting x_i, and for $Z \subseteq X$ let L^Z be the augmentation of L by the constants $\{\underline{x_i} \mid x_i \in Z\}$. Then $\mathbf{C}^{\ddagger}$ has the property that if for any finite $Z \subseteq X$ and any set Γ of L^Z-formulae, if any finite $\Gamma_0 \subseteq \Gamma$ is satisfiable in $(\mathbf{C}^{\ddagger}, \{\underline{x_i} \mid x_i \in Z\})$ - in other words if Γ is ***finitely satisfiable*** there - then so is Γ itself.

Given this elementary extension, define σ by $\sigma(x) = \{Y \subseteq X \mid \underline{Y}(x)\}$, and note that the following first-order sentences of L are true in $\mathbf{C}'$ and therefore also in $\mathbf{C}^{\ddagger}$:

1. $\forall x \underline{X}(x)$.

2. $\neg \exists x \underline{\emptyset}(x)$.

3. $\forall x(\underline{Y \cap Z}(x) \Longleftrightarrow \underline{Y}(x) \wedge \underline{Z}(x))$.

4. $\forall x(\underline{Y \cup Z}(x) \Longleftrightarrow \underline{Y}(x) \vee \underline{Z}(x))$.

5. $\neg \exists x(\underline{Y}(x) \wedge \underline{-Y}(x))$.

6. $\forall x(\underline{Y}(x) \vee \underline{-Y}(x))$.

7. $\forall x(\underline{Y^*}(x) \Longleftrightarrow \underline{Y}(x^*))$.

$1 - 4$ show that $\sigma(x)$ is a prime filter of $\mathbf{C}^+$ and so that σ is well-defined; and $5 - 7$ show that in $\mathbf{C}^{\ddagger}$, $x^* \in \underline{Y}$ iff $x \notin \underline{-Y^*}$, so that $\sigma(x^*) = \sigma(x)^*$, given the definition of $*$ in $(\mathbf{C}^+)_+$.

To show that σ is surjective, let G be a prime filter of $\mathbf{C}^+$ and let Γ be the set of sentences

$$\{\underline{Y}(x) \mid Y \in G\} \cup \{\neg \underline{Y}(x) \mid Y \subseteq X \text{ and } Y \notin G\}.$$

The $\neg$ here of course belongs to the first-order language L. Then it can be seen that Γ is finitely satisfiable in $\mathbf{C}'$: take any finite subsets $I \subseteq G$ and $J \subseteq -G$ of $\mathbf{C}^+$ and consider

$$\{\underline{Y}(x) \mid Y \in I\} \cup \{\neg\underline{Y}(x) \mid Y \in J\}.$$

Because G is prime, $\bigcap I \not\subseteq \bigcup J$, and so there is some $y \in \bigcap I$ such that $y \notin \bigcup J$. So y satisfies the above set of sentences for $\mathbf{C}'$. This shows that Γ is finitely satisfiable in $\mathbf{C}'$ and so also in $\mathbf{C}^\ddagger$; but $\mathbf{C}^\ddagger$ is ω-saturated, so Γ is satisfiable in $\mathbf{C}^\ddagger$, say by x. Then $\sigma(x) = G$.

To show that if xRy then $\sigma(x)R\,\sigma(y)$, observe that if $Y \subseteq X$ and $\nu_R(Y) \in \sigma(x)$, then we have $x \in \underline{\nu_R(Y)}$. But it is easy to show that in $\mathbf{C}'$, and so in $\mathbf{C}^\ddagger$, we have

$$\forall z(\underline{\nu_R(Y)}(\underline{x}) \wedge \underline{xRz} \Longrightarrow \underline{Y}(z))$$

true. But because $\underline{\nu_R(Y)}(x)$ and xRy, it then follows that $\underline{Y}(y)$, showing that $Y \in \sigma(y)$ and $\sigma(x)R\,\sigma(y)$.

For the final condition on a p-morphism, suppose that $\sigma(x)RG$ and let $\Gamma = \{\underline{xRy}\} \cup \{\underline{Y}(y) \mid Y \in G\}$. If there were some $y \in \mathbf{C}^\ddagger$ which satisfied Γ, then we would have $G \subseteq \sigma(y)$; but clearly prime filters are discretely ordered in $\mathbf{C}^+$ so we would have $G = \sigma(y)$. So to complete the proof it need only be demonstrated that Γ is satisfied in $\mathbf{C}^\ddagger$. So take arbitrary $Y \in G$. Then $-Y \notin G$ and given that $\sigma(x)RG$ we have $x \notin \underline{\nu_R(-Y)}$ and so $\mathbf{C}^\ddagger$ satisfies the open sentence $\underline{-\nu_R(-Y)}(z)$. But it is easy to see that in $\mathbf{C}'$ and so in $\mathbf{C}^\ddagger$ the sentence

$$\forall z(\underline{-\nu_R(-Y)}(z) \Longrightarrow \exists z'(z\underline{R}z' \wedge \underline{Y}(z')))$$

is true, and so $\{\underline{xRy}, \underline{Y}(y)\}$ is satisfiable in $\mathbf{C}^\ddagger$ for arbitrary $Y \in G$. To show that Γ is satisfiable there it need only be shown to be finitely satisfiable, so for arbitrary $\{Y_1, \ldots Y_n\} \subseteq G$ let Γ_0 be the set $\{\underline{xRy}, \underline{Y_1}(y), \ldots \underline{Y_n}(y)\}$. But if we let Y denote their (finite) intersection, we have $Y \in G$ since G is a filter; and using the first part of the proof we know that $\{\underline{xRy}, \underline{Y}(y)\}$ is satisfied by some element $y \in \mathbf{C}^{\ddagger}$. This same element shows that Γ_0 is itself satisfied, completing the proof. $\square$

Now suppose that $\mathbf{A} \in V(G_n^+)$. Then $\mathbf{A} \in HS(C^+)$ for some $C \in K$ which is a disjoint union of G_n^+, which in turn implies that $(\mathbf{A}_+)^+ \in HS(((C^+)_+)^+)$. But $(C^+)_+ \in K$ so $((C^+)_+)^+ \in V(G_n^+)$, implying that $(\mathbf{A}_+)^+ \in V(G_n^+)$. This completes the proof of

Theorem 105 $\mathbf{A} \in V(G_n^+)$ *iff* $(\mathbf{A}_+)^+ \in V(G_n^+)$. $\square$

Because $V(G_n^+)$ is obviously generated by complemented algebras, we have thus shown this variety to be canonical. A standard model-theoretic argument shows that the splitting variety $V(\varepsilon_{G_n^+})$ is also canonical, for the frames *not* in K also form a basic elementary class. Thus we have an alternative description of the splitting pairs by means of their corresponding frames.

Chapter 7

Complemented Algebras

7.1 The Relation Between Algebras And Complemented Algebras

In this chapter modal algebras in general are compared with the complemented modal algebras which have been seen to have stronger algebraic properties. In particular, it is shown that any modal algebra has a free complemented completion into which it can be embedded, and some properties and consequences of this result are explored, especially with regard to areas in which complementation has proved useful. There is a parallel in the theory of distributive lattices, where any distributive lattice $\mathbf{L}$ can be embedded into a complemented distributive lattice $\mathbf{L}'$ which it generates, if complementation is regarded as an operation. This $\mathbf{L}'$ is said to to be *free* with respect to this property; moreover, $\mathbf{L}'$ is uniquely determined, and any congruence on $\mathbf{L}$ has a unique extension to a congruence on $\mathbf{L}'$. $\mathbf{L}'$ is said to be R-generated by $\mathbf{L}$, because $\mathbf{L}$ generates it as a ring. Here we explore the extent to which these results extend with the addition of $\neg$ and the modal operators.

Recall that $\mathcal{MA}$ is the category of modal algebras and homomorphisms, and further let $\mathcal{MA}^-$ be the category the objects of which are modal algebras $\mathbf{B}$ with a further unary operation $-$ such that for $b \in B$, $b \vee -b = 1$ and $b \wedge -b = 0$; the morphisms are again algebraic homomorphisms. With any $\mathbf{B} \in \mathcal{MA}^-$ we can naturally associate a modal algebra $\mathcal{G}(\mathbf{B}) \in \mathcal{MA}$ which has the same universe with the operations $\wedge, \vee, \neg, \nu, \mu, 0$ and 1 acting on it in exactly the same way, but which 'forgets' that $-$ is an operation. It should be clear that the modal algebras isomorphic to some such $\mathcal{G}(\mathbf{B})$, for $\mathbf{B} \in \mathcal{MA}^-$, are precisely the complemented modal algebras: clearly $\mathcal{G}(\mathbf{B})$ is complemented, and if $\mathbf{A} \in \mathcal{MA}$ is complemented then it admits a complementation operation, which if added then forgotten leaves $\mathbf{A}$ unchanged. And if for $\mathbf{B}_1$, $\mathbf{B}_2 \in \mathcal{MA}^-$ and $\mathbf{B}_1 \xrightarrow{f} \mathbf{B}_2$ a homomorphism, $\mathcal{G}(\mathbf{B}_1) \xrightarrow{\mathcal{G}f} \mathcal{G}(\mathbf{B}_2)$ is defined to be the same function from the universe B_1 to the universe B_2, then it is immediate that $\mathcal{G}f$ is a homomorphism and $\mathcal{G}$ satisfies the conditions for being a functor from $\mathcal{MA}^-$ to $\mathcal{MA}$: that $\mathcal{G}1_{\mathbf{B}} = 1_{\mathcal{G}(\mathbf{B})}$ and $\mathcal{G}(g \circ f) = \mathcal{G}g \circ \mathcal{G}f$. The main aim here is to show that something similar is possible in the other direction: to associate with *every* $\mathbf{A} \in \mathcal{MA}$ some $\mathcal{F}(\mathbf{A}) \in \mathcal{MA}^-$ which is natural in a sense which will shortly be made technically precise, and which has the following property: homomorphisms from $\mathbf{A}$ to $\mathcal{G}(\mathbf{B})$ are in one-to-one correspondence with those from $\mathcal{GF}(\mathbf{A})$ to $\mathcal{G}(\mathbf{B})$, and there is a homomorphism $\mathbf{A} \xrightarrow{\eta_A} \mathcal{GF}(\mathbf{A})$ which in a sense pairs them off: for any $\mathbf{A} \xrightarrow{f} \mathcal{G}(\mathbf{B})$ there is a unique homomorphism $\mathcal{GF}(\mathbf{A}) \xrightarrow{f'} \mathcal{G}(\mathbf{B})$ with $f = f' \circ \eta_A$; also, f' determines f in a similar fashion. Thus if $\mathbf{A}$ is complemented, we should have $\mathcal{GF}(\mathbf{A}) \cong \mathbf{A}$ and η_A the identity isomorphism. In the general case η_A will turn out to be injective, and so the above property will provide a limited form of the congruence extension property.

In the following, for reasons of clarity technical definitions will be kept to the minimum required to demonstrate what we want, but the liberty will occasionally be taken of referring to theorems and concepts which have not been fully explained. MacLane[1] is a useful reference if these are wanted: only the central notions will be repeated here. Thus an *adjunction* between categories is defined in the following way.

---

[1]MacLane 1971.

Definition 106 *If $\mathcal{X}$ and $\mathcal{A}$ are categories, then an adjunction from $\mathcal{X}$ to $\mathcal{A}$ is a triple*

$$< \mathcal{F}, \mathcal{G}, \varphi >: \mathcal{X} \rightharpoonup \mathcal{A}$$

where $\mathcal{F} : \mathcal{X} \to \mathcal{A}$ and $\mathcal{G} : \mathcal{A} \to \mathcal{X}$ are functors and φ is a function assigning to each pair $x \in \mathcal{X}$ and $a \in \mathcal{A}$ a bijection $\varphi_{x,a}$ from morphisms $\{f \mid \mathcal{F}(x) \xrightarrow{f} a\}$ in $\mathcal{A}$ to morphisms $\{g \mid x \xrightarrow{f} \mathcal{G}(a)\}$ in $\mathcal{X}$, such that φ has the following property: take any $x' \xrightarrow{g} x$ in $\mathcal{X}$ and $a \xrightarrow{h} a'$ in $\mathcal{A}$, thus giving the diagrams

$$\mathcal{F}(x') \xrightarrow{\mathcal{F}g} \mathcal{F}(x) \xrightarrow{f} a \xrightarrow{h} a'$$

$$x' \xrightarrow{g} x \xrightarrow{\varphi f} \mathcal{G}(a) \xrightarrow{\mathcal{G}h} \mathcal{G}(a').$$

Then we have

$$\varphi(f \circ \mathcal{F}g) = \varphi f \circ g \quad \text{and} \quad \varphi(h \circ f) = \mathcal{G}h \circ \varphi f.$$

This holds just in case the inverse φ^{-1} of φ has the same property: that is, for any such g, h and for $x \xrightarrow{k} \mathcal{G}(a)$

$$\varphi^{-1}(k \circ g) = \varphi^{-1}k \circ \mathcal{F}g \quad \text{and} \quad \varphi^{-1}(\mathcal{G}h \circ k) = h \circ \varphi^{-1}k$$

hold. □

$\mathcal{F}$ is called a *left adjoint* for $\mathcal{G}$, and $\mathcal{G}$ is a a *right adjoint* for $\mathcal{F}$. These are unique if they exist. The aim now is to prove the existence of a left adjoint to the forgetful functor $\mathcal{G} : \mathcal{MA}^- \to \mathcal{MA}$ by appealing to a theorem called the Freyd Adjoint Functor Theorem, the premises of which will now be shown to be satisfied in our case.

To show the first of these conditions, recall that products in algebra are *universal* in the sense that if for algebras $\mathbf{A}'$ and $\{\mathbf{A}_i \mid i \in I\}$ there are homomorphisms $\mathbf{A}' \xrightarrow{f_i} \mathbf{A}_i$, then there is a unique homomorphism $\mathbf{A}' \xrightarrow{f} \prod_{i \in I} \mathbf{A}_i$ such that for each $i \in I$ $f_i = \pi_i \circ f$, where π_i is the projection mapping from $\prod_{i \in I} \mathbf{A}_i$ to $\mathbf{A}_i$: f is defined by $f(a)(i) = f_i(a)$. This shows that the categories $\mathcal{MA}$ and

$\mathcal{MA}^-$ have (small) products in a category-theoretical sense, this being an instance of the more general phenomenon of a *limit*; and a category possessing all (reasonable) limits is called ***small-complete***. Another limit is that of an equaliser, which in this case is as follows: given a pair of homomorphisms $f, g : \mathbf{A}_1 \to \mathbf{A}_2$, it consists of an algebra $\mathbf{A}'$ and a homomorphism $\mathbf{A}' \xrightarrow{e} \mathbf{A}_1$ with the property that $f \circ e = g \circ e$, and such that if $\mathbf{B} \xrightarrow{h} \mathbf{A}_1$ and $f \circ h = g \circ h$ then there is a unique homomorphism $\mathbf{B} \xrightarrow{h'} \mathbf{A}'$ with $e \circ h' = h$. These limits also always exist in both $\mathcal{MA}$ and $\mathcal{MA}^-$, for $\mathbf{A}'$ can be defined to be the subalgebra of $\mathbf{A}_1$ with universe $\{a \in A \mid f(a) = g(a)\}$. This set is clearly closed under the operations of $\mathbf{A}_1$; and if h is as above, then for $b \in \mathbf{B}$ define $h'(b) = h(b)$. $h(b) \in B$, because if not then by definition of $\mathbf{B}$ we have $f \circ h(b) \neq g \circ h(b)$, contrary to the assumption that $f \circ h = g \circ h$. So the image of h is contained in B and h' is a homomorphism. To show a category small-complete it suffices to show that it has products and equalisers; having identified and shown the existence of these, we have shown that $\mathcal{MA}$ and $\mathcal{MA}^-$ are small complete - they have all limits.

The second thing to show is that $\mathcal{G}$ ***preserves*** these limits, which for our purposes we may take to mean that in the case of products, if $\{\mathbf{B}_i \mid i \in I\}$ are in $\mathcal{MA}^-$ then

$$\mathcal{G}(\prod_{i \in I} \mathbf{B}_i) \cong \prod_{i \in I} \mathcal{G}(\mathbf{B}_i);$$

and for equalisers, if $\mathbf{B} \xrightarrow{e} \mathbf{B}_1$ is the equaliser of the pair $f, g : \mathbf{B}_1 \to \mathbf{B}_2$, then $\mathcal{G}e$ is the equaliser of the pair $\mathcal{G}f$, $\mathcal{G}g$. The first case is easy to see, for if $\{\mathbf{A}_i \mid i \in I\}$ is a set of complemented modal algebras and $a \in \prod_{i \in I} \mathbf{A}_i$, then the complement of a is the element a' such that for each $i \in I$ $a'(i)$ is the complement of $a(i)$. So complement exists in the product and may be defined pointwise, showing that the product of the algebras $\{\mathbf{B}_i \mid i \in I\}$ with $\mathcal{G}(\mathbf{B}_i) = \mathbf{A}_i$ - which are unique up to isomorphism - is isomorphic to the result of adding a complementation operation to $\prod_{i \in I} \mathbf{A}_i$. As for preservation of equalisers, observe that these were defined set-theoretically in the same way in both categories $\mathcal{MA}$ and $\mathcal{MA}^-$: then the fact that for $\mathbf{B} \in \mathcal{MA}^-$, $\mathbf{B}$ and $\mathcal{G}(\mathbf{B})$ have the same universe proves that $\mathcal{G}$

preserves equalisers. From this it follows that $\mathcal{G}$ preserves all small limits, or is *continuous*.

To demonstrate that the third and final condition holds, a further definition is required. For any $\mathbf{A} \in \mathcal{MA}$ the category $(\mathbf{A} \downarrow \mathcal{G})$ is defined to have as objects pairs $< f, \mathbf{B} >$ with $\mathbf{B} \in \mathcal{MA}^-$ and $\mathbf{A} \xrightarrow{f} \mathcal{G}(\mathbf{B})$ in $\mathcal{MA}$; and if $\mathbf{A} \xrightarrow{f} \mathcal{G}(\mathbf{B})$ and $\mathbf{A} \xrightarrow{f'} \mathcal{G}(\mathbf{B}')$ then a morphism $< f, \mathbf{B} > \xrightarrow{h} < f', \mathbf{B}' >$ in $(\mathbf{A} \downarrow \mathcal{G})$ is a homomorphism $\mathbf{B} \xrightarrow{h} \mathbf{B}'$ in $\mathcal{MA}^-$ such that $f' = \mathcal{G}h \circ f$. Now pick out certain objects of $(\mathbf{A} \downarrow \mathcal{G})$ by defining

Definition 107 $\mathbf{A} \xrightarrow{f} \mathcal{G}(\mathbf{B})$ *is said to* span $\mathbf{B}$ *if there is no proper - that is, non-isomorphic - injective* $\mathbf{B}' \xrightarrow{h} \mathbf{B}$ *in* $\mathcal{MA}^-$ *such that* f *factors through* $\mathcal{G}(\mathbf{B}') \xrightarrow{\mathcal{G}h} \mathcal{G}(\mathbf{B})$. *Abbreviating,* f *simply* spans *if it spans* $\mathbf{B}$. □

In our case, the definition implies that if f spans $\mathbf{B}$, then the image of the homomorphism f, regarded as a subset of $\mathbf{B}$ rather than of $\mathcal{G}(\mathbf{B})$, generates $\mathbf{B}$. For if not then the least subalgebra $\mathbf{B}'$ of $\mathbf{B}$ containing the image of f would be a proper subalgebra of $\mathbf{B}$, contradicting the definition. This also shows why for any $< f, \mathbf{B} > \in (\mathbf{A} \downarrow \mathcal{G})$, there is some $< f', \mathbf{B}' > \in (\mathbf{A} \downarrow \mathcal{G})$ such that f' spans and f factors through f': the least subalgebra of $\mathbf{B}$ containing the image of f provides the required $\mathbf{B}'$, while for $a \in A$, f' is defined as $f'(a) = f(a)$.

The set[2] of objects $< f, \mathbf{B} > \in (\mathbf{A} \downarrow \mathcal{G})$ such that f spans satisfy what is called the *solution set condition* for the category $(\mathbf{A} \downarrow \mathcal{G})$, which is the last of the conditions to be shown given which Freyd's Adjoint Functor Theorem shows the existence of a left adjoint to $\mathcal{G}$. In the abstract, what is happening

[2]For reasons not gone into here, this assumption is legitimate. Essentially, choose one from each isomorphism class; then the number of such objects is limited by the fact that the size of the universe of these algebras $\mathbf{B}$ is manageably constrained, being generated by the universe of $\mathbf{A}$.

here is that, given that $\mathcal{MA}^-$ has all limits and these are preserved by $\mathcal{G}$, these limits induce limits in $(\mathbf{A} \downarrow \mathcal{G})$; then the presence of the solution set is used to show that $(\mathbf{A} \downarrow \mathcal{G})$ has an initial object - a pair $< f, \mathbf{B} >$ through which *any* $\mathbf{A} \xrightarrow{f'} \mathcal{G}(\mathbf{B}')$ may be factored. The left adjoint then maps $\mathbf{A}$ to this algebra $\mathbf{B}$. More specifically, these limits are induced in the following way. Suppose that $\{< f_i, \mathbf{B}_i > | \ i \in I\}$ is such that $\{f_i \mid i \in I\}$ are the homomorphisms from $\mathbf{A}$ which span. Then $\prod_{i \in I} < f_i, \mathbf{B}_i >$ is $< f, \prod_{i \in I} \mathbf{B}_i >$, where for $a \in \mathbf{A}$ $f(a)(i) = f_i(a)$. Because

$$\mathcal{G}(\prod_{i \in I} \mathbf{B}_i) \cong \prod_{i \in I} \mathcal{G}(\mathbf{B}_i),$$

this is well-defined. Let $\mathbf{B}$ be the least subalgebra of $\prod_{i \in I} \mathbf{B}_i$ containing $\{f(a) \mid a \in A\}$, where these are regarded as being in the universe of $\prod_{i \in I} \mathbf{B}_i$ rather than that of $\mathcal{G}(\prod_{i \in I} \mathbf{B}_i)$. Then defining $\mathbf{A} \xrightarrow{g} \mathcal{G}(\mathbf{B})$ by $g(a) = f(a)$, $< g, \mathbf{B} >$ can be seen to be initial. For suppose $< g', \mathbf{B}' > \in (\mathbf{A} \downarrow \mathcal{G})$ and for $< h, \mathbf{B}_j > \in (\mathbf{A} \downarrow \mathcal{G})$, $\mathbf{B}_j \xrightarrow{k} \mathbf{B}'$ is the embedding in $\mathbf{B}'$ of its least subalgebra containing $\{g'(a) \mid a \in A\}$. Then h spans, and it is not too difficult to see that, where i is the obvious injective morphism, the composite of the arrows

$$< g, \mathbf{B} > \xrightarrow{i} < f, \prod_{i \in I} \mathbf{B}_i > \xrightarrow{\pi_j} < h, \mathbf{B}_j > \xrightarrow{k} < g', \mathbf{B}' >,$$

is indeed a morphism in $(\mathbf{A} \downarrow \mathcal{G})$, showing that $< g, \mathbf{B} >$ is initial. The functor $\mathcal{F}$ then takes $\mathbf{A}$ to this algebra $\mathbf{B} \in \mathcal{MA}^-$, and we have an adjunction $< \mathcal{F}, \mathcal{G}, \varphi >$: $\mathcal{MA} \rightharpoonup \mathcal{MA}^-$: abbreviate $\mathcal{GF}(\mathbf{A})$ to $\eta\mathbf{A}$ and let η_A be the homomorphism from $\mathbf{A}$ to $\eta\mathbf{A}$ given by the above construction. Then if $\mathcal{F}(\mathbf{A}) \xrightarrow{f} \mathbf{B}$, let φf be defined as $\mathcal{G}f \circ \eta_A$. This completes the proof of the existence of a left adjoint.

We already know some things about $\eta\mathbf{A}$. By the construction of $\eta\mathbf{A}$, if complementation is regarded as an operation then A generates $\eta\mathbf{A}$. Also, the homomorphism η_A from $\mathbf{A}$ to $\eta\mathbf{A}$, is injective: consider the injective $\mathbf{A} \xrightarrow{f} (\mathbf{A}_+)^+$, where $\mathbf{A}_+$ is as usual the discrete frame. We know that $(\mathbf{A}_+)^+$ is complemented, so for some $\mathbf{B} \in \mathcal{MA}^-$ we have $\mathcal{G}(\mathbf{B}) \cong (\mathbf{A}_+)^+$. Because $< \eta_A, \mathcal{F}(\mathbf{A}) >$ is initial, the injective mapping f must factor through η_A: for some $g : \mathcal{F}(\mathbf{A}) \to \mathbf{B}$ we have

$f = \mathcal{G}g \circ \eta_A$, and so η_A must be injective. Also, it can be seen that for $\mathbf{B} \in \mathcal{MA}^-$, $\mathbf{B} \cong \mathcal{F}\mathcal{G}(\mathbf{B})$: $\mathcal{G}(\mathbf{B})$ is closed under complementation, so the description of $\mathcal{F}\mathcal{G}(\mathbf{B})$ as a subalgebra generated by the elements of $\mathcal{G}(\mathbf{B})$ including complementation as an *operation* obviously does not expand the universe of $\mathcal{G}(\mathbf{B})$.

More can also be said about the relation between $\mathcal{MA}$ and $\mathcal{MA}^-$. For homomorphisms f, g: $\mathbf{B_1} \to \mathbf{B_2}$ in $\mathcal{MA}^-$, the definition of the functor $\mathcal{G}$ shows that $f \neq g$ implies that $\mathcal{G}f \neq \mathcal{G}g$: that is to say, $\mathcal{G}$ is *faithful.* But $\mathcal{G}$ also has the property that if $\mathcal{G}(\mathbf{B}_1) \xrightarrow{h} \mathcal{G}(\mathbf{B}_2)$ is a homomorphism in $\mathcal{MA}$, then there is some $\mathbf{B}_1 \xrightarrow{f} \mathbf{B}_2$ with $\mathcal{G}f = h$: in other words, $\mathcal{G}$ is *full.* This is true because any homomorphism between modal algebras respects complements, where these exist: if the complement of a is a', then $h(a \wedge a') = h(a) \wedge h(a') = 0$ and $h(a \vee a') = h(a) \vee h(a') = 1$, so $h(a')$ is the complement of $h(a)$. So with f defined as the same function from B_1 to B_2, we have $\mathcal{G}f = h$. Because $\mathcal{G}$ is full and faithful, this means that for $\mathbf{A} \in \mathcal{MA}$ and $\mathbf{B} \in \mathcal{MA}^-$, $\mathcal{G}$ is a bijection from $\{f \mid \mathcal{F}(\mathbf{A}) \xrightarrow{f} \mathbf{B}\}$ to $\{g \mid \eta\mathbf{A} \xrightarrow{g} \mathcal{G}(\mathbf{B})\}$. But because of the adjunction there is also a bijection $\varphi_{\mathbf{A},\mathbf{B}}$ from $\{f \mid \mathcal{F}(\mathbf{A}) \xrightarrow{f} \mathbf{B}\}$ to $\{g \mid \mathbf{A} \xrightarrow{g} \mathcal{G}(\mathbf{B})\}$. Putting these two facts together, any homomorphism $\mathbf{A} \xrightarrow{h} \mathcal{G}(\mathbf{B})$ has a unique extension $\eta\mathbf{A} \xrightarrow{h'} \mathcal{G}(\mathbf{B})$: h' is the unique homomorphism such that $h' \circ \eta_A = h$.[3] It is not too difficult to see that this extension h' is in fact $\mathcal{G}\varphi^{-1}h$. Technically, we have in fact shown that because $\mathcal{G}$ is full and faithful, $\mathcal{MA}^-$ is equivalent to a subcategory of $\mathcal{MA}$ - the image of $\mathcal{MA}^-$ under $\mathcal{G}$. These of course are simply the complemented modal algebras, and the parallel adjunction is given be the composite functor $\mathcal{G}\mathcal{F}$ together with the inclusion of complemented modal algebras in modal algebras in general. Moreover, this is what is called a *reflective* subcategory: with $\mathcal{F}$ the reflector, the adjunction gives a reflection of $\mathcal{MA}$ in its complemented subcategory. To sum up so far:

[3]And so η_A is *epi*, although in general it is not surjective.

Theorem 108 *$\mathcal{MA}$ is adjoint to $\mathcal{MA}^-$ by $< \mathcal{F}, \mathcal{G}, \varphi >: \mathcal{MA} \rightharpoonup \mathcal{MA}^-$. $\mathcal{G}$ is full and faithful, making $\mathcal{MA}^-$ equivalent to a reflective subcategory of $\mathcal{MA}$, the* complemented *modal algebras. Also, any $\mathbf{A} \in \mathcal{MA}$ is a subalgebra of $\mathcal{GF}(\mathbf{A}) = \eta\mathbf{A}$, which it generates if complementation is regarded as an operation.* □

Incidentally, it can also be shown that $\mathcal{G}$ has a *right* adjoint, although this is not as interesting here as its left adjoint. It will, however, be encountered again at the end of the chapter. For $\mathbf{A} \in \mathcal{MA}$ define the algebra $\mathcal{H}(\mathbf{A})$ in $\mathcal{MA}^-$ to be the result of adding a complementation operation to the greatest complemented subalgebra of $\mathbf{A}$. This subalgebra exists, and its universe is the set of *all* complemented elements in the universe of $\mathbf{A}$, because

Theorem 109 *If $\mathbf{A} \in \mathcal{MA}$ and $\{a_1, \dots a_n\} \subseteq A$ are complemented then so is $\omega(a_1, \dots a_n)$, where ω is an n-ary operator derived from the operators of $\mathbf{A}$.*

Proof. It suffices to check each of the primitive operators in turn. First we have $-\neg a_i = \neg - a_i$: $a_i \wedge -a_i = 0$, so $\neg(a_i \wedge -a_i) = \neg a_i \vee \neg - a_i = 1$; and $a_i \vee -a_i = 1$, so $\neg(a_i \vee -a_i) = \neg a_i \wedge \neg - a_i = 0$. Also, $-(a_i \wedge a_j) = -a_i \vee -a_j$: that $a_i \wedge a_j \wedge (-a_i \vee -a_j) = 0$ can be seen by using the distribution of $\wedge$ over $\vee$, and the other distribution rule shows that their join is the element 1. The fact that $-(a_i \vee a_j) = -a_i \wedge -a_j$ is shown similarly to this, and it is obvious that νa_i and μa_i are complemented. □

So all the complemented elements of $\mathbf{A}$ form a subalgebra of $\mathbf{A}$: obviously this is the greatest complemented subalgebra of $\mathbf{A}$. For a homomorphism $\mathbf{A}_1 \xrightarrow{f} \mathbf{A}_2$ in $\mathcal{MA}$ define $\mathcal{H}f$ simply to be the restriction of the underlying function of f to the complemented elements of its domain - the universe of $\mathcal{H}(\mathbf{A}_1)$. The image of $\mathcal{H}f$ is contained in the universe of $\mathcal{H}(\mathbf{A}_2)$, because this image is complemented: for $a_i \in A_1$ complemented, then the complement of $f(a_i)$ is $f(-a_i)$. Given this definition of $\mathcal{H}f$ as the restriction of f to the complemented elements, it is almost immediate that $\mathcal{H}f$ is a homomorphism, $\mathcal{H}1_A = 1_{\mathcal{H}A}$ and $\mathcal{H}(g \circ f) = \mathcal{H}g \circ \mathcal{H}f$, and so $\mathcal{H}$ is a functor from $\mathcal{MA}$ to $\mathcal{MA}^-$.

Theorem 110 *There is an adjunction* $< \mathcal{G}, \mathcal{H}, \varphi' >: \mathcal{MA}^- \rightarrow \mathcal{MA}$.

Proof. We show that for any algebras $\mathbf{A} \in \mathcal{MA}$ and $\mathbf{B} \in \mathcal{MA}^-$, there is a bijection from $\{f \mid \mathcal{G}(\mathbf{B}) \xrightarrow{f} \mathbf{A}\}$ to $\{g \mid \mathbf{B} \xrightarrow{g} \mathcal{H}(\mathbf{A})\}$. Now obviously $\mathbf{B} \cong \mathcal{H}\mathcal{G}(\mathbf{B})$ - $\mathcal{G}(\mathbf{B})$ is complemented, so $\mathbf{B}$, $\mathcal{G}(\mathbf{B})$ and $\mathcal{H}\mathcal{G}(\mathbf{B})$ all have the same universe; the isomorphism then follows from the way in which the functors were defined. Letting η'_B be this isomorphism, for a homomorphism $\mathcal{G}(\mathbf{B}) \xrightarrow{f} \mathbf{A}$ φ' is defined to be $\mathcal{H}f \circ \eta'_B$: Clearly this a homomorphism from $\mathbf{B}$ to $\mathcal{H}(\mathbf{A})$. And if $f' \neq f$ they must differ only in the complemented subuniverse of $\mathbf{A}$, so $\varphi' f' \neq \varphi' f$. It is surjective as a mapping because for any $\mathbf{B} \xrightarrow{g} \mathcal{H}(\mathbf{A})$, the same function on the universe of $\mathbf{B}$ to the perhaps expanded universe of $\mathbf{A}$ is readily seen to be a homomorphism from $\mathcal{G}(\mathbf{B})$ to $\mathbf{A}$ which φ' maps back to g. Spelling out the final conditions for an adjunction given our definition of φ', it can be seen that they are satisfied once it is shown that for $\mathbf{B}_1 \xrightarrow{g} \mathbf{B}_2$ in $\mathcal{MA}^-$, $\mathcal{H}\mathcal{G}g \circ \eta'_{B_1} = \eta'_{B_2} \circ g$. But this is easy to see, especially since η'_{B_1} and η'_{B_2} are isomorphisms. □.

Abbreviate $\mathcal{G}\mathcal{H}(\mathbf{A})$ to $\varepsilon\mathbf{A}$. This time for any $\mathcal{G}(\mathbf{B}) \xrightarrow{f} \mathbf{A}$ we are given what might be called a unique *restriction* $\mathcal{G}\varphi' f$ of f to the subalgebra $\varepsilon\mathbf{A}$ of $\mathbf{A}$; and conversely any homomorphism $\mathcal{G}(\mathbf{B}) \xrightarrow{g} \varepsilon\mathbf{A}$ has a unique extension to $\mathbf{A}$. The awkwardness of the locution illustrates the fact that this was already obvious, however, with the same function defining the other homomorphism in each case. Now because the complemented modal algebras form a subcategory of $\mathcal{MA}$, equivalent to $\mathcal{MA}^-$, the inclusion and ε also form an adjoint pair: technically, this means that in addition to being a reflective subcategory,

Theorem 111 *The complemented modal algebras in* $\mathcal{MA}$ *form a co-reflective subcategory of* $\mathcal{MA}$. □

The functor ε - or in a parallel fashion $\mathcal{H}$ - may be called a *co-reflector.*

7.2 Their Congruences

Before examining whether these results can illuminate congruences in any way, we need the next theorem, which explains why $\eta\mathbf{A}$ is called the complemented algebra *freely* generated by $\mathbf{A}$.

Theorem 112 *Suppose that for $\mathbf{A} \in \mathcal{MA}$ and $\mathbf{B} \in \mathcal{MA}^-$ there is an injective $\mathbf{A} \xrightarrow{f} \mathcal{G}(\mathbf{B})$ which spans. Then there is a surjective homomorphism f' from $\eta\mathbf{A}$ onto $\mathcal{G}(\mathbf{B})$.*

Proof. Let f' be the unique homomorphism such that $f' \circ \eta_A = f$: that is, $f' = \mathcal{G}\varphi^{-1}f$. We show that f' is surjective. So let b be any element in the universe of $\mathcal{G}(\mathbf{B})$, and also regard it as an element of $\mathbf{B}$. Because the image of f generates $\mathbf{B}$, there is some n-ary operator ω, and some $a_1, \dots a_n \in A$ such that $b = \omega(f(a_1), \dots f(a_n))$, where ω is derived from the operators of $\mathbf{B}$. Thus ω may involve the operation of complementation. Now $\eta\mathbf{A}$ is also complemented and generated in the above sense by $\mathbf{A}$; so we may consider the elements $\eta_A(a_1), \dots \eta_A(a_n)$ in $\eta\mathbf{A}$ as elements of $\mathcal{F}(\mathbf{A})$, find the element $c = \omega(\eta_A(a_1), \dots \eta_A(a_n))$ there, and then regard c as being in $\eta\mathbf{A}$. Now because $\mathcal{G}$ is full and faithful, f' preserves the operation ω:

$$f'(c) = f'(\omega(\eta_A(a_1), \dots \eta_A(a_n))) = \omega(f' \circ \eta_A(a_1), \dots f' \circ \eta_A(a_n)).$$

But $f' \circ \eta_A = f$, so

$$f'(c) = \omega(f(a_1), \dots f(a_n)) = b,$$

showing that f' is surjective. □

If $\mathbf{A} \xrightarrow{f} \mathcal{G}(\mathbf{B})$ is surjective then obviously f', the unique extension of f to $\eta\mathbf{A}$ is also surjective: this follows from the facts that $f' \circ \eta_A = f$ and η_A is injective. Thus the congruences on $\mathbf{A}$ with a complemented quotient have unique extensions to congruences on $\eta\mathbf{A}$. It is then natural to ask if this extends to arbitrary congruences on a modal algebra $\mathbf{A}$, especially since we have already characterised the congruences of $\eta\mathbf{A}$ because this algebra is complemented. For then the congruences of the arbitrary algebra $\mathbf{A}$ would have a useful characterisation. The answer to the question is no, but the approach is not wholly uninformative:

Theorem 113 *If θ is an arbitrary congruence on a modal algebra $\mathbf{A} \in \mathcal{MA}$, then there is a least congruence on $\eta\mathbf{A}$ the restriction of which to $\mathbf{A}$ is θ.* □

This will follow from the next two theorems. First it is shown that there does exist such a congruence, for we know this much more more about extending surjective homomorphisms:

Theorem 114 *If $\mathbf{A}_1 \xrightarrow{f} \mathbf{A}_2$ is surjective then so is $\mathcal{G}\varphi^{-1}(\eta_{A_2} \circ f)$.*

Proof. We assume that surjective f defines the congruence θ, in the sense that for $a, b \in A_1$, $a\theta b$ iff $f(a) = f(b)$. Then because η_{A_2} is injective, θ is also defined by $\eta_{A_2} \circ f$. Let f' abbreviate $\mathcal{G}\varphi^{-1}(\eta_{A_2} \circ f)$, the unique extension of $\eta_{A_2} \circ f$ to $\eta\mathbf{A}_1$. Obviously we have $f' \circ \eta_{A_1} = \eta_{A_2} \circ f$, so $f' \circ \eta_{A_1}$ also defines θ. Now factor f' as $i \circ s$, where $\eta\mathbf{A}_1 \xrightarrow{s} \mathbf{A}'$ is the surjective mapping of $\eta\mathbf{A}_1$ onto its image under f' given by $s(a) = f'(a)$ for $a \in A_1$; $\mathbf{A}' \xrightarrow{i} \eta\mathbf{A}_2$ is then given by the subalgebra connection between the two algebras. Any $\mathcal{MA}$-homomorphism can be factored in this way into a surjective one followed by an injective one. Note that, as has already been seen, because $\eta\mathbf{A}_1$ is a complemented modal algebra, so is its homomorphic image $\mathbf{A}'$. Since we have $f' = i \circ s$, also $f' \circ \eta_{A_1} = i \circ s \circ \eta_{A_1}$, so $i \circ s \circ \eta_{A_1}$ also defines θ; and because i is injective so does $s \circ \eta_{A_1} : \mathbf{A}_1 \to \mathbf{A}'$. But this means that the image of $\mathbf{A}_1$ under $s \circ \eta_{A_1}$ is isomorphic to $\mathbf{A}_2$ because $\mathbf{A}_2$ is isomorphic to the quotient $\mathbf{A}_1/\theta$. So let $\mathbf{A}_2 \xrightarrow{k} \mathbf{A}'$ be the embedding; thus, $k \circ f$ is a surjective-injective factorisation of $s \circ \eta_{A_1}$ and $s \circ \eta_{A_1} = k \circ f$. Now we show that $\eta_{A_2} = i \circ k$. Because $k \circ f = s \circ \eta_{A_1}$ we also have $i \circ k \circ f = i \circ s \circ \eta_{A_1}$;

but $i \circ s = f'$, so $i \circ k \circ f = f' \circ \eta_{A_1}$; and $f' \circ \eta_{A_1} = \eta_{A_2} \circ f$, so $i \circ k \circ f = \eta_{A_2} \circ f$. But the fact that f is surjective then implies that $i \circ k = \eta_{A_2}$. And because η_{A_2} spans, i cannot be a proper injective morphism: $\mathbf{A}'$ and $\eta\mathbf{A}_2$ are complemented algebras and $\mathcal{G}$ is full, so for some $\mathcal{MA}^-$-homomorphism j we have $\mathcal{G}j = i$. And if $\mathcal{G}j$ were properly injective then j would also be properly injective, since it is the same function. So i must be an isomorphism. Finally, because $f' = i \circ s$, i is an isomorphism and s is surjective, this means that f' is surjective, completing the proof. □

Corollary 115 *For any congruence θ on $\mathbf{A} \in \mathcal{MA}$ there is a congruence θ' on $\eta\mathbf{A}$, the restriction of which to $\mathbf{A}$ is θ.* □

Theorem 116 *For any congruence θ on $\mathbf{A} \in \mathcal{MA}$ there is a smallest congruence on $\eta\mathbf{A}$ which extends θ.*

Proof. This is obvious in the abstract: at least one congruence extending θ exists, so simply take the intersection of all those congruences extending θ. But a more direct method of proof will tell us more about this congruence, so let θ, f and f' be as in the previous proof. Because the quotient of any congruence on $\eta\mathbf{A}_1$ is complemented, we may assume that any such congruence θ' the restriction of which to $\mathbf{A}$ is θ is given by a homomorphism onto $\mathcal{G}(\mathbf{B})$, for some $\mathbf{B} \in \mathcal{MA}^-$. So let $\eta\mathbf{A}_1 \xrightarrow{g} \mathcal{G}(\mathbf{B})$ be such a surjective homomorphism defining θ', the restriction of which to $\mathbf{A}_1$ defines θ there in this way. To prove the theorem it suffices to show that there is a surjective homomorphism from $\eta\mathbf{A}_2$ onto $\mathcal{G}(\mathbf{B})$. For then the congruence defined by f' will be contained in θ', and so will be the least such congruence, given that θ' was arbitrary. Now because the restriction of θ' to $\mathbf{A}_1$ is θ, as was seen in the previous proof this means that $\mathbf{A}_2$ is isomorphic to a subalgebra of $\mathcal{G}(\mathbf{B})$: the morphism $g \circ \eta_{A_1}$ can be factored through f and an injective $\mathbf{A}_2 \xrightarrow{h} \mathcal{G}(\mathbf{B})$ with $g \circ \eta_{A_1} = h \circ f$. Let h' abbreviate $\mathcal{G}\varphi^{-1}h$: so $h' \circ \eta_{A_2} = h$ and h' is a homomorphism from $\eta\mathbf{A}_2$

to $\mathcal{G}(\mathbf{B})$. We now show that h is surjective, as required by the proof. Take any element $b \in \mathcal{G}(B)$: then for some $a \in \eta A_1$ we have $g(a) = b$. But because the set $\{\eta_{A_1}(a_i) \mid a_i \in A_1\}$ generates $\eta\mathbf{A}_1$ as a $\mathcal{MA}^-$-algebra, we may assume that there is some n-ary operator ω derived from the primitive operators of the $\mathcal{MA}^-$-algebras, and some $\{a_1, \ldots a_n\} \subseteq A_1$ such that

$$g(\omega(\eta_{A_1}(a_1), \ldots \eta_{A_1}(a_n))) = b.$$

But it has been seen that g preserves ω, so

$$b = \omega(g \circ \eta_{A_1}(a_1), \ldots g \circ \eta_{A_1}(a_n)).$$

Now $g \circ \eta_{A_1} = h \circ f$, and because $h = h' \circ \eta_{A_2}$ also $g \circ \eta_{A_1} = h' \circ \eta_{A_2} \circ f$. But $\eta_{A_2} \circ f = f' \circ \eta_{A_1}$ so $g \circ \eta_{A_1} = h' \circ f' \circ \eta_{A_1}$, and

$$b = \omega(h' \circ f' \circ \eta_{A_1}(a_1), \ldots h' \circ f' \circ \eta_{A_1}(a_n)).$$

But $\eta\mathbf{A}_2$ is complemented, so $\omega(f' \circ \eta_{A_1}(a_1), \ldots, f' \circ \eta_{A_1}(a_n))$ exists in its universe and h' preserves ω; thus

$$h'(\omega(f' \circ \eta_{A_1}(a_1), \ldots, f' \circ \eta_{A_1}(a_n))) = b,$$

and so h' is surjective. □

If θ is a congruence on $\mathbf{A}$, denote by $\mathcal{E}(\theta)$ its least extension to $\eta\mathbf{A}$; and for θ a congruence on $\eta\mathbf{A}$ let $\mathcal{R}(\theta)$ be its restriction to $\mathbf{A}$. Then by the proof of the last theorem we have

Corollary 117 $\eta(\mathbf{A}/\theta) \cong \eta\mathbf{A}/\mathcal{E}(\theta)$. □

It can now be shown that a *precise* characterisation of the congruences of $\mathbf{A}$ cannot be achieved by exploiting in this manner our previous characterisation of the congruences of $\eta\mathbf{A}$.

Theorem 118 *It is possible that for distinct congruences θ and θ' on $\eta\mathbf{A}$, nevertheless $\mathcal{R}(\theta) = \mathcal{R}(\theta')$.*

Proof. This can be proved by producing a simple counterexample. Let $\mathbf{A}$ be the four element modal algebra with $0 < a < \neg a < 1$, $\nu a = 0$ and $\nu\neg a = 1$. Also, where $\mathbf{4}$ and $\mathbf{2}$ are as previously described, consider their product $\mathbf{4} \times \mathbf{2}$ together with the mapping from $\mathbf{A}$ to $\mathbf{4} \times \mathbf{2}$ given by

$$0 \longmapsto < 0,0 >; \quad a \longmapsto < a,0 >; \quad \neg a \longmapsto < a,1 >; \quad 1 \longmapsto < 1,1 > .$$

We now define two modal algebras $\mathbf{B}_1$ and $\mathbf{B}_2$ on $\mathbf{4} \times \mathbf{2}$ by specifying the modalities.

$$\text{In } \mathbf{B}_1, \qquad \nu x = \begin{cases} < 1,1 > & \text{if } x \in \{< a,1 >, < 1,1 >\}; \\ < 0,0 > & \text{otherwise.} \end{cases}$$

$$\text{In } \mathbf{B}_2, \qquad \nu x = \begin{cases} < 1,1 > & \text{if } x \in \{< a,1 >, < 1,1 >\}; \\ < 0,1 > & \text{if } x \in \{< 0,1 >, < b,1 >\}; \\ < 0,0 > & \text{otherwise.} \end{cases}$$

Let g_1 and g_2 be the mappings into $\mathbf{B}_1$ and $\mathbf{B}_2$ respectively given above: then it is easy to see $\mathbf{B}_1$ and $\mathbf{B}_2$ are both modal algebras and that g_1 and g_2 are both modal homomorphisms. The algebras $\mathbf{B}_1$ and $\mathbf{B}_2$ are complemented since $\mathbf{4} \times \mathbf{2}$ is, and with complementation as an operation, they are generated by the subalgebra $\mathbf{A}$: we have $< b,0 >= -g_i(\neg a)$, $< b,1 >= -g_i(a)$, $< 1,0 >= g_i(a) \vee -g_i(\neg a)$ and $< 0,1 >= g_i(\neg a) \wedge -g_i(a)$. In fact, both g_1 and g_2 span. Neither $\mathbf{B}_1$ nor $\mathbf{B}_2$ is a homomorphic image of the other - consider what to do with the element $< 0,1 >\in \mathbf{B}_2$ in attempting to construct a surjective homomorphism in either direction - which means that $\eta\mathbf{A}$ is not isomorphic to either of these algebras: if it were isomorphic to $\mathbf{B}_1$ say, this implies that there is a surjective homomorphism f from $\eta\mathbf{A} \cong \mathbf{B}_1$ to $\mathbf{B}_2$ with $f \circ \eta_A = g_2$, which is false. So now we have surjective homomorphisms f_i from $\eta\mathbf{A}$ to $\mathbf{B}_i$ with $f_i \circ \eta_A = g_i$, and these cannot be injective since g_1 and g_2 span. This means that the congruences they determine on $\eta\mathbf{A}$ are not equal to the identity congruence and are distinct; but $f_i \circ \eta_A = g_i$ and g_i is injective, so the restriction of both g_1 and g_2 to $\mathbf{A}$ is the identity congruence Δ on $\mathbf{A}$. □

Not surprisingly, $\varepsilon\mathbf{A}$ cannot be used either. Again, an example proves the point.

Theorem 119 *Distinct congruence θ and θ' on $\mathbf{A}$ can have the same restriction to $\varepsilon\mathbf{A}$.*

Proof. Let $\mathbf{A}$ be the six-element algebra with $0, 1, a = \neg a, b = \neg b, a \wedge b$ and $a \vee b$ all distinct; and for all $x \in A$, $\nu x = 1$. Then the only complemented members of A are 0 and 1, so $\varepsilon\mathbf{A} \cong \underline{\mathbf{2}}^{+}$ and its only congruences are Δ and ∇. But the congruence $\theta_{(a\vee b,1)}$ on $\mathbf{A}$ is distinct from both Δ and ∇: $\mathbf{A}/\theta_{(a\vee b,1)} \cong S_1$. So $\theta_{(a\vee b,1)} \neq \Delta$, but both have the same restriction to $\varepsilon\mathbf{A}$, namely Δ. □

The relation between $\mathbf{Con}(\mathbf{A})$ and $\mathbf{Con}(\eta\mathbf{A})$ can be described more formally in the following way. A lattice $\mathbf{L}$ may be regarded as a category with objects the elements of its universe L; and for $a, b \in L$ with $a \to b$ a morphism iff $a \leq b$. Regarding $\mathbf{Con}(\mathbf{A})$ and $\mathbf{Con}(\eta\mathbf{A})$ in this way, we have

Theorem 120 $< \mathcal{E}, \mathcal{R} >$: $\mathbf{Con}(\mathbf{A}) \rightharpoonup \mathbf{Con}(\eta\mathbf{A})$, *with* $\mathcal{R}\mathcal{E} = 1_{\mathbf{Con}(\mathbf{A})}$.

Proof. It is easy to see that $\mathcal{E}$ and $\mathcal{R}$ are monotone - they preserve the lattice ordering. That $\mathcal{R}\mathcal{E} = 1_{\mathbf{Con}(\mathbf{A})}$ is clear from the above discussion, as is $\mathcal{E}\mathcal{R} \leq 1_{\mathbf{Con}(\eta\mathbf{A})}$: we saw that for $\theta \in \mathbf{Con}(\mathbf{A})$, $\mathcal{E}(\theta) = \bigwedge\{\mathcal{R}^{-1}(\theta)\}$, so if for $\psi \in \mathbf{Con}(\eta\mathbf{A})$ we have $\mathcal{R}(\psi) = \theta$, then $\psi \in \mathcal{R}^{-1}(\theta)$ and so $\mathcal{E}\mathcal{R}(\psi) \leq \psi$. Then for $\theta \in \mathbf{Con}(\mathbf{A})$ and $\psi \in \mathbf{Con}(\eta\mathbf{A})$, it follows that $\mathcal{E}(\theta) \leq \psi$ iff $\theta \leq \mathcal{R}(\psi)$: for if $\mathcal{E}(\theta) \leq \psi$ then $\mathcal{R}\mathcal{E}(\theta) \leq \mathcal{R}(\psi)$. But $\theta \leq \mathcal{R}\mathcal{E}(\theta)$ so $\theta \leq \mathcal{R}(\psi)$. The other direction is similar, using the fact that $\mathcal{E}\mathcal{R}(\psi) \leq \psi$. This shows the adjunction exists. □

From this it follows by general theory that $\mathcal{E}$ is injective and preserves arbitrary sups, and $\mathcal{R}$ is surjective and preserves arbitrary infs, although it was already fairly obvious in the context.

It is now shown that whatever description of filters is arrived at, it is not enough to characterise congruences: the filter $1/\theta$ need not determine θ. For $\mathbf{A}$ a modal algebra, let $\mathbf{Filt}(\mathbf{A})$ be the lattice under inclusion of $\{1/\theta \mid \theta \in \mathbf{Con}(\mathbf{A})\}$. As we have seen, $\mathbf{Filt}(\eta\mathbf{A})$ is precisely the lattice of strongly $\neg$-consistent and open filters of $\eta\mathbf{A}$, and $\mathbf{Filt}(\eta\mathbf{A}) \cong \mathbf{Con}(\eta\mathbf{A})$. For $1/\theta \in \mathbf{Filt}(\mathbf{A})$, let $\mathcal{L}(1/\theta) = 1/\mathcal{E}(\theta)$. This is the least strongly $\neg$-consistent and open filter of $\eta\mathbf{A}$ containing $\{\eta_A(a) \mid a \in 1/\theta\}$. In the other direction, for $1/\psi \in \mathbf{Filt}(\eta\mathbf{A})$ let $\mathcal{M}(1/\psi) = 1/\mathcal{R}(\psi)$: regarding $\mathbf{A}$ as a subalgebra of $\eta\mathbf{A}$, we have $\mathcal{M}(1/\psi) = 1/\psi \cap A$. Then it can easily be shown that

Theorem 121 $<\mathcal{L}, \mathcal{M}>: \mathbf{Filt}(\mathbf{A}) \rightharpoonup \mathbf{Filt}(\eta\mathbf{A})$, *with* $\mathcal{M}\mathcal{L} = 1_{\mathbf{Filt}(\mathbf{A})}$. □

Because of the isomorphism $\mathbf{Filt}(\eta\mathbf{A}) \cong \mathbf{Con}(\eta\mathbf{A})$ we may regard this as an adjunction $<\mathcal{L}, \mathcal{M}>: \mathbf{Filt}(\mathbf{A}) \rightharpoonup \mathbf{Con}(\eta\mathbf{A})$. Then even though the direction of the two adjunctions we have do not match naturally, we are still able to compose them in the following way:

Theorem 122 $<\mathcal{R}\mathcal{L}, \mathcal{M}\mathcal{E}>: \mathbf{Filt}(\mathbf{A}) \rightharpoonup \mathbf{Con}(\mathbf{A})$, *with* $\mathcal{M}\mathcal{E}\mathcal{R}\mathcal{L} = 1_{\mathbf{Filt}(\mathbf{A})}$.

Proof. First, it can be seen that $\mathcal{M}\mathcal{E}(\theta) = 1/\theta$: for $a \in A$, with a also regarded as an element of $\eta\mathbf{A}$, $a\theta 1$ iff $a\mathcal{E}(\theta)1$ iff $a \in 1/\mathcal{E}(\theta)$; and $\mathcal{M}\mathcal{L} = 1_{\mathbf{Filt}(\mathbf{A})}$, so this is so iff $a \in 1/\mathcal{E}(\theta) \cap A = \mathcal{M}\mathcal{E}(\theta)$. Also, $\mathcal{R}\mathcal{L}(F)$ is the least congruence $\theta \in \mathbf{Con}(\mathbf{A})$ with $1/\theta = F$: $\mathcal{L}(F)$ is the smallest congruence ψ on $\eta\mathbf{A}$ such that for all $a \in F$, $a\psi 1$; and $\mathcal{M}\mathcal{L}(F) = F$, so $a\mathcal{L}(F)1$ iff $a \in F$, which means that its restriction to $\mathbf{A}$ is the smallest on $\mathbf{A}$ with this property. This shows that $\mathcal{M}\mathcal{E}\mathcal{R}\mathcal{L} = 1_{\mathbf{Filt}(\mathbf{A})}$; and the fact that $\mathcal{R}\mathcal{L}(1/\theta)$ is the *smallest* congruence ψ with $1/\psi = 1/\theta$, means that $\mathcal{R}\mathcal{L}\mathcal{M}\mathcal{E} \leq 1_{\mathbf{Con}(\mathbf{A})}$. Together, these complete the proof of the adjunction. □

It can now be seen that even if the congruences of $\eta\mathbf{A}$ can be used to provide a characterisation of $\mathbf{Con}(\mathbf{A})$, then this does not mean that for $\theta \in \mathbf{Con}(\mathbf{A})$, $1/\theta$ determines θ, as happens if $\mathbf{A} \cong \eta\mathbf{A}$. It is easier to see that the converse is also true: the fact that it may happen that $\mathbf{Filt}(\mathbf{A}) \cong \mathbf{Con}(\mathbf{A})$ does not mean that congruences on $\mathbf{A}$ have a unique extension to $\eta\mathbf{A}$. A case in point is the four element algebra $\mathbf{A}$ above with $0 < a < \neg a < 1$: the three congruences are determined by the sets $\{0, a, \neg a, 1\}$, $\{\neg a, 1\}$ and $\{1\}$, and these are the only possible restrictions to $\mathbf{A}$ of strongly $\neg$-consistent and open filters on $\eta\mathbf{A}$ since these are the only filters of $\mathbf{A}$. Yet it was shown that distinct congruences on $\eta\mathbf{A}$ had the same restriction to $\mathbf{A}$.

To complete the proof that neither equivalence entails the other,

Theorem 123 *It is possible that* $\mathbf{Con}(\mathbf{A}) \cong \mathbf{Con}(\eta\mathbf{A})$*, and yet that for* $\theta \in \mathbf{Con}(\mathbf{A})$*,* $1/\theta$ *does not determine* θ*.*

Proof. Consider again the six-element algebra $\mathbf{A}$ given above with universe $0, 1, a = \neg a, b = \neg b, a \wedge b$ and $a \vee b$. Now, $\eta\mathbf{A}$ happens to be the sixteen element algebra $S_1 \times S_1$: there is more than one embedding of $\mathbf{A}$ in $S_1 \times S_1$, so fix η_A to be the homomorphism determined by $\eta_A(a) =< a, a >$ and $\eta_A(b) =< a, b >$ - the latter a and b name the elements of the universe of S_1 given in the previous chapter. $\mathbf{A}$ has four congruences, namely Δ, ∇, $\theta_{(a,b)}$ and $\theta_{(a\vee b,1)}$, so $\eta\mathbf{A}$ has at least four: the extensions of these congruences. But these are in fact the only congruences on $\eta\mathbf{A}$, as consideration of the strongly $\neg$-consistent and open filters of $\eta\mathbf{A}$ shows:

$$\mathcal{E}\left(\theta_{(a,b)}\right) = \theta_{(\eta_A(a),\eta_A(b))} = \theta_{(\eta_A(a)+\eta_A(b),1)} = \theta_{(<1,0>,1)};$$

and for $\mathcal{E}\left(\theta_{(a\vee b,1)}\right)$, since $\eta_A(a \vee b) =< a, 1 >$ and $-\neg < a, 1 >=< b, 1 >$, the least strongly $\neg$-consistent and open filter containing $\eta_A(a \vee b)$ must contain $\eta_A(a \vee b) \wedge -\neg\eta_A(a \vee b) =< 0, 1 >$. So $\mathcal{E}\left(\theta_{(a\vee b,1)}\right) = \theta_{(<0,1>,1)}$. Because both $-\neg < 1, 0 >=< 1, 0 >$ and $-\neg < 0, 1 >=< 0, 1 >$ the filters in the two cases are the principal filters generated by $< 1, 0 >$ and $< 0, 1 >$. Given how negation was defined on S_1, apart form these two filters there are no strongly $\neg$-consistent filters other that the whole universe ηA containing either of these. Such filters contain an element c just in case they also contain $c \wedge -\neg c$, and the only elements of this form in $S_1 \times S_1$ are $< 1, 1 >$, $< 1, 0 >$, $< 0, 1 >$ and $< 0, 0 >$. But the intersection of these filters is the filter $\{1\}$, so these are the only four congruences on $\eta\mathbf{A}$, and $\mathbf{Con}(\mathbf{A}) \cong \mathbf{Con}(\eta\mathbf{A})$. Thus Δ and $\theta_{(a,b)}$ provide the counterexample showing that $\mathbf{Con}(\mathbf{A}) \not\cong \mathbf{Filt}(\mathbf{A})$, even though $\mathbf{Con}(\mathbf{A}) \cong \mathbf{Con}(\eta\mathbf{A})$. $\square$

A trivial but almost useless characterisation of $\mathbf{Con}(\mathbf{A})$ *is* possible, and was implicit in the adjunctions described. Let a strongly $\neg$-consistent and open filter F in $\eta\mathbf{A}$ be called *A-generated* if it is the least such filter containing $\{a+b \mid a+b \in F \text{ and } a,b \in A\}$. Then such filters are in one-one correspondence with the congruences of $\mathbf{A}$. This follows from the fact that, because $\mathcal{E}$ is a left adjoint and so preserves sups, using the correspondence between $\mathbf{Con}(\eta\mathbf{A})$ and $\mathbf{Filt}(\eta\mathbf{A})$ we have

$$\mathcal{E}(\theta) = \mathcal{E}(\bigvee\{\theta_{(a,b)} \mid a\theta b\}) = \bigvee\{\mathcal{E}(\theta_{(a,b)}) \mid a\theta b\} = \bigvee\{(\theta_{(a+b,1)}) \mid a\theta b\}.$$

7.3 Their Logics

We conclude with a brief word on one of the more global connections that arises from the connection between modal algebras and complemented modal algebras, namely the relation between their logics. In frames, let the complementation operation be understood in the obvious way, with $\mathbf{C}, v, x \models -\alpha$ iff $\mathbf{C}, v, x \not\models \alpha$. The usual classical sequent rules for complement may then be added to the logic of $\mathcal{MA}$ to give the logic of $\mathcal{MA}^-$, and the relations between frames for this logic and algebras in $\mathcal{MA}^-$ can be worked out in just the way previously done for the algebras of $\mathcal{MA}$. For example for $\mathbf{C}^+$ a $\mathcal{MA}^-$-algebra and α a formula of the extended language of $\mathcal{MA}^-$, we again have $\mathbf{C} \models \alpha$ iff $\mathbf{C}^+ \models \alpha$: so the same algebraic characterisation of complete varieties will apply to varieties $\mathcal{MA}^-$-algebras.

With this in mind, let $L^{V(\mathcal{MA})}$ and $L^{V(\mathcal{MA}^-)}$ denote the lattices of varieties of $\mathcal{MA}$-algebras and of varieties of $\mathcal{MA}^-$-algebras respectively. Also, for $K \subseteq \mathcal{MA}$ let $\mathcal{H}(K)$ abbreviate $\{\mathcal{H}(\mathbf{A}) \mid \mathbf{A} \in K\}$; and for $K' \subseteq \mathcal{MA}^-$ let $\mathcal{G}(K')$ abbreviate

$\{\mathcal{G}(\mathbf{B}) \mid \mathbf{B} \in K'\}$. We sometimes write $V\mathcal{G}(K')$ for $V(\mathcal{G}(K'))$, the smallest $\mathcal{MA}$-variety containing $\mathcal{G}(K')$: other such omissions of brackets should not give rise to ambiguity. Then we have the following result.

Theorem 124 $< V\mathcal{G}, \mathcal{H} >: L^{V(\mathcal{MA}^-)} \rightharpoonup L^{V(\mathcal{MA})}$.

Proof. Obviously for $W \in L^{V(\mathcal{MA}^-)}$ we have $V(\mathcal{G}(W)) \in L^{V(\mathcal{MA})}$, and if $W \leq W'$ then $\mathcal{G}(W) \subseteq \mathcal{G}(W')$ and so $V(\mathcal{G}(W)) \leq V(\mathcal{G}(W'))$, showing that $V\mathcal{G}$ is a functor. To show that $\mathcal{H}$ is a mapping from $L^{V(\mathcal{MA})}$ to $L^{V(\mathcal{MA}^-)}$ it must be shown that for $V \in L^{V(\mathcal{MA})}$, $\mathcal{H}(V)$ is a $\mathcal{MA}^-$-variety. Now we have seen that $\mathcal{H}$ is a right adjoint and so preserves limits: that is

$$\mathcal{H}(\prod_{i \in I} \mathbf{A}_i) \cong \prod_{i \in I} \mathcal{H}(\mathbf{A}_i).$$

Given the definition of $\mathcal{H}(\mathbf{A}_i)$ this can be seen to be so because $a \in \prod_{i \in I} \mathbf{A}_i$ is complemented precisely if for all $i \in I$, $a(i)$ is complemented in $\mathbf{A}_i$. Then because varieties are obviously closed under products, if $\mathbf{B} \in HSP(\mathcal{H}(V))$ then $\mathbf{B} \in HS(\mathcal{H}(V))$. Now because $\mathcal{G}$ is the forgetful functor it is easy to see that for any $\mathcal{MA}^-$-homomorphism g we have g injective iff $\mathcal{G}g$ is injective, and g is surjective iff $\mathcal{G}g$ is surjective. So if $\mathbf{B}$ is a subalgebra of $\mathcal{H}(\mathbf{A})$ then $\mathcal{G}(\mathbf{B})$ is a subalgebra of $\mathcal{G}\mathcal{H}(\mathbf{A}) = \varepsilon\mathbf{A}$, which as we saw is in turn a subalgebra of $\mathbf{A}$. Varieties are closed under subalgebras, so if $\mathbf{B} \in S(\mathcal{H}(V))$ then $\mathcal{H}\mathcal{G}(\mathbf{B}) \cong \mathbf{B} \in \mathcal{H}(V)$. It may be shown in a similar fashion that if $\mathbf{B} \in H(\mathcal{H}(V))$ then $\mathbf{B} \in (\mathcal{H}(V))$, completing the proof that $\mathcal{H}(V)$ is a $\mathcal{MA}^-$-variety. Clearly if $V \leq V'$ then $\mathcal{H}(V) \leq \mathcal{H}(V')$, so $\mathcal{H}$ is a functor between the two lattices. To prove the adjunction, suppose that $W \in L^{V(\mathcal{MA}^-)}$ and $V \in L^{V(\mathcal{MA})}$. Then if $V(\mathcal{G}(W)) \leq V$ and $\mathbf{B} \in W$, we have $\mathbf{B} \cong \mathcal{H}\mathcal{G}(\mathbf{B}) \in \mathcal{H}(V(\mathcal{G}(W)))$, so $W \leq \mathcal{H}(V(\mathcal{G}(W))) \leq \mathcal{H}(V)$. And if $W \leq \mathcal{H}(V)$, note that $\mathcal{G}\mathcal{H}(V) = \varepsilon(V)$ are the complemented algebras in V, so $V(\varepsilon(V)) \leq V$; so $V(\mathcal{G}(W)) \leq V(\varepsilon(V)) \leq V$. □

If $V \in L^{V(\mathcal{MA})}$ is complete, then it is generated by its powerset algebras and so $V = V(\varepsilon(V))$: this shows that complete logics 'admit complementation as a rule'. If complement is added to the language, then closing a complete logic under its rules does not provide any new theorems in the original language. This shows that the modal part of these logics can be presented and discussed in a base language of classical logic, as recommended by our methodology. In general, however, we do not have $V = V(\varepsilon(V))$. Nor do we have, for all $W \in L^{V(\mathcal{MA}^-)}$, $W = \mathcal{H}(V(\mathcal{G}(W)))$: this fails essentially because of the failure in $\mathcal{MA}$ of the congruence extension property. To arrive at a reflective or co-reflective subcategory relation, as was done above with $\mathcal{MA}$ and $\mathcal{MA}^-$, it is necessary to weaken slightly the classes of algebras by introducing the definition of a *quasi-variety*. This move is parallel to the logical move required to interpret $\mathcal{MA}^-$-logics as $\mathcal{MA}$-logics by introducing certain easily determined higher-order axioms, or *rules*.

Definition 125 *A quasi-variety is a class of algebras closed under S, P, P_U and isomorphism, and containing a trivial algebra.* □

Let $L^{qV(\mathcal{MA})}$ and $L^{qV(\mathcal{MA}^-)}$ be the lattices of quasi-varieties of $\mathcal{MA}$-algebras and $\mathcal{MA}^-$-algebras respectively. We are interested in extending $\mathcal{H} : L^{V(\mathcal{MA})} \longrightarrow L^{V(\mathcal{MA}^-)}$ to $V\mathcal{H} : L^{qV(\mathcal{MA})} \longrightarrow L^{V(\mathcal{MA}^-)}$, but there are also adjunctions involving $L^{qV(\mathcal{MA}^-)}$ of which those involving $L^{V(\mathcal{MA}^-)}$ and the above functor may be seen as restrictions.

Theorem 126 *Let $V \in L^{V(\mathcal{MA}^-)}$ and $Q \in L^{qV(\mathcal{MA}^-)}$. Then the following relations hold:*

1. $<\mathcal{H}, \mathcal{H}^{-1}>: L^{qV(\mathcal{MA})} \rightharpoonup L^{qV(\mathcal{MA}^-)}$ *with* $\mathcal{H}(\mathcal{H}^{-1}(Q)) = Q$;

2. $<V\mathcal{H}, \mathcal{H}^{-1}>: L^{qV(\mathcal{MA})} \rightharpoonup L^{V(\mathcal{MA}^-)}$ *with* $V\mathcal{H}(\mathcal{H}^{-1}(V)) = V$;

3. $<S\mathcal{G}, \mathcal{H}>: L^{qV(\mathcal{MA}^-)} \rightharpoonup L^{qV(\mathcal{MA})}$ *with* $\mathcal{H}(S\mathcal{G}(Q)) = Q$;

4. $<S\mathcal{G}, V\mathcal{H}>: L^{V(\mathcal{MA}^-)} \rightharpoonup L^{qV(\mathcal{MA})}$ *with* $V\mathcal{H}(S\mathcal{G}(V)) = V$.

Proof. For the first adjunction, it should be clear that $\mathcal{H}$ and $\mathcal{H}^{-1}$ preserve the lattice ordering; to see that they are mappings, it must be shown that, for $Q \in L^{qV(\mathcal{MA}^-)}$ and $R \in L^{qV(\mathcal{MA})}$, $\mathcal{H}^{-1}(Q)$ and $\mathcal{H}(R)$ are quasi-varieties. It has already been shown that $\mathcal{H}(\prod_{i\in I}\mathbf{A}_i) \cong \prod_{i\in I}\mathcal{H}(\mathbf{A}_i)$, and so these classes are closed under products. To see that both are closed under ultraproducts, it is enough to show that $\mathcal{H}(\prod_{i\in I}\mathbf{A}_i/U) \cong \prod_{i\in I}\mathcal{H}(\mathbf{A}_i)/U$. First, using the subalgebra relations $\varepsilon\mathbf{A}_i \to \mathbf{A}_i$, it is straightforward to verify that the natural mapping

$$\prod_{i\in I}\mathcal{H}(\mathbf{A}_i)/U \xrightarrow{h} \mathcal{H}(\prod_{i\in I}\mathbf{A}_i/U)$$

is a well-defined injective homomorphism - this may be more easily seen by considering the equivalent

$$\mathcal{G}(\prod_{i\in I}\mathcal{H}(\mathbf{A}_i)/U) \longrightarrow \varepsilon(\prod_{i\in I}\mathbf{A}_i/U),$$

since $\mathcal{G}(\prod_{i\in I}\mathcal{H}(\mathbf{A}_i)/U) \cong \prod_{i\in I}\varepsilon\mathbf{A}_i/U$ is a complemented subalgebra of $\prod_{i\in I}\mathbf{A}_i/U$. To show that h is surjective, it is enough to show that if $a/U \in \prod_{i\in I}\mathbf{A}_i/U$ is complemented, then for some complemented $a' \in \prod_{i\in I}\mathbf{A}_i$, $a/U = a'/U$. So let $a/U \in \prod_{i\in I}\mathbf{A}_i/U$ have a complement b/U, and note that in any algebra c and d are complements just in case $(c \vee d) \wedge (\neg c \vee \neg d) = 1$. From this it follows that $J = \{i \in I \mid (a(i) \vee b(i)) \wedge (\neg a(i) \vee \neg b(i)) = 1\} \in U$, and that for each $i \in J$, $a(i)$ is complemented. Defining a' such that $a'(i) = a(i)$ for $i \in J$, and $a'(i) = 1$, say, otherwise, a' is complemented and $a'/U = a/U$. So regarding a' as an element of $\prod_{i\in I}\mathcal{H}(\mathbf{A}_i)$, we have $h(a'/U) = a/U$. So h is an isomorphism and $\mathcal{H}^{-1}(Q)$ and $\mathcal{H}(R)$ are closed under ultraproducts.

It remains to show that they are closed under subalgebras. To show this true of $\mathcal{H}^{-1}(Q)$, let $\mathcal{H}(\mathbf{A}) \in Q$ and let $\mathbf{B}$ be a subalgebra of $\mathbf{A}$. Then any complemented element of $\mathbf{B}$ may also be regarded as a complemented element of $\mathbf{A}$, so $\varepsilon\mathbf{B}$ is a subalgebra of $\varepsilon\mathbf{A}$; but $\mathcal{H}$ and $\mathcal{H}\varepsilon$ are identical morphisms, so $\mathcal{H}(\mathbf{B})$ is a subalgebra of $\mathcal{H}(\mathbf{A})$. Also, Q is closed under subalgebras, so $\mathcal{H}(\mathbf{B}) \in Q$. To show that $\mathcal{H}(R)$ is closed under subalgebras, let $\mathbf{B}$ be a subalgebra of $\mathcal{H}(\mathbf{A})$ for some $\mathbf{A} \in R$: then clearly $\mathcal{G}(\mathbf{B})$ is a subalgebra of $\varepsilon\mathbf{A}$. But R is closed under subalgebras so $\mathcal{G}(\mathbf{B}) \in R$; and $\mathcal{H}\mathcal{G}(\mathbf{B}) = \mathbf{B}$, so $\mathbf{B} \in \mathcal{H}(R)$.

To complete the proof of the first adjunction, the fact that for $\mathbf{B} \in Q$, also $\mathbf{B} = \mathcal{H}\mathcal{G}(\mathbf{B}) \in Q$ shows that $Q \leq \mathcal{H}(\mathcal{H}^{-1}(Q))$. The other direction is equally obvious: if $\mathbf{B} \in \mathcal{H}(\mathcal{H}^{-1}(Q))$, then $\mathbf{B} = \mathcal{H}(\mathbf{A})$, for some $\mathbf{A} \in \mathcal{H}^{-1}(Q)$: but this just means that $\mathcal{H}(\mathbf{A}) = \mathbf{B} \in Q$, proving the first adjunction. Given that varieties are quasi-varieties, the second adjunction now follows almost immediately; note that here and in the fourth adjunction $V\mathcal{H}$ may be simplified to $H\mathcal{H}$.

All that really needs to be shown for the third part of the proof is that $S\mathcal{G}$ is a mapping and that for $Q \in L^{qV(\mathcal{MA}^-)}$, $\mathcal{H}(S\mathcal{G}(Q)) = Q$. But given that $\mathcal{G}(\prod_{i\in I}\mathbf{A}_i) \cong \prod_{i\in I}\mathcal{G}(\mathbf{A}_i)$ and $\mathcal{G}(\prod_{i\in I}\mathbf{A}_i/U) \cong \prod_{i\in I}\mathcal{G}(\mathbf{A}_i)/U$ are immediate, it is not difficult to see that $S\mathcal{G}$ is a mapping. For the second part, the fact that $\mathcal{H}\mathcal{G}(\mathbf{A}) = \mathbf{A}$ shows that $Q \leq \mathcal{H}(S\mathcal{G}(Q))$, so for the other direction let $\mathbf{B} \in Q$ and $\mathbf{A} = \mathcal{H}(\mathbf{B}')$, where $\mathbf{B}'$ is a subalgebra of $\mathcal{G}(\mathbf{B})$. But then $\varepsilon\mathbf{B}'$ is a subalgebra of $\mathcal{H}\mathcal{G}(\mathbf{B}) = \mathbf{B}$ and so $\mathcal{H}(\varepsilon\mathbf{B}') = \mathcal{H}(\mathbf{B}') = \mathbf{A}$ is a subalgebra of $\mathcal{H}\mathcal{G}(\mathbf{B}) = \mathbf{B}$: but Q is closed under subalgebras, so $\mathbf{A} \in Q$, showing the third adjunction. The fourth then follows from this in a straightforward manner. □

We now close as promised by hinting at the change of logical direction implicit in the above. The theorem shows that $L^{V(\mathcal{MA}^-)}$ partitions $L^{qV(\mathcal{MA})}$ into intervals $[S\mathcal{G}(V), \mathcal{H}^{-1}(V)]$, for $V \in L^{V(\mathcal{MA}^-)}$, each quasi-variety in which has the same complemented algebras. The embedding given by the fourth adjunction, however, provided the more logically natural representation of $L^{V(\mathcal{MA}^-)}$ within $L^{qV(\mathcal{MA})}$. First note that every defining identity of a variety of $\mathcal{MA}^-$-algebras is of the form $t^\alpha = 1$, for some formula α in the language of $\mathcal{MA}^-$. Then an easy induction shows that for any such α we may assume that all occurrences of the connective $-$ have narrowest scope: the proof uses the fact that it distributes over the other propositional connectives, and the facts that $-\mu x = \neg\nu x$ and $-\nu x = \neg\mu x$. This gives the 'essential' occurrences of the complementation connective in α. An $\mathcal{MA}^-$-formula α in n variables in this way determines a

$\mathcal{MA}$-formula α' in $n+k$ variables, where for each of the k pairs $\{p_i, -p_i\}$ occurring essentially in α a fresh variable p_i' is substituted for $-p_i$ in α. Suppose for simplicity this happened in the argument places $1, \ldots k$, and so the result is the $\mathcal{MA}$-formula α' in the variables $\{p_1, \ldots p_n, p_1', \ldots p_k'\}$. These argument places can then in a sense be abstracted to define the conditional identity - or ***quasi-identity*** - $t^{q(\alpha)}$ by

$$\bigwedge_{1 \leq i \leq k} (x_i \vee y_i) \wedge (\neg x_i \vee \neg y_i) = 1 \implies t^{\alpha'}(x_1, \ldots, x_n, y_1, \ldots y_k) = 1.$$

Quasi-identities bear to quasi-varieties the relation that identities bear to varieties, and it is not too difficult to see that a variety $V \in L^{V(\mathcal{MA}^-)}$ defined by the theorems $\{\alpha_i \mid i \in I\}$ is equivalently defined by the quasi-identities $\{t^{q(\alpha_i)} \mid i \in I\}$. From this it can be shown that the same set of quasi-identities defines the quasi-variety $S\mathcal{G}(V)$. So $\mathcal{MA}^-$-logics can be represented within the language and models appropriate to $\mathcal{MA}$, these logics being defined by allowing as proper assumptions the ***rules*** - rather than simple axioms - determined by the quasi-identities. In this way the relation between the logics of the two languages is represented in a uniform, extended proof system corresponding to the move to quasi-varieties, but further examination of these issues would mean straying outside the domain of the study of models for four-valued logic.

Bibliography

[1] Anderson, A. R. and Belnap, N. D. (1975) *Entailment,* vol. 1. Princeton, N.J.: Princeton University Press.

[2] Apt, K. and van Emden, M. H. (1982) Contributions to the Theory of Logic Programming. *Journal of the ACM,* **29**. 841-862.

[3] Asher, N. and Kamp, J. (1986) The Knower's Paradox and Representational Theories of Attitudes. In Halpern, J.Y. (ed.) *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the 1986 Conference,* Los Altos: Morgan Kaufmann. 131-148.

[4] Barwise, J. and Perry, J. (1983) *Situations and Attitudes.* Cambridge, Mass.: MIT Press.

[5] van Benthem, J. (1985) *A Manual of Intensional Logic.* CSLI.

[6] van Benthem, J. (1986) Partiality and Nonmonotonicity in Classical Logic. *Logique et Analyse,* **114**. 225-247.

[7] Blamey, S. (1986) Partial Logic. Chapter 1 of Gabbay, D. and Guenthner, F. (eds.) *Handbook of Philosophical Logic.* vol. 3. Dordrecht: D. Reidel. 1-70.

[8] Blok, W. J. and Kohler, P. (1983) Algebraic Semantics for Quasi-Classical Modal Logics. *Journal of Symbolic Logic,* **48**. 941-964.

[9] Blok, W. J. (1980) The Lattice of Modal Logics: An Algebraic Investigation. *Journal of Symbolic Logic,* **45;2**. 221-236.

[10] Boolos, G. S. and Jeffrey, R. C. (1980) *Computability and Logic,* second edition. Cambridge: Cambridge University Press.

[11] Bull, R. and Segerberg, K. (1984) Basic Modal Logic. Chapter 1 of Gabbay, D. and Guenthner, F. (eds.) (1984) *Handbook of Philosophical Logic,* vol. 2. Dordrecht: D. Reidel. 1-88.

[12] Chellas, B. F. (1980) *Modal Logic.* Cambridge: Cambridge University Press.

[13] Church, A. (1951) The Weak Theory of Implication. In *Kontrolliertes Denken,* Munich.

[14] Cresswell, M. J. (1973) *Logics and Languages.* London: Methuen.

[15] Cresswell, M. J. (1985) *Structured Meanings.* Cambridge, Mass.: MIT Press.

[16] Deutsch, H. (1984) Paraconsistent Analytic Implication. *Journal of Philosophical Logic,* **13**. 1-11.

[17] Dunn, J. M. (1976) Intuitive Semantics for First-degree Entailments and 'Coupled Trees'. *Philosophical Studies,* **29**. 149-168.

[18] Dunn, J. M. (1986) Relevance Logic and Entailment. Chapter 3 of Gabbay, D. and Guenthner, F. (eds.) *Handbook of Philosophical Logic.* vol. 3. Dordrecht: D. Reidel. 117-224.

[19] Fagin, R. and Halpern, J. (1985) Belief, Awareness and Limited Reasoning: Preliminary Report. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence.* University of California at Los Angeles, Los Angeles, Ca.. 491-501.

[20] Fine, K. (1974) Models for Entailment. *Journal of Philosophical Logic,* **3**. 347-372.

[21] Fine, K. (1986) Analytic Implication. *Notre Dame Journal of Formal Logic,* **27**. 169-179.

[22] Fine, K. (1988) Semantics for Quantified Relevance Logic. *Journal of Philosophical Logic,* **17**. 27-59.

[23] Fodor, J. A. (1978) Propositional Attitudes. *The Monist,* **61**. 501-523.

[24] Gabbay, D. and Guenthner, F. (eds.) (1984) *Handbook of Philosophical Logic,* vol. 2. Dordrecht: D. Reidel.

[25] Gabbay, D. and Guenthner, F. (eds.) (1986) *Handbook of Philosophical Logic,* vol. 3. Dordrecht: D. Reidel.

[26] Gallin, D. (1975) *Intensional and Higher-Order Modal Logic with Applications to Montague Semantics.* Amsterdam: North-Holland.

[27] Garson, J. W. (1985) Quantification In Modal Logic. Chapter 5 of Gabbay, D. and Guenthner, F. (eds.) *Handbook of Philosophical Logic.* vol. 2. Dordrecht: D. Reidel. 249-307.

[28] Goldblatt, R. (1989) Varieties of Complex Algebras. *Annals of Pure and Applied Logic,* **44**. 173-242.

[29] Gupta, A. (1982) Truth and Paradox. *Journal of Philosophical Logic,* **11**. 1-60.

[30] Halpern, J. and Moses, Y. (1985) Guide to Modal Logics of Knowledge and Belief: Preliminary Draft. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence.* University of California at Los Angeles, Los Angeles, Ca.. 480-490.

[31] Helman, G. (1977) Completeness of the Normal Typed Fragment of the λ-system *U*. *Journal of Philosophic Logic,* **6**. 33-46.

[32] Herzberger, H. (1982) Naive Semantics and the Liar Paradox. *Journal of Philosophy,* **79**. 479-497.

[33] Hintikka, J. (1962) *Knowledge and Belief.* Ithaca, N.Y.: Cornell University Press.

[34] Hintikka, J. (1975) Impossible Possible Worlds Vindicated. In *Game-Theoretical Semantics.* Dordrecht: D. Reidel.

[35] Humberstone, I. L. (1981) From Worlds to Possibilities. *Journal of Philosophical Logic,* **10**. 313-344.

[36] Humberstone, I. L. (1985) The Formalities of Collective Omniscience. *Philosophical Studies,* **48**. 401-423.

[37] Jackendoff, R. (1976) Toward an Explanatory Semantic Representation. *Linguistic Inquiry,* **7**. 89-150.

[38] Katz, J. (1977) The Advantage of Semantic Theory over Predicate Calculus in the Representation of Logical Form in Natural Language. *The Monist,* **60**. 380-405.

[39] Kearns, J. T. (1979) The Strong Completeness of a System for Kleene's Three-valued Logic. *Zeitschrift für mathematische Logik und Grundlagen der Arithmetik,* **25**. 61-68.

[40] Kleene, S. C. (1950) *Introduction to Metamathematics.* Princeton: von Nostrand.

[41] Konolige, K (1983) A Deductive Model of Belief. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence,* Karlsruhe, West Germany. 377-381.

[42] Konolige, K. (1986) What Awareness Isn't: A Sentential View of Implicit and Explicit Belief. In Halpern, J.Y. (ed.) *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the 1986 Conference,* Los Altos: Morgan Kaufmann. 241-250.

[43] Kripke, S. (1975) Outline of a Theory of Truth. *Journal of Philosophy,* **72**. 690-716.

[44] Lakemeyer, G. (1986) Steps Towards a First-order Logic of Explicit and Implicit Belief. In Halpern, J.Y. (ed.) *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the 1986 Conference,* Los Altos: Morgan Kaufmann. 325-340.

[45] Lakemeyer, G. (1987) Tractable Meta-reasoning in Propositional Logics of Belief. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence,* Milan, Italy. 402-408.

[46] Levesque, H. J. (1984) A Logic of Implicit and Explicit Belief. In *Proceedings of the Annual Meeting of the American Association for Artificial Intelligence,* Austin, Texas. 1984.

[47] Löb, M. (1955) Solution of Problem of Leon Henkin. *Journal of Symbolic Logic,* **20**. 115-118.

[48] Loureiro, I. (1985) Principal Congruences of Tetravalent Modal Algebras. *Notre Dame Journal of Formal Logic,* **26;1**. 76-80.

[49] MacLane, S. (1971) *Categories for the Working Mathematician.* Springer-Verlag.

[50] Martin, R. L. (ed.) (1984) *Recent Essays on Truth and Paradox*. Oxford: Oxford University Press.

[51] Mitchell, J. C. and O'Donnell, M. J. (1986) Realizability Semantics for Error-tolerant Logics. In Halpern, J.Y. (ed.) *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the 1986 Conference,* Los Altos: Morgan Kaufmann. 363-381.

[52] Montague, R. (1963) Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability. *Acta Philosophica Fennica,* **16**. 153-167.

[53] Montague, R. (1968) Pragmatics. In Klibansky, R. (ed.) *Contemporary Philosophy: A Survey,* Florence: La Nuova Italia Editrice. Reprinted in Thomason, R. H. (ed.) *Formal Philosophy: Selected Papers of Richard Montague.* 95-118.

[54] Montague, R. (1970) Universal Grammar. *Theoria,* **36**. 373-398. Reprinted in Thomason, R. H. (ed.) *Formal Philosophy: Selected Papers of Richard Montague.* 222-246.

[55] Montague, R. (1970) The Proper Treatment of Quantification in Ordinary English. In Hintikka, J., Moravcsik, J. M. E. and Suppes, P. (eds.) *Approaches to Natural Language.* Dordrecht: D. Reidel. Reprinted in Thomason, R. H. (ed.) *Formal Philosophy: Selected Papers of Richard Montague.* 247-270.

[56] Moore, R. C. and Hendrix, G. G. (1979) Computational Models Of Beliefs And The Semantics Of Belief-Sentences. Technical Note No. 187, SRI International, Menlo Park, Ca..

[57] Moore, R. C. (1985) Formal Theory of Knowledge and Action. Report No. CSLI-85-31, Center for the Study of Language and Information.

[58] Patel-Schneider, P. F. (1985) A Decidable First-order Logic for Knowledge and Belief. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence,* University of California at Los Angeles, Los Angeles, Ca.. 455-458.

[59] Priestley, H. A. (1970) Representation of Distributed Lattices by means of Ordered Stone Spaces. *Bull. London Math. Soc.,* **2**. 186-190.

[60] Prawitz, D. (1965) *Natural Deduction: A Proof-Theoretic Study.* Uppsala: Almqvist and Wiksell.

[61] Quine, W. V. O. (1979) Intensions Revisited. In French, P. A., Uehling, T. E. and Wettstein, H. K. *Contemporary Perspectives in the Philosophy of Language.* Minneapolis: University of Minnesota Press. 268-274.

[62] Rantala, V. (1975) Urn Models: A New Kind of Non-Standard Model for First-Order Logic. In *Game-Theoretical Semantics.* Dordrecht: D. Reidel.

[63] Rescher, N. and Brandon, R. (1979) *The Logic of Inconsistency.* Rowman and Littlefield.

[64] des Rivières, J. and Levesque, H. (1986) The Consistency of Syntactical Treatments of Knowledge. In Halpern, J.Y. (ed.) *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the 1986 Conference,* Los Altos: Morgan Kaufmann. 115-130.

[65] Rosenschein, S. J. and Kaelbing, L. P. (1986) The Synthesis of Digital Machines with Provable Epistemic Properties. In Halpern, J.Y. (ed.) *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the 1986 Conference,* Los Altos: Morgan Kaufmann. 83-98.

[66] Rosenschein, S. J. (1987) Formal Theories of Knowledge in AI and Robotics. Report No. CSLI-87-84, Centre for the Study of Language and Information.

[67] Scott, D. (1971) On Engendering an Illusion of Understanding. *The Journal of Philosophy,* **68**. 787-807.

[68] Tarski, A., Mostowski, A. and Robinson, R. (1953) *Undecidable Theories.* Amsterdam.

[69] Thistlewaite, P. B., Meyer, R. K. and McRobbie, M. A. (1986) Advanced Theorem-proving Techniques for Relevant Logics. *Logique et Analyse,* **110/111**. 233-256.

[70] Thomason, R. H. (ed.) (1974) *Formal Philosophy: Selected Papers of Richard Montague.* New Haven, Conn.: Yale University Press.

[71] Thomason, R. H. (1980) Model Theory for Propositional Attitudes. *Linguistics and Philosophy,* **4**. 47-70.

[72] Thomason, R. H. (1980) A Note on Syntactical Treatments of Modality. *Synthèse,* **44**. 391-395.

[73] Urquhart, A. (1973) A Semantical Theory of Analytic Implication. *Journal of Philosophical Logic,* **2**. 212-219.

[74] Vardi, M. Y. (1985) A Model-Theoretic Analysis of Monotonic Knowledge. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence,* University of California at Los Angeles, Los Angeles, Ca.. 509-512.

[75] Vardi, M. Y. (1986) On Epistemic Logic and Logical Omniscience. In Halpern, J.Y. (ed.) *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the 1986 Conference,* Los Altos: Morgan Kaufmann. 293-306.