Coalgebraic Logic

Lawrence S. Moss
Indiana University

Dedicated to Rohit Parikh on his 60th birthday.

Abstract

We present a generalization of modal logic to logical systems which are interpreted on coalgebras of functors on sets. The leading idea is that infinitary modal logic contains *characterizing formulas*. That is, every model-world pair is characterized up to bisimulation by an infinitary formula. The point of our generalization is to understand this on a deeper level. We do this by studying a frangment of infinitary modal logic which contains the characterizing formulas and is closed under infinitary conjunction and an operation called $\triangle$. This fragment generalizes to a wide range of *coalgebraic logics*. We then apply the characterization result to get representation theorems for final coalgebras in terms of maximal elements of ordered algebras. The end result is that the formulas of coalgebraic logics can be viewed as approximations to the elements of the final coalgebra.

Keywords: infinitary modal logic, characterization theorem, functor on sets, coalgebra, greatest fixed point.

1 Introduction

The starting point of this paper is the characterization phenomenon of infinitary modal logic. The basic result of this type is that $\langle M, w\rangle$ and $\langle M, v\rangle$ are model-world pairs which satisfy the same formulas of infinitary modal logic (over some fixed set of atomic propositions) iff there is a bisimulation of M relating w and v. This highlights the relation of infinitary modal logic to bisimulation. But infinitary modal logic also has a relation to the power set operation of set theory. Actually, the relation is to a closely related operation which we now describe. Fix a set AtProp of atomic propositions. From AtProp we obtain finitary and infinitary modal logic in the usual way. Kripke models for these logics are *coalgebras* for the functor $\mathsf{F}(A) = \mathcal{P}(A) \times \mathcal{P}(\mathsf{AtProp})$; that is, pairs $\langle A, e\rangle$, where A is a set and $e : A \to \mathcal{P}(A) \times \mathcal{P}(\mathsf{AtProp})$. For each world $a \in A$, $e(a)$ gives the set of worlds accessible from a, together with the set of atomic propositions true of a in the model.

The connection of modal logic to this functor seems to be fairly new. In recent years, a number of authors have considered coalgebras of functors in connection with modeling issues in theoretical computer science; for example, see Aczel [A], Barwise and Moss [BM1], Rutten [R96], and Turi and Rutten [TR]. One important line of work concerns *final coalgebra* theorems. A final coalgebra for a functor often gives a "most abstract" model of a specified kind, and this is the interest of such a result. Now coalgebraic logic is connected to final coalgebras in the following way: we will generalize our fragment of infinitary modal logic to a logic we call $\mathcal{L}_{\mathsf{F}}$,

for each functor F on sets satisfying a few conditions. The logic $\mathcal{L}_{\mathsf{F}}$ will be interpreted on F-coalgebras. One of our goals is that $\mathcal{L}_{\mathsf{F}}$ should be expressive enough so that distinct points in the final coalgebra should differ on some formula of $\mathcal{L}_{\mathsf{F}}$. On the other hand, it should be weak enough so that bisimilar model-world pairs should satisfy the the same formulas of $\mathcal{L}_{\mathsf{F}}$. Another of our goals is to get generalizations of the characterization result.

While practically every functor on sets has a final coalgebra (Aczel and Mendler [AM]), the more concrete a representation one has of the final coalgebra, the better. Final coalgebra theorems of various types may be found in the works cite above, and also in Barr [Bar93, Bar94], Moss and Danner[MD], and in Paulson [P]. The results of this paper also give a final coalgebra: the maximal formulas of $\mathcal{L}_{\mathsf{F}}$ in the semantic preorder. This connects coalgebraic logic with the issue of getting representations of the final coalgebras in terms of some sort of entities that served as approximations. Coalgebraic logics show how this can be done in a general way for a wide class of functors on sets.

Contents of this paper We begin in Section 2 with an overview of the characterization results from infinitary modal logic. Although we present all of the details of this work, one should be advised that we are attempting to be as brief as possible. Also, we are not really presenting the standard approach to modal logic, but rather a reformulation that is suggestive of the generalization provided later in the paper. This is half of the background of the paper, and the other half is found in Section 3. That section reviews some material on functors on sets. The new work of the paper begins in Section 4, where we define coalgebraic logics and their semantics. We study these logics in Section 5, and then Sections 6 and 7 generalize the characterization results of infinitary modal logic to coalgebraic logics. A few remarks on sound logical principles appear in Section 8, but the work there is just the beginning of a study of the logical aspects of the machinery introduced in this paper. In Section 9 we apply our theory to show how formulas of coalgebraic logics can be seen as approximations of members of final coalgebras.

Dedication It is a pleasure to dedicate this paper to Rohit Parikh. Some of the results in this area were first presented at the Seminar in Applications of Logic which he organizes at the CUNY Graduate Center. I am grateful to Rohit and to the other participants in this seminar. I wish Rohit many more years of insightful work in logic and philosophy.

2 Background on Infinitary Modal Logic

In this section, we present a reformulation of infinitary modal logic in terms which we generalize in the remainder of the paper. The work here is not new. We omit most of the routine details, and so the reader unfamiliar with the ideas below might like to compare our reformulation with a standard presentation.

Fix a set AtProp of *atomic propositions*. We use letters like p and q to range over atomic propositions. The infinitary modal language $\mathcal{L}(\mathsf{AtProp})$ is the smallest class such that $\mathsf{AtProp} \subseteq \mathcal{L}(\mathsf{AtProp})$; if $\varphi \in \mathcal{L}$, then $\neg\varphi \in \mathcal{L}(\mathsf{AtProp})$ and also $\Diamond\varphi \in \mathcal{L}(\mathsf{AtProp})$; and if S is any subset of $\mathcal{L}$, then $\bigwedge S \in \mathcal{L}(\mathsf{AtProp})$. Note that the empty set is a subset of $\mathcal{L}(\mathsf{AtProp})$, and we denote $\bigwedge \emptyset$ by *true*.

A *model* is a pair $\mathcal{E} = \langle A, e\rangle$, where A is a set and $e : A \to \mathcal{P}(A) \times \mathcal{P}(\mathsf{AtProp})$. That is, a model associates to every $a \in A$ a set $\pi_1(e_a) \subseteq A$ and another set $\pi_2(e_a) \subseteq \mathsf{AtProp}$. For such a

model, we define the *satisfaction relation* $a \models_{\mathcal{E}} \varphi$ by recursion on $\varphi \in \mathcal{L}(\mathsf{AtProp})$ as follows:

$$\begin{array}{ll} a \models_{\mathcal{E}} p & \text{if } p \in \pi_2(e_a) \\ a \models_{\mathcal{E}} \neg\varphi & \text{if } a \not\models_{\mathcal{E}} \varphi \\ a \models_{\mathcal{E}} \Diamond\varphi & \text{if for some } b \in \pi_1(e_a),\ b \models_{\mathcal{E}} \varphi \\ a \models_{\mathcal{E}} \bigwedge S & \text{if for all } \varphi \in S,\ a \models_{\mathcal{E}} \varphi \end{array}$$

The finitary boolean connective $\wedge$ is a special case of $\bigwedge$. We also introduce $\bigvee$, $\vee$, $\Box$, etc., as abbreviations in the standard way.

Example 2.1 Let $\mathsf{AtProp} = \{p, q\}$. Consider a model $\mathcal{E} = \langle A, e\rangle$, where $A = \{x, y\}$ and $e : A \to \mathsf{F}(A)$ is given by $e_x = \langle\{x, y\}, \{p, q\}\rangle$ and $e_y = \langle\{x\}, \{p\}\rangle$. (This corresponds to a Kripke model based on X whose accessibility relation is given by xRx, xRy, and yRx; and where the atomic formula p is true at both points while q is true at x alone.) Let

$$\psi \quad = \quad \Box \mathit{true} \ \wedge\ \Diamond(p \wedge \neg q) \ \wedge\ p \wedge q.$$

We claim first that $x \models_{\mathcal{E}} \psi$. The semantics of *true* implies that every point in every model satisfies *true*. Since $x \models_{\mathcal{E}} \mathit{true}$ and $x \in \pi_1(e_x)$, $x \models_{\mathcal{E}} \Diamond \mathit{true}$. The rest of the verification is similar.

Now define formulas φ_n^x and φ_n^y by recursion on n, so that $\varphi_0^x = \varphi_0^y = \mathit{true}$, and

$$\begin{array}{lcl} \varphi_{n+1}^x & = & \Box(\varphi_n^x \vee \varphi_n^y) \ \wedge\ \Diamond\varphi_n^x \ \wedge\ \Diamond\varphi_n^y \ \wedge\ p \ \wedge\ q \\ \varphi_{n+1}^y & = & \Box\varphi_n^x \ \wedge\ \Diamond\varphi_n^x \ \wedge\ p \ \wedge\ \neg q \end{array}$$

Then it is not hard to show by induction on n that $x \models_{\mathcal{E}} \varphi_n^x$ and $y \models_{\mathcal{E}} \varphi_n^y$.

Bisimulation A key ingredient in our reformulation of modal logic is the categorical notion of bisimulation of Aczel and Mendler [AM]. To get started, recall that we are considering the functor $\mathsf{F} : \mathsf{F} \to$ given by $\mathsf{F}(A) = \mathcal{P}(A) \times \mathcal{P}(\mathsf{AtProp})$. We regard as a functor on the category of sets in a natural way: Given $f : A \to B$, we define $\mathsf{F}f : \mathsf{F}(A) \to \mathsf{F}(B)$ by

$$\mathsf{F}f(s, t) \quad = \quad \langle\{f(c') : c' \in c\}, t\rangle.$$

It is easy to see that F preserves compositions and identity maps.

When F is viewed as a functor, Kripke models become *coalgebras*. (A coalgebra for a functor F on a category C is just a map $\alpha : c \to Fc$.) Note that the accessibility relation $a\,R\,b$ of a model is recoverable as the relation $b \in \pi_1(e_a)$. In modal logic, one also finds the notion of a p-morphism of models; indeed, this is one of the main types of mappings between models. For us, p-morphisms are the morphisms of coalgebras. That is, let $\mathcal{E} = \langle A, e\rangle$ and $\mathcal{F} = \langle B, f\rangle$ be models. Then a morphism $\alpha : \mathcal{E} \to \mathcal{F}$ is a map $\alpha : A \to B$ so that $\mathsf{F}\alpha \circ e = f \circ \alpha$.

Let $\mathcal{E} = \langle A, e\rangle$ and $\mathcal{F} = \langle B, f\rangle$ be models. A *bisimulation between $\mathcal{E}$ and $\mathcal{F}$* is a relation R on $A \times B$ (that is, a subset of $A \times B$) such that there is some coalgebra $\mathcal{R} = \langle R, h\rangle$ so that the projections $\pi_1 : \mathcal{R} \to \mathcal{E}$ and $\pi_2 : \mathcal{R} \to \mathcal{F}$ are morphisms.

All of the definitions introduced above agree with their standard formulations. We verify this for bisimulation because the idea is important for this paper. The usual concept of a bisimulation between $\mathcal{E}$ and $\mathcal{F}$ is a relation R so that if aRb, then

(1) $\pi_2(e_a) = \pi_2(e_b)$

(2) For all $a' \in \pi_1(e_a)$, there is some $b' \in \pi_1(e_b)$ so that $a'Rb'$.

(3) For all $b' \in \pi_1(e_b)$, there is some $a' \in \pi_1(e_a)$ so that $a'Rb'$.

If R is a bisimulation in the coalgebraic sense, then we check (1) and (2); (3) is similar to (2). Suppose aRb, so that $\langle a, b\rangle \in R$. Then

$$\begin{array}{rcl} \pi_2(e_a) & = & \pi_2((\mathsf{F}\pi_1)h(\langle a, b\rangle)) \\ & = & \pi_2((\mathsf{F}\pi_2)h(\langle a, b\rangle)) \\ & = & \pi_2(e_b) \end{array}$$

For (2), the morphism condition tells us that

$$\pi_1(e_a) \quad = \quad \pi_1((\mathsf{F}\pi_1)h(\langle a, b\rangle)) \quad = \quad \{a' : (\exists b')\langle a', b'\rangle \in R \cap h(\langle a, b\rangle)\}.$$

Now suppose that $a' \in e_a$. So there is some b' such that $a'Rb'$ and $\langle a', b'\rangle \in h(\langle a, b\rangle)$. This last condition implies that $b \in e'_b$, since we calculate $\pi_1(e'_b) = \pi_1((\mathsf{F}\pi_2)h(\langle a, b\rangle))$. This concludes this part of the proof.

On the other hand, suppose R is a bisimulation in the usual sense. We need to define a map $h : R \to \mathsf{F}(R)$ such that the projections π_1 and π_2 are morphisms. We set

$$h(\langle a, b\rangle) \quad = \quad \langle(\pi_1(e_a) \times \pi_1(e_b)) \cap R, \pi_2(e_a)\rangle.$$

Then the usual bisimulation conditions imply that both projections give morphisms.

2.1 The Characterization Result for Modal Logic

The basic result of infinitary modal logic is the following:

Proposition 2.2 *Let $\mathcal{E} = \langle A, e\rangle$ and $\mathcal{F} = \langle B, f\rangle$ be models. For all $a \in A$ and $b \in B$, a and b satisfy the same formulas of infinitary modal logic iff there is a bisimulation $\mathcal{R}$ between $\mathcal{E}$ and $\mathcal{F}$ such that $a\,R\,b$.*

This result may be strengthened in a number of ways. Let $S \subseteq \mathcal{L}(\mathsf{AtProp})$ and let $T \subseteq \mathsf{AtProp}$. We define $\triangle(S, T)$ to be the following formula:

$$\triangle(S, T) \quad = \quad \Box \bigvee S \ \wedge \ \bigwedge \Diamond S \ \wedge \ \bigwedge_{p \in T} p \ \wedge \ \bigwedge_{p \notin T} \neg p.$$

where $\Diamond S = \{\Diamond\varphi : \varphi \in S\}$.

Let $\mathcal{L}(\triangle)$ be the least collection $\mathcal{L}(\triangle)$ such that for every $S \subseteq \mathcal{L}(\triangle)$ and $T \subseteq \mathsf{AtProp}$, $\bigwedge S \in \mathcal{L}(\triangle)$ and $\triangle(S, T) \in \mathcal{L}(\triangle)$. Then the proof of Proposition 2.2 shows that a and b satisfy the same formulas of $\mathcal{L}(\triangle)$ iff there is a bisimulation $\mathcal{R}$ between $\mathcal{E}$ and $\mathcal{F}$ such that $a\,R\,b$. This result itself can be significantly strengthened: for a fixed $\langle\mathcal{E}, a\rangle$, the bisimulation type of $\langle\mathcal{E}, a\rangle$ may be determined by a *single* infinitary formula in $\mathcal{L}(\triangle)$ depending only on $\langle A, a\rangle$: it is not necessary to use the whole proper class of infinitary formulas.

Definition A *model-world pair* is is a model $\mathcal{E} = \langle A, e\rangle$ together with some $a \in A$. A formula

θ of $\mathcal{L}(\mathsf{AtProp})$ *characterizes* $\langle \mathcal{E}, a\rangle$ if for all $\mathcal{F} = \langle B, f\rangle$ and all $b \in B$, $b \models_{\mathcal{F}} \theta$ iff there is a bisimulation $\mathcal{R}$ between $\mathcal{E}$ and $\mathcal{F}$ such relating a to b.

We go into detail on how characterizing formulas are obtained. Let $\langle \mathcal{E}, a\rangle$ be a model-world pair. We define a formula φ^a_α by recursion on α as follows:

$$\begin{array}{lcl} \varphi^a_0 & = & \mathit{true} \\ \varphi^a_{\alpha+1} & = & \triangle(\{\varphi^b_\alpha : b \in \pi_1(e_a)\}, \pi_2(e_a)) \\ \varphi^a_\lambda & = & \bigwedge\{\varphi^a_\beta : \beta < \lambda\} \end{array}$$

Note that these formulas depend on $\mathcal{E}$ but that we have suppressed $\mathcal{E}$ in the notation.

Theorem 2.3 *For all $\langle \mathcal{E}, a\rangle$, there is an ordinal α with the following property: if $\langle \mathcal{F}, b\rangle$ is any model-world pair and $b \models_{\mathcal{F}} \varphi^a_\alpha$, then there is a bisimulation of $\mathcal{E}$ and $\mathcal{F}$ relating a to b.*

Proof We sketch Gerbrandy's proof of this result from [G]. First, consider the submodel $\mathcal{E}_0$ of $\mathcal{E}$ generated by a. This submodel is easily seen to be a set. The approximations φ^a_α have several important properties: if $\alpha \leq \beta$, then $\varphi^a_\beta \to \varphi^a_\alpha$; and for all fixed α, the two formulas φ^α_a and φ^α_b are either equivalent or contradictory. We will not prove these facts here, but versions of them will follow from our more general work in Section 6. So we can fix an ordinal β large enough so that for all $b, c \in A_0$, if $\varphi^\beta_b \neq \varphi^\beta_c$, then for all $\gamma \geq \beta$, we have $\varphi^\gamma_b \neq \varphi^\gamma_c$.

Next, let $\alpha = \beta + \omega$, so that α is the least limit ordinal greater than β. We claim that the following relation is a bisimulation between $\mathcal{E}$ and $\mathcal{F}$: cRd iff $\varphi^\alpha_c = \varphi^\alpha_d$. (It then will follow that if $b \models_{\mathcal{F}} \varphi^\alpha_a$, then R is a bisimulation of $\mathcal{E}$ and $\mathcal{F}$ relating a to b.) Suppose that cRd. Then it is easy to see that c and d satisfy the same atomic formulas. We claim that for every $c' \in \pi_1(e_c)$ there is some $d' \in \pi_1(e'_d)$ such that $c'Rd'$. To see this, fix c'. Note first that since $\varphi^{\beta+2}_c = \varphi^{\beta+2}_d$, there is some d' so that $\varphi^{\beta+1}_{c'} = \varphi^{\beta+1}_{d'}$. We claim that in fact $\varphi^\alpha_{c'} = \varphi^\alpha_{d'}$. For this, fix $n < \omega$. For d' there is some c'' such that $\varphi^{\beta+n}_{c''} = \varphi^{\beta+n}_{d'}$. But then $\varphi^{\beta+1}_{c''} = \varphi^{\beta+1}_{d'} = \varphi^{\beta+1}_{c'}$. Hence $\varphi^\gamma_{c''} = \varphi^\gamma_{c'}$ for all γ. In particular, this holds for $\gamma = \beta + n$. Thus $\varphi^{\beta+n}_{c'} = \varphi^{\beta+n}_{d'}$. Since n is arbitrary, $\varphi^\alpha_{c'} = \varphi^\alpha_{d'}$.

This verifies one direction of the bisimulation conditions, and the other part is similar. $\dashv$

There is a refinement of Theorem 2.3 that gives a bound on the ordinals involved. For each $\mathcal{E}$, let R^* be the transitive closure of the relation R on A. Further, for each model-world pair $\langle \mathcal{E}, a\rangle$, let $\kappa = \kappa(\mathcal{E}, a)$ be the smallest infinite regular cardinal such that for all b such that $a\,R^*\,b$, w has $< \kappa$ many successors up to bisimulation.

Theorem 2.4 *Let $\langle \mathcal{E}, a\rangle$ and $\langle \mathcal{F}, b\rangle$ be model-world pairs. If and $b \models_{\mathcal{F}} \varphi^a_\kappa$, then there is a bisimulation of $\mathcal{E}$ and $\mathcal{F}$ relating a to b.*

Example 2.5 We continue to study $\mathcal{E}$ from Example 2.1. In this case, $\kappa(\mathcal{E}, x) = \kappa(\mathcal{E}, y) = \omega$, the first infinite ordinal. For all natural numbers n, φ^x_n and φ^y_n are written out in Example 2.1. Let $\theta^x = \bigwedge_n \varphi^n_x$. Theorem 2.4 tells us that if we have any model, say $\mathcal{E}' = \langle B, e'\rangle$, and some $b \in B$ such that $b \models_{\mathcal{E}'} \theta^x$, then there is a bisimulation R between $\mathcal{E}$ and $\mathcal{F}$ which relates x to b.

The proof of Theorem 2.4 may be found in Chapter 11 of [BM1]. Applications are discussed there, in [BM2], and in [Bal]. Both Baltag [Bal] and Gerbrandy [G] note that Fagin [F] also

obtains characterization results for semantics of infintary modal logic based on α-worlds. As it happens, those worlds are essentially what we are calling the canonical formulas.

Having introduced $\mathcal{L}(\triangle)$ as a fragment of infinitary modal logic, we would like to pose one open question about it. As shown by van Benthem [Ben], infinitary modal logic is equivalent to the fragment of infinitary first-order logic which contains the formulas invariant under bisimulation of models. Is there a similar result for $\mathcal{L}(\triangle)$, isolating $\mathcal{L}(\triangle)$ as a logic corresponding to invariance for some relation on first-order models?

Towards a generalization Beginning in Section 4, we present a generalization of $\mathcal{L}(\triangle)$. This generalization is actually quite broad, and it covers other logics as well. As a result, the reader might not at first glace see the connection of the work to come with modal logic. So we sketch briefly a development of our ideas more closely connected to modal logic.

The basic idea is to work directly with the syntax of $\mathcal{L}(\triangle)$, dropping the connection to modal logic. In that case, we'd have formulas like true, $\triangle(\{\mathit{true}\},\{p\})$, $\triangle(\{\triangle(\mathit{true},\{p\})\},\{p,q\})$, etc. Note that the modal operators $\Box$ and $\Diamond$ are gone. This is a general feature (or bug) of our approach. Now suppose we have sets $S \subseteq \mathcal{L}(\triangle)$ and $T \subseteq \mathsf{AtProp}$. Let's see what the definition of the semantics of $\triangle(S,T)$ should say. To get started, fix a coalgebra $\mathcal{E} = \langle A, e\rangle$. Let

$$\mathit{Sat} \quad = \quad \{\langle b, \psi\rangle : b \models_{\mathcal{E}} \psi\}.$$

Our first rendering of the semantics is

$$\begin{array}{ll} a \models_{\mathcal{E}} \triangle(S,T) & \text{if there is } w_0 \subseteq \mathit{Sat} \text{ so that } \{\pi_1(x) : x \in w_0\} = e_a, \\ & \{\pi_2(x) : x \in w_0\} = S, \text{ and } T = \pi_2(e_a). \end{array}$$

In other words, subsets of Sat are connected to the semantics. However, we can go further. If we had a set $S \subseteq \mathit{Sat}$, then we naturally get sets $\pi_1[S] \subseteq A$ and $\pi_2[S] \subseteq \mathcal{L}(\triangle)$. Furthermore consider $\mathsf{F}(\mathit{Sat}) = \mathcal{P}(\mathit{Sat}) \times \mathcal{P}(\mathsf{AtProp})$. The importance of this is that $\langle w_0, T\rangle \in \mathsf{F}(\mathit{Sat})$. Indeed, now we can reformulate the semantics to read

$$\begin{array}{ll} a \models_{\mathcal{E}} \triangle(S,T) & \text{if there is } w \in \mathsf{F}(\mathit{Sat}) \text{ so that } (\mathsf{F}\pi_1)w = e_a \\ & \text{and } (\mathsf{F}\pi_2)w = \langle S, T\rangle. \end{array}$$

The maps $\mathsf{F}\pi_1$ and $\mathsf{F}\pi_2$ are the same that we saw above, concerning bisimulation. Finally, we can go one step further. Suppose we reformulate $\mathcal{L}(\triangle)$ itself so as to simply write $\langle S, T\rangle$ instead of $\triangle(S,T)$. Then for every formula φ of $\mathcal{L}(\triangle)$ which is not a conjunction,

$$a \models_{\mathcal{E}} \varphi \quad \text{if there is } w \in \mathsf{F}(\mathit{Sat}) \text{ so that } (\mathsf{F}\pi_1)w = e_a \text{ and } (\mathsf{F}\pi_2)w = \varphi.$$

This is the leading idea behind our logics.

3 Background on Functors on Sets

From our work in Section 2, we have seen the operation $\mathsf{F}(a) = \mathcal{P}(a) \times \mathcal{P}(\mathsf{AtProp})$ come up several times. In fact, we not only applied this to sets but also to functions (when we wrote $\mathsf{F}\pi_1$ and $\mathsf{F}\pi_2$) and even to classes (when we applied it to the satisfaction relation Sat of some model).

So to get a generalization of $\mathcal{L}(\triangle)$, it makes sense to consider the category $\mathtt{SET}$ of classes and set-continuous functions. These are functions $f : C \to D$ between classes with the property that

$$f(C) \quad = \quad \bigcup\{f(c) : c \subseteq C \text{ and } c \text{ is a set}\}. \tag{1}$$

In working with classes, we have the usual choice of either working in set theory and treating classes as a *facon de parler*, or working in a theory that has classes as first-class objects. None of our results hinge on the choice here. For some results, we assume the Global Axiom of Choice. Except for this, we work in standard set theory in this paper. We will not need either the Foundation or Anti-Foundation Axioms, though we discuss the connections of *AFA* to our work at certain points.

Returning to F, we are concerned with functors on `SET`. We will work with functors which are *set-based*, *standard* and which *preserve weak pullbacks.*

Definition F is *set-based* if for all classes C and all $a \in \mathsf{F}(C)$, there is some set C_0 and some $a_0 \in \mathsf{F}(C_0)$ such that $a = (\mathsf{F}i)a_0$, where i is the inclusion of C_0 in C.

F is *standard* if whenever $f : A \hookrightarrow B$ is an inclusion, then also $\mathsf{F}f : \mathsf{F}(A) \hookrightarrow \mathsf{F}(B)$ is also an inclusion.

As a consequence, we see F commutes with restrictions: If $f : A \to B$ and $A_0 \subseteq A$, then $\mathsf{F}(f \restriction A_0) = (\mathsf{F}f) \restriction \mathsf{F}(A_0)$. For another consequence, note that that as an operator on classes, F is *monotone*: if $A \subseteq B$, then $\mathsf{F}(A) \subseteq \mathsf{F}(B)$. Also, given a monotone operator on sets (such as the power set operation), then we can extend it to classes via (1). As a result, we have the following:

Theorem 3.1 *If* F *is an operator on classes which is monotone and set-based, then* F *has a least fixed point* F_* *and a greatest fixed point* F^*.

The assertion about least fixed points is a well-known consequence of definition by recursion on the ordinals. The greatest fixed point result is due to Aczel [A]. Indeed, there is a nice representation for the class F^*:

$$\mathsf{F}^* \quad = \quad \bigcup\{b \mid \ b \text{ is a set and } b \subseteq \mathsf{F}(b)\}.$$

Definition Let $g : C \to D$ and $v : B \to D$ be two morphisms with the same codomain D. A *weak pullback (of* g *and* v*)* is a pair $\langle f : A \to B, u : A \to C\rangle$ so that $v \circ f = g \circ u$ and with the extra property that if $b \in B$ and $c \in C$ are such that $v(b) = g(c)$, then there is some $a \in A$ such that $f(a) = b$, and $u(a) = c$. This a need not be unique. (If there were a unique a for each such b and c, the diagram would be a *pullback.*)

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "u"'] & B \arrow[d, "v"] \\ C \arrow[r, "g"'] & D \end{tikzcd}$$

F *preserves weak pullbacks* if the image under F of every weak pullback square is itself a weak pullback square.

Essentially all of the natural functors on `SET` are set-based, standard, and preserve weak pullbacks. These include the identity, every constant functor, and the power set $\mathcal{P}$. This last

functor sends a map $f : A \to B$ to $\mathcal{P}f : \mathcal{P}A \to \mathcal{P}B$ given by $(\mathcal{P}f)s = f[s]$. Incidentally, the power set functor preserves weak pullbacks, but it does not preserve pullbacks. For every fixed set C, the functors $a \mapsto a + C$, $a \mapsto a \times C$, and $a \mapsto C \to a$ also have all these properties. In writing $a + C$, we have in mind the disjoint sum, defined in general by

$$X + Y \quad = \quad (\{0\} \times X) \cup (\{1\} \times Y), \tag{2}$$

and the cartesian pair operation is based on the Kuratowski pair $\langle x, y\rangle = \{\{x\}, \{x, y\}\}$. Since we are mentioning disjoint sums at this point, we should also descibe some related notation. If A and B are any sets, then we write $\mathsf{inl} : A \to A + B$ for the function $a \mapsto \langle 0, a\rangle$, and $\mathsf{inr} : B \to A + B$ for $b \mapsto \langle 1, b\rangle$. If $f : A \to C$ and $g : B \to C$, then $[f, g] : A + B \to C$ in the obvious way, by cases.

We remark that there are functors on $\mathtt{SET}$ which do not have these properties. For example, the operation $x \mapsto \{x\}$ is not monotone, so as a functor it is not standard. There are examples of functors which do not preserve weak pullbacks: one may be found in Aczel and Mendler [AM], and another (mentioned in Rutten [R96]) is the composition of the contravariant power set functor with itself.

Before we turn to a background section on coalgebras, we note a few consequences of the assumption that F preserve weak pullbacks.

Example 3.2 We need a fact about the composition of relations. Let $R \subseteq A \times B$ and $S \subseteq B \times C$. Then the domain of the pullback of $\pi_2^R : R \to B$ and $\pi_1^S : S \to B$ is *not* the composition $S \circ R$ but rather $S \star R$, where

$$S \star R \quad = \quad \{\langle\langle a, b\rangle, \langle b', c\rangle\rangle \in (A \times B) \times (B \times C) : b = b'\}.$$

The pullback square is $\pi_2^R \circ \pi_1^{S\star R} = \pi_1^S \circ \pi_2^{S\star R}$. It is possible to find maps $k_1 : S \circ R \to A$ and $k_2 : S \circ R \to B$ so that $\pi_2^R \circ k_1 = \pi_1^S \circ k_2$ is a weak pullback. However, k_1 and k_2 are not unique, and we need the Axiom of Choice to get them at all. Note also that there is a natural map $h : S \star R \to S \circ R$ given by

$$h \quad = \quad \langle \pi_1^{A\times B} \circ \pi_1^{S\star R}, \pi_2^{B\times C} \circ \pi_2^{S\star R}\rangle.$$

With this definition, for all $c \in S \star R$,

$$\begin{array}{lcl} (\pi_1^{S\circ R} \circ h)c & = & (\pi_1^{A\times B} \circ \pi_1^{S\star R})c \\ (\pi_2^{S\circ R} \circ h)c & = & (\pi_2^{B\times C} \circ \pi_2^{S\star R})c \end{array} \tag{3}$$

Definition A *top-full square* in the category of sets is a commutative square $v \circ f = g \circ u$ with the extra property that if $b \in B$ and $c \in C$ are such that $v(b) = g(c)$, then there is some $a \in A$ such that $f(a) = b$. Note that we do not require the symmetric condition that there be some $a \in A$ such that $u(a) = c$. F *preserves top-full squares* if the image of every top-full square is again a top-full square.

Proposition 3.3 *If* $\mathsf{F} : \mathtt{SET} \to \mathtt{SET}$ *preserves weak pullbacks, then* F *preserves top-full squares.*

Proof First, an observation. Consider a square $v \circ f = g \circ u$. If there is a weak pullback $v \circ f' = g \circ u'$ and a map k such that $f \circ k = f'$, then our original square is a weak pullback. This observation is easy to check. Based on it and the fact that (weak) pullbacks exist in $\mathtt{SET}$, the overall result follows easily. $\dashv$

Summary All functors F in this paper are endofunctors on $\mathtt{SET}$ and are assumed to be set-based, standard, and to preserve weak pullbacks.

3.1 Coalgebras

A *coalgebra* for an endofunctor $\mathsf{F} : \mathtt{SET} \to \mathtt{SET}$ is a pair $\mathcal{E} = \langle A, e\rangle$ such that A is a set or class, together with a map $e : A \to \mathsf{F}A$. This is the dual concept of an *algebra*: a pair $\mathcal{E} = \langle A, e\rangle$, where $e : \mathsf{F}A \to A$.

Any fixed point of F may be considered as an algebra or coalgebra of F; we use the identity map. For concrete examples of coalgebras, note first that a coalgebra for $a \mapsto \mathcal{P}(a)$ is just a system of set equations $\langle A, e\rangle$ using A as a set of variables. We can also consider a fixed set B and the functor $a \mapsto \mathcal{P}(a) + B$. Then the coalgebras are systems of equations $\langle A, e\rangle$ where the right hand sides might have elements of B, especially tagged as such.

Example 3.4 For a different kind of example, let $\mathsf{F}(b) = 2 \times \mathcal{P}(Act \times b)$. The coalgebras for this functor are essentially non-deterministic automata, together with a tagging of states with either 0 or 1. (The tagging might be used to indicate whether a state is an accepting state, for example.) Indeed, suppose that $Act = \{a, b\}$. Consider the coalgebra $\langle S, \delta\rangle$, where $S = \{s, t, u\}$, and δ is given as follows:

$$\begin{array}{lcl} \delta(s) & = & \langle 0, \{\langle a, t\rangle, \langle b, u\rangle\}\rangle \\ \delta(t) & = & \langle 0, \{\langle a, t\rangle\}\rangle \\ \delta(u) & = & \langle 1, \{\langle a, t\rangle, \langle b, t\rangle, \langle b, u\rangle\}\rangle \end{array}$$

Formally, coalgebras are pairs. But when we have a coalgebra $\mathcal{E}$, we often write $a \in \mathcal{E}$ to mean that $a \in \pi_1(\mathcal{E})$. This will save on some notation. Also, when we apply a coalgebra map e to some $a \in \mathcal{E}$, we often write e_a rather than $e(a)$.

Definition Let $\mathcal{E} = \langle A, e\rangle$ and $\mathcal{E}' = \langle B, e'\rangle$ be coalgebras. An F-*morphism* $f : \mathcal{E} \to \mathcal{E}'$ is a function $f : A \to B$ such that $\mathsf{F}f \circ e = e' \circ f$. $\mathcal{E}$ is a *subcoalgebra* of $\mathcal{E}'$ if $A \subseteq B$ and if the inclusion map $i_{A,B} : A \hookrightarrow B$ is a F-morphism from $\mathcal{E}$ to $\mathcal{E}'$.

Lemma 3.5 (Small Subcoalgebra Lemma, Aczel and Mendler [AM]) *If $\mathcal{E} = \langle A, e\rangle$ is a coalgebra and A_0 is a subset of A, then there is a subcoalgebra $\mathcal{E}' = \langle B, e'\rangle$ of $\mathcal{E}$ such that B is a set and $A_0 \subseteq B$.*

Coalgebras for a functor with the morphisms above constitute a category, and so we get the notion of *final* F-*coalgebra*: this is an F-coalgebra $\mathcal{E}$ such that for all $\mathcal{E}'$ there is a unique F-morphism $f : \mathcal{E}' \to \mathcal{E}$.

Theorem 3.6 (Aczel and Mendler [AM]) *There is a final F-coalgebra $\overline{\mathcal{F}}$.*

This result was proved on the sole assumption on F be set-based. The proof involves taking a quotient of the colimit of all small coalgebras. As a result, by itself it does not give a general method of getting detailed information about a final coalgebra. The greatest fixed point F^* is *usually* a final coalgebra, but this takes some kind of assumption on F. There are a number of

different conditions that one could impose in order to show that F^* is the a final coalgebra; we discuss this matter further in Section 7.

Given two F-coalgebras $\mathcal{E} = \langle A, e\rangle$ and $\mathcal{E}' = \langle A', e'\rangle$, an F-*bisimulation* between $\mathcal{E}$ and $\mathcal{E}'$ is an F-coalgebra $\mathcal{R} = \langle R, f\rangle$ such that $R \subseteq A \times A'$ and such that the projections π_1^R and π_2^R give F-morphisms $\pi_1 : \mathcal{R} \to \mathcal{E}$ and $\pi_2 : \mathcal{R} \to \mathcal{E}'$. This notion generalizes what we have seen of bisimulations in modal logic in Section 2.

There are a number of results in this paper which only hold for functors which preserve weak pullbacks. Here is one such result, a fact due to Aczel and Mendler [AM] (see also [BM1], Theorem 17.11):

Proposition 3.7 *Let F preserve weak pullbacks, and let $\mathcal{E} = \langle A, e\rangle$ be an F-coalgebra with final coalgebra map s. Let t be a function with domain A with the property that*

$$\textit{whenever } t_b = t_c, \quad (\mathsf{F}t)e_b = (\mathsf{F}t)e_c.$$

Then whenever $t_b = t_c$, we also have $s_b = s_c$.

4 From Functors to Coalgebraic Logics

Having introduced our subject from two sides, we can now state the goals of this paper:

1. To generalize the $\mathcal{L}(\triangle)$ from modal logic to a family of logics given in terms of functors F on sets which have the basic properties mentioned in Section 3. The logical language for F should be determined entirely by F, and is should be interpreted on F-coalgebras.

2. The generalizations should have the property that bisimilar objects should satisfy the same formulas.

3. The languages should be strong enough to get a characterization result.

In this section we begin this work. We remind the reader that we assume that our functors are assumed to be set-based, standard, and preserve weak-pullbacks.

First, here is our definition of a *coalgebraic logic* $\mathcal{L}_\mathsf{F}$ corresponding to such a functor F.

Definition We define $\mathcal{L}_\mathsf{F} = (\mathcal{P} + \mathsf{F})_*$, and we call this the F-*language*. That is, $\mathcal{L}_\mathsf{F}$ is the least class such that

$$\mathcal{L}_\mathsf{F} \quad = \quad \mathcal{P}\mathcal{L}_\mathsf{F} + \mathsf{F}\mathcal{L}_\mathsf{F} \tag{4}$$

We therefore have two injections which we write as $\bigwedge : \mathcal{P}\mathcal{L}_\mathsf{F} \to \mathcal{L}_\mathsf{F}$ and $\mathsf{inr} : \mathsf{F}\mathcal{L}_\mathsf{F} \to \mathcal{L}_\mathsf{F}$. We write also *true* for $\bigwedge\emptyset$. Given an F-coalgebra $\mathcal{E} = \langle A, e\rangle$, we define the *satisfaction relation* $Sat(= Sat(\mathsf{F}, \mathcal{E}))$ to be the least fixed point of the following operator $\Phi : \mathcal{P}(A \times \mathcal{L}_\mathsf{F}) \to \mathcal{P}(A \times \mathcal{L}_\mathsf{F})$:

$$\begin{array}{rcl} \Phi(R) & = & \{\langle a, \bigwedge S\rangle : S \subseteq \mathcal{L}_\mathsf{F} \text{ and } \langle a, \varphi\rangle \in R \text{ for all } \varphi \in S\} \\ & \cup & \{\langle a, \varphi\rangle : \text{ for some } w \in \mathsf{F}(R),\ e_a = \mathsf{F}\pi_1^R(w) \ \&\ \varphi = (\mathsf{inr} \circ \mathsf{F}\pi_2^R)w\} \end{array} \tag{5}$$

We shall show below that Φ is monotone. We also write Sat as $\models$, especially when we use it as a binary relation. The fixed point property means that we have the following characterization:

$b \models_{\mathcal{E}} \bigwedge S$ if $b \models_{\mathcal{E}} \varphi$ for all $\varphi \in S$
$b \models_{\mathcal{E}} \varphi$ if there is some $w \in \mathsf{F}(Sat)$ such that $\mathsf{F}\pi_1^{Sat}(w) = e_b$ and $(\mathsf{inr} \circ \mathsf{F}\pi_2^{Sat})w = \varphi$

The second clause applies when $\varphi \in (\mathsf{inr} \circ \mathsf{F})\mathcal{L}_{\mathsf{F}}$. When w is as in that clause, we call w a *reason why* $b \models_{\mathcal{E}} \varphi$.

This concludes the definition of the languages and their semantics. The remainder of this section contains examples, beginning with the syntax, together with a discussion of the semantics.

Example 4.1 Consider $\mathsf{F}(a) = \{p, q\} \times \mathcal{P}a$. The language $\mathcal{L}_{\mathsf{F}}$ contains $\bigwedge \emptyset = \mathit{true}$, as do all of our languages. We also get formulas in $\mathcal{L}_{\mathsf{F}}$ such as the following:

$$\begin{array}{lcl} \psi_1 & = & \mathsf{inr}(\langle q, \bigwedge\{\mathit{true}\}\rangle) \\ \psi_2 & = & \bigwedge\{\psi_1, \mathsf{inr}(\langle p, \bigwedge\{\psi_1, \mathit{true}\}\rangle)\} \end{array}$$

One might well drop the injections, use infix notation for binary conjunctions, and abbreviate these as $\psi_1 = \langle q, \bigwedge\{\mathit{true}\}\rangle$ and $\psi_2 = \psi_1 \wedge \langle p, \psi_1 \wedge \mathit{true}\rangle$.

To understand the semantics of $\mathcal{L}$, note that we have projections from $A \times \mathcal{L}_{\mathsf{F}}$ to A and to $\mathcal{L}_{\mathsf{F}}$. Consider the restrictions $\pi_1 : Sat \to A$ and $\pi_2 : Sat \to \mathcal{L}_{\mathsf{F}}$. Recall that F is the object part of an endofunctor on classes. So we have maps $\mathsf{F}\pi_1 : \mathsf{F}(Sat) \to \mathsf{F}A$ and $\mathsf{F}\pi_2 : \mathsf{F}(Sat) \to \mathsf{F}\mathcal{L}_{\mathsf{F}}$. By using $\mathsf{F}\pi_1$ and $\mathsf{F}\pi_2$, $\mathsf{F}(Sat)$ can be regarded as a relation on $\mathsf{F}(A) \times \mathsf{F}(\mathcal{L}_{\mathsf{F}})$. This is the leading idea for the semantics.

In the definition of the semantics, we asserted that Φ in (5) is a monotone operator. To verify this, suppose that $R \subseteq S$. We need only consider a pair such as $\langle a, \varphi\rangle$ which is in $\Phi(R)$ via the second clause. So for some $w \in \mathsf{F}(R)$, $e_a = (\mathsf{F}\pi_1^R)w$ and $\varphi = (\mathsf{F}\pi_2^R)w$. Now clearly $w \in \mathsf{F}(S)$. The only thing to check is that $e_a = (\mathsf{F}\pi_1^S)w$ and $\varphi = (\mathsf{F}\pi_2^S)w$. And for this, let $i : R \hookrightarrow S$ be the inclusion. Note that $\pi_1^S \circ i = \pi_1^R$ and $\pi_2^S \circ i = \pi_2^R$. Therefore, $(\mathsf{F}\pi_1^S) \circ (\mathsf{F}i) = \mathsf{F}\pi_1^R$. By standardness, $\mathsf{F}i$ is the inclusion $\mathsf{F}R \hookrightarrow \mathsf{F}S$. This implies that $(\mathsf{F}\pi_1^R)w = (\mathsf{F}\pi_1^S)w$. The same is true for π_2.

As this argument shows, the various superscripts on the projections π_i do not really matter. Therefore, we occasionally will drop them.

The last clause in the definition of the semantics is intended for the case of $\varphi \in \mathsf{F}(\mathcal{L})$. It is the heart of the definition. So we present a few examples to make the concept clear.

Example 4.2 We return to Example 2.1 from Section 2. This concerned $\mathsf{F}(a) = \mathcal{P}(a) \times \mathcal{P}(\{p, q\})$ and the F-coalgebra $\mathcal{E} = \langle A, e\rangle$, where $A = \{x, y\}$, $e_x = \langle\{x, y\}, \{p, q\}\rangle$, and $e_y = \langle\{x\}, \{p\}\rangle$. Let $\chi = \mathsf{inr}(\langle\{\mathit{true}\}, \{p, q\}\rangle)$. (This corresponds to the modal formula $\Diamond \mathit{true} \wedge \Box \mathit{true} \wedge p \wedge q$.) We claim first that $x \models \chi$. For a reason w we take

$$w \quad = \quad \langle\{\langle x, \mathit{true}\rangle, \langle y, \mathit{true}\rangle\}, \{p, q\}\rangle.$$

We first have to see that w really does belong to $\mathsf{F}(Sat)$. This is because $\langle x, \mathit{true}\rangle$ and $\langle y, \mathit{true}\rangle$ belong to Sat, and because $\{p, q\} \subseteq \{p, q\}$. Next, we need to calculate $(\mathsf{F}\pi_1)w$ and $(\mathsf{F}\pi_2)w$. We see that $(\mathsf{F}\pi_1)w = \langle\{x, y\}, \{p, q\}\rangle = e_x$, and $(\mathsf{F}\pi_2)w = \langle\{\mathit{true}\}, \{p, q\}\rangle$, so that $(\mathsf{inr} \circ \mathsf{F}\pi_2) = \chi$. This w is a reason that $x \models \chi$. (The formula ψ from Example 2.1 does not correspond to a formula in $\mathcal{L}_{\mathsf{F}}$. It is exactly the formulas of $\mathcal{L}(\triangle)$ that have translations into $\mathcal{L}_{\mathsf{F}}$.)

It might also be useful to note that for this F, we have a formula $\mathsf{inr}(\emptyset)$ in addition to $\bigwedge \emptyset = \mathit{true}$. This semantics of $\mathsf{inr}(\emptyset)$ is: $a \models_{\mathcal{E}} \mathsf{inr}(\emptyset)$ iff $e_a = \emptyset$.

Example 4.3 As in Example 3.4, we consider $\mathsf{F}(b) = 2 \times \mathcal{P}(Act \times b)$. To make the notation simpler, in this example we'll omit the injections inr which ought to be applied to each pair. Consider the following formulas:

$$\begin{array}{lcl} \varphi_1^s & = & \langle 0, \{\langle a, \mathit{true}\rangle, \langle b, \mathit{true}\rangle\}\rangle \\ \varphi_1^t & = & \langle 0, \{\langle a, \mathit{true}\rangle\}\rangle \\ \varphi_1^u & = & \langle 1, \{\langle a, \mathit{true}\rangle, \langle b, \mathit{true}\rangle\}\rangle \end{array}$$

Then we check that $s \models \varphi_s^1$; the same hold for t and u with similar proofs. For s, we let

$$w \quad = \quad \langle 0, \{\langle a, \langle t, \mathit{true}\rangle\rangle, \langle b, \langle u, \mathit{true}\rangle\rangle\}\rangle.$$

We claim that $w \in \mathit{Sat}$. For this, it is enough to see that both $\langle a, \langle t, \mathit{true}\rangle\rangle$ and $\langle b, \langle u, \mathit{true}\rangle\rangle$ belong to $\mathit{Act} \times \mathit{Sat}$. And this is clear, since $t \models_{\mathcal{E}} \mathit{true}$ and $u \models_{\mathcal{E}} \mathit{true}$. Also, it is possible to check that $s \not\models \varphi_1^t$. Further, $t \models \langle 0, \{\langle a, \varphi_1^t\rangle, \langle b, \varphi_1^u\rangle\}\rangle$. We will call this last formula φ_2^t.

Example 4.4 If I is the identity functor, then we obtain a triviality: every element of every coalgebra satisfies every formula of $\mathcal{L}_I$. This shows that to get languages which are minimally expressive, we must impose some condition on F. We do this in Section 7 when we consider *uniform* functors.

Example 4.5 If C is a fixed set and $\mathsf{F}(a) = C \times a$, then $\mathcal{L}_{\mathsf{F}}$ is essentially the smallest set closed under all conjunctions and under prefixing by elements of C. (That is, we can change the notation to read: if $\varphi \in \mathcal{L}_{\mathsf{F}}$, then $c : \varphi \in \mathcal{L}_{\mathsf{F}}$.) The semantics on a coalgebra would be the natural one. We could also work with $\mathsf{F}(a) = 1 + (C \times a)$ to get a logic more closely related those studied in theories of processes. One should view these examples in the light of the first goal stated at the beginning of this section. The languages $\mathcal{L}_{\mathsf{F}}$ are determined from the functor alone, and they correspond to logics of independent interest.

The theory would have been simpler if reasons were required to be functions, but it is possible to get examples of assertions $a \models_{\mathcal{E}} \varphi$ where the reason w cannot be a function. It is also possible to get assertions $a \models_{\mathcal{E}} \varphi$ which have more than one reason.

Example 4.6 Consider $\mathsf{F}(x) = \mathcal{P}(x)$. Let $X = \{x\}$ and $e_x = \{x\}$. We claim first that $x \models \mathsf{inr}(\{\mathit{true}, \mathsf{inr}(\{\mathit{true}\})\})$. To see this, note that $x \models \mathit{true}$, so $\langle x, \mathit{true}\rangle \in \mathit{Sat}$. Then $\{\langle x, \mathit{true}\rangle\}$ is a reason why $x \models \mathsf{inr}(\{\mathit{true}\})$. And then $\{\langle x, \mathsf{inr}(\{\mathit{true}\})\rangle\} \in \mathsf{F}(\mathit{Sat})$. Let $w = \{\langle x, \mathit{true}\rangle, \langle x, \mathsf{inr}(\{\mathit{true}\})\rangle\}$. Then $w \in \mathsf{F}(\mathit{Sat})$, $(\mathsf{F}\pi_1)w = \{x\}$, and $(\mathsf{F}\pi_2)w = \{\mathit{true}, \mathsf{inr}(\{\mathit{true}\})\}$. So w is a reason why $x \models \mathsf{inr}(\{\mathit{true}, \mathsf{inr}(\{\mathit{true}\})\})$. But there cannot be any *function* w with this property, since then w could contain only one pair and $\mathsf{F}\pi_2$ could never be a set with two elements.

For the example of an assertion with more than one reason, consider $\mathcal{E} = \langle B, e\rangle$, where $B = \{x, y, z, w\}$, and e is given by $e_x = \{y, z\}$, $e_y = \{y, w\}$, $e_z = \{w\}$, and $e_w = \emptyset$. Then $\emptyset$ is a reason why $w \models \mathsf{inr}(\emptyset)$. Let $\psi = \mathsf{inr}(\emptyset)$, and let $\chi = \mathsf{inr}(\{\mathit{true}, \psi\})$. So $y \models \chi$, and also $z \models \chi$; the reasons are different, of course. Now $x \models \mathsf{inr}(\{\mathit{true}, \chi\})$. There are two reasons for this:

$$w_1 = \{\langle y, \mathit{true}\rangle, \langle z, \chi\rangle\} \quad \text{and} \quad w_2 = \{\langle y, \chi\rangle, \langle z, \mathit{true}\rangle\}.$$

Remark The language $\mathcal{L}_{\mathsf{F}}$ can be extended in two natural ways. First, one can add negation

and use the classical semantics. We'll call this extended language $\mathcal{L}_{\mathsf{F}}(\neg)$. Infinitary disjunction follows, of course, but it would be possible to add disjunction without negation. Second, one can take the greatest class $\mathcal{L}$ satisfying $\mathcal{L} = \mathcal{P}(\mathcal{L}) + \mathsf{F}(\mathcal{L})$ instead of the least. For example, if $\mathsf{F}(a) = Act \times a$ for a fixed set Act, then the greatest fixed point language would have formulas that contain themselves, such as $\psi = \langle b, \psi \rangle$. We'll write this greatest fixed point as as $\mathcal{L}^*_{\mathsf{F}}$. To define the semantics of $\mathcal{L}^*_{\mathsf{F}}$ we would take Sat to be the greatest fixed point of the operator Φ from (5).

With these notions, we can say in what way our logic generalizes infinitary modal logic.

Proposition 4.7 *Let $\mathsf{F}(a) = \mathcal{P}(a) \times \mathcal{P}(\mathsf{AtProp})$. $\mathcal{L}_{\mathsf{F}}$ corresponds to $\mathcal{L}(\triangle)$ in the sense that every formula of one language is semantically equivalent to some formula of the other, and vice-versa. In the same way, $\mathcal{L}_{\mathsf{F}}(\neg)$ corresponds to $\mathcal{L}(\mathsf{AtProp})$. The languages $\mathcal{L}_{\mathsf{F}}$ and $\mathcal{L}_{\mathsf{F}}(\neg)$ are not equivalent, since if $p \neq q$, the formula $p \vee q$ is not expressible in $\mathcal{L}_{\mathsf{F}}$.*

Proof Consider the following map $\varphi \mapsto \varphi^t$ of $\mathcal{L}(\triangle)$ to $\mathcal{L}_{\mathsf{F}}$:

$$\begin{array}{lcl} (\bigwedge S)^t & = & \bigwedge S^t \\ (\triangle(S,T))^t & = & \mathsf{inr}(\langle S^t, T \rangle) \end{array}$$

In these $S^t = \{\varphi^t : \varphi \in S\}$. Suppose $\langle A, e \rangle$ is a Kripke model (i.e., an F-coalgebra). We show by induction on φ that $a \models \varphi$ iff $a \models \varphi^t$. The proof is by induction on φ. The case for infinitary conjunction is trivial, and assume the result for all $\psi \in S$. If $a \models \triangle(S,T)$, then we consider

$$w \quad = \quad \langle \{\langle b, \psi^t \rangle : b \in \pi_1(e_a), \psi \in S, \text{ and } b \models \psi\}, T \rangle.$$

Then $w \in \mathsf{F}(Sat)$, and the semantics of $\triangle$ implies that w is a reason why $a \models (\triangle(S,T))^t$.

Going the other way, suppose that $a \models (\triangle(S,T))^t$ via the reason w. Then the fact that $(\mathsf{F}\pi_2)w = \mathsf{inr}(\langle S^t, T \rangle)$ implies easily that $a \models p$ for $p \in T$, and $a \models \neg p$ for $p \notin T$. Suppose that $b \in \pi_1(e_a)$. Then since $e_a = \pi_1((\mathsf{F}\pi_1)w)$, there is some pair $\langle b, \psi^t \rangle$ in $\pi_1(w)$. But this means that $b \models \psi$. And for all $\psi \in S$ there is some such $\langle b, \psi^t \rangle \in \pi_1(w)$. So for some b, $b \models \psi$. Thus $a \models \Diamond\psi$. It follows from all of this that $a \models \triangle(S,T)$.

This proves that every formula of $\mathcal{L}(\triangle)$ is equivalent to a formula of $\mathcal{L}(\mathsf{F})$. The translation above is a bijection, so we have the converse assertion also.

Now we turn to the full languages. Obviously, everything formula of $\mathcal{L}_{\mathsf{F}}(\neg)$ can be expressed in $\mathcal{L}(\mathsf{AtProp})$. In the language $\mathcal{L}_{\mathsf{F}}(\neg)$, one can express infinitary disjunction. Hence we can translate the atomic formula p via

$$p^t \quad = \quad \bigvee \{\langle S, T \rangle : S \subseteq \{\mathit{true}\} \ \& \ p \in T \subseteq \mathsf{AtProp}\}.$$

This extends to a translation of all of $\mathcal{L}_{\mathsf{F}}(\neg)$, via

$$(\Diamond\varphi)^t \quad = \quad \bigvee \{\langle \{\varphi^t, \mathit{true}\}, T \rangle : T \subseteq \mathsf{AtProp}\}.$$

Up to equivalence, $\mathcal{L}_{\mathsf{F}}(\neg)$ contains the atomic formulas and is closed under all $\neg$, $\Diamond$, and $\bigwedge$. Therefore, $\mathcal{L}_{\mathsf{F}}(\neg)$ and $\mathcal{L}(\mathsf{AtProp})$ are of equal expressive power.

Finally, we argue that if $p \neq q$, then $p \vee q$ is not expressible in $\mathcal{L}_{\mathsf{F}}$. Actually, something even stronger is true. Let $\mathcal{E} = \langle A, e \rangle$, where $A = \{a, b\}$, $e_a = \langle \{a\}, \{p\} \rangle$ and $e_b = \langle \{b\}, \{q\} \rangle$. We claim that the only formula of $\mathcal{L}_{\mathsf{F}}$ satisfied by both a and b is *true*. (So *true* is the only formula of $\mathcal{L}_{\mathsf{F}}$ which is semantically implied by $p \vee q$.) We argue by induction on $\mathcal{L}_{\mathsf{F}}$. Suppose

a and b both satisfy a formual $\varphi = \mathsf{inr}(\langle S, T\rangle)$, where $S \subseteq \mathcal{L}_{\mathsf{F}}$ and $T \subseteq \mathsf{AtProp}$. Let w_a and w_b be reasons. Thus $(\mathsf{F}\pi_2)w_a = \langle S, T\rangle = (\mathsf{F}\pi_2)w_b$. This means that w_a and w_b are pairs whose second component is T. But the definition of $\mathsf{F}\pi_1$ tells us that this second component is also the second component of both $(\mathsf{F}\pi_1)w_a$ and $(\mathsf{F}\pi_1)w_a$. Therefore, $\{p\} = T = \{q\}$. This contradicts $p \neq q$. Now suppose we have a conjunction $\bigwedge S$ satisfied by both a and b. If $S \neq \emptyset$, let $\varphi \in S$. Then a and b both satisfy φ. This contradicts our induction hypothesis. $\dashv$

5 Basic Properties of the Semantics

In this section we mention some of the fundamental properties of the semantics. The first properties have to do with the question of where reasons for various semantic assertions are to be found. Although Lemmas 5.1 and 5.2 appear to be technical, they are key results in our development. The point is that we do not have access to *elements* of sets in the usual sense, so we need results which allow us to manipulate *reasons* instead.

Lemma 5.1 *Let $\mathcal{E} = \langle A, e\rangle$ be an F-coalgebra, let $f : A \to \mathcal{L}_{\mathsf{F}}$, and let $e \in \mathsf{F}(A)$, so that $(\mathsf{inr} \circ \mathsf{F}f)e$ is a formula of $\mathcal{L}_{\mathsf{F}}$. Suppose that $a \models_{\mathcal{E}} (\mathsf{inr} \circ \mathsf{F}f)e$. Then there is reason w for this assertion which belongs to $\mathsf{F}(X)$, where*

$$X \quad = \quad \{\langle b, f(b)\rangle \in A \times \mathcal{L}_{\mathsf{F}} : b \models_{\mathcal{E}} f(b)\}.$$

Proof We have a commutative diagram

$$\begin{tikzcd} X \arrow[r, "\pi_1"] \arrow[d, "i"'] & A \arrow[d, "f"] \\ Sat \arrow[r, "\pi_2"'] & \mathcal{L}_{\mathsf{F}} \end{tikzcd}$$

The left side of the square is the inclusion of the set X in Sat. Here is why this is a top-full square: Suppose $f(b) = \varphi = \pi_2(\langle b', \varphi\rangle)$. Then $\langle b, \varphi\rangle \in X$, and $\pi_1(\langle b, \varphi\rangle) = b$. Now when we apply F to the square, then again we have a top-full square. The hypothesis that $a \models_{\mathcal{E}} (\mathsf{inr} \circ \mathsf{F}f)e$ tells us that there is some $w_0 \in \mathsf{F}(Sat)$ so that $(\mathsf{F}\pi_1)w_0 = e_a$ and $(\mathsf{F}\pi_2)w_0 = (\mathsf{F}f)e$. We can thus apply the top-full square condition with $e \in \mathsf{F}(A)$ and $w_0 \in \mathsf{F}(Sat)$. So there is some $w \in \mathsf{F}(X) \subseteq \mathsf{F}(Sat)$ such that $\mathsf{F}\pi_1(w) = e_a$. Since the square commutes, $(\mathsf{F}\pi_2 \circ \mathsf{F}i)w = (\mathsf{F}f)e$. By standardness, $\mathsf{F}i$ is an inclusion, so $(\mathsf{F}\pi_2)w = (\mathsf{F}f)e$. This verifies that $w \in \mathsf{F}(f)$ is a reason why $a \models_{\mathcal{E}} (\mathsf{inr} \circ \mathsf{F}f)e$. $\dashv$

Lemma 5.2 *Let $\mathcal{E} = \langle A, e\rangle$ be an F-coalgebra, let $S \subseteq \mathcal{L}_F$, let $x \in \mathsf{F}(S)$ so that $\mathsf{inr}(x) \in \mathcal{L}_{\mathsf{F}}$, and suppose that $a \models_{\mathcal{E}} \mathsf{inr}(x)$. Then there is a reason w for this assertion which belongs to $F(Sat(S))$, where*

$$Sat(S) \quad = \quad \{\langle b, \varphi\rangle \in A \times S : b \models_{\mathcal{E}} \varphi\}.$$

Proof This time we consider

$$\begin{array}{ccc} Sat(S) & \xrightarrow{\pi_2} & S \\ {\scriptstyle i}\downarrow & & \downarrow{\scriptstyle \mathsf{id}_S} \\ Sat & \xrightarrow[\pi_2]{} & S \end{array}$$

The map i on the left side of the square is the inclusion of $Sat(S)$ in Sat. This square is a pullback. Its image under F is therefore a weak pullback. By hypothesis on a, there is some $w_0 \in \mathsf{F}(Sat)$ such that $(\mathsf{F}\pi_1)w_0 = e_a$ and $(\mathsf{F}\pi_2)w_0 = x$. Recall that $x \in \mathsf{F}(S)$. So by the weak pullback condition, there is some $w \in \mathsf{F}(Sat(S))$ such that $(\mathsf{F}\pi_2)w = x$ and $(\mathsf{F}i)w = w_0$. But $\mathsf{F}i$ is the inclusion map, so $w = w_0$. And then $(\mathsf{F}\pi_1)w = e_a$. This shows that w is what we want. $\dashv$

5.1 Morphisms preserve semantics

The next property of the semantics is that it behaves nicely with respect to coalgebra morphisms.

Lemma 5.3 *Suppose that $\mathcal{E}$ and $\mathcal{E}'$ are F-coalgebras, and $f : \mathcal{E} \to \mathcal{E}'$ is an F-morphism. Then for all $\varphi \in \mathcal{L}_{\mathsf{F}}$ and all $a \in \mathcal{E}$, $a \models_{\mathcal{E}} \varphi$ iff $f(a) \models_{\mathcal{E}'} \varphi$.*

Proof By induction on $\mathcal{L}$. That is, let S be the set of those formulas φ for which the lemma holds. Clearly S is closed under $\bigwedge$; we show that S is closed under F in the sense that $\mathsf{inr}[\mathsf{F}(S)] \subseteq S$. It then follows from the least fixed point characterization of $\mathcal{L}_{\mathsf{F}}$ that $\mathcal{L}_{\mathsf{F}} \subseteq S$, and this proves our result.

Suppose that $\varphi \in \mathsf{inr}[\mathsf{F}(S)]$. Let $Sat_{\mathcal{E}}(S)$ be the restriction of $Sat_{\mathcal{E}}$ to $A \times S$, where A is the domain of $\mathcal{E}$. Let $Sat_{\mathcal{E}'}(S)$ be defined similarly. We write π_1' and π_2' for the natural projections. The induction hypothesis tells us that we have a map $g : Sat_{\mathcal{E}}(S) \to Sat_{\mathcal{E}'}(S)$ such that $f \circ \pi_1 = \pi_1' \circ g$, and $\pi_2 = \pi_2' \circ g$. In fact, we have a (weak) pullback:

$$\begin{array}{ccc} Sat_{\mathcal{E}}(S) & \xrightarrow{g} & Sat_{\mathcal{E}'}(S) \\ {\scriptstyle \pi_1}\downarrow & & \downarrow{\scriptstyle \pi_1'} \\ A & \xrightarrow[f]{} & A' \end{array}$$

When we apply F, we therefore have a weak pullback.

Suppose first that $a \models_{\mathcal{E}} \varphi$. By Lemma 5.2, we have a reason w for this assertion which belongs to $\mathsf{F}(Sat_{\mathcal{E}}(S))$. This means that $(\mathsf{F}\pi_1)w = e_b$ and $(\mathsf{F}\pi_2)w = \varphi$. Let $w' = (\mathsf{F}g)w$. Then

$$(\mathsf{F}\pi_1')w' \quad = \quad (\mathsf{F}(\pi_1' \circ g))w \quad = \quad (\mathsf{F}f \circ \mathsf{F}\pi_1)w \quad = \quad e'_{f(b)}.$$

(Here e' is the coalgebra map of $\mathcal{E}'$.) And $(\mathsf{F}\pi_2')w' = (\mathsf{F}\pi_2)w = \varphi$. So w' shows that $f(b) \models_{\mathcal{E}'} \varphi$.

Going the other way, assume that $f(a) \models_{\mathcal{E}'} \varphi$. By Lemma 5.2 again, we have some $w' \in \mathsf{F}(Sat_{\mathcal{E}'}(S))$ such that $(\mathsf{F}\pi_1')w' = e'_{f(a)} = \mathsf{F}f(e_a)$ and $(\mathsf{F}\pi_2')w' = \varphi$. We apply the preservation of weak pullbacks, with w' and e_a. Thus there is some $w \in \mathsf{F}(Sat_{\mathcal{E}}(S))$ such that $(\mathsf{F}g)w = w'$ and $\mathsf{F}\pi_1(w) = e_a$. Then

$$(\mathsf{F}\pi_2)w \quad = \quad (\mathsf{F}(\pi_2' \circ g))w \quad = \quad (\mathsf{F}\pi_2')w' \quad = \quad \varphi.$$

So we see that w is a reason why $b \models_{\mathcal{E}} \varphi$.

Having shown both of these facts for an arbitrary $a \in A$, we conclude that $\varphi \in S$. $\dashv$

As a corollary, we see that our syntax and semantics satisfies one of our overal goals: If R is an F-bisimulation between F-coalgebras $\mathcal{E}$ and $\mathcal{E}'$, and aRa', then a and a' satisfy all of the same formulas of $\mathcal{L}_{\mathsf{F}}$. Also, if $\mathcal{E}$ is a subcoalgebra of $\mathcal{E}'$, then the satisfaction relation for $\mathcal{E}$ is the restriction of the satisfaction relation for $\mathcal{E}'$.

5.2 The Monotonicity Lemma

Let $\leq$ be the natural semantic preorder on $\mathcal{L}_{\mathsf{F}}$: $\varphi \leq \psi$ iff for all coalgebras $\mathcal{E}$ and all $a \in \mathcal{E}$ such that $a \models_{\mathcal{E}} \psi$, we also have $a \models_{\mathcal{E}} \varphi$. We also write this relation as $\psi \models \varphi$. In the sequel, we shall be concerned with properties of this order. Here is one of the most basic of those properties:

Lemma 5.4 (The Monotonicity Lemma) *Let C be any set or class. Suppose that $f : C \to \mathcal{L}_{\mathsf{F}}$ and $g : C \to \mathcal{L}_{\mathsf{F}}$, and assume that $f \leq g$ in the pointwise order on $C \to \mathcal{L}_{\mathsf{F}}$. Then as elements of $\mathsf{F}(C) \to \mathcal{L}_{\mathsf{F}}$, $\mathsf{inr} \circ (\mathsf{F}f) \leq \mathsf{inr} \circ (\mathsf{F}g)$.*

Proof Let $e \in \mathsf{F}(X)$ be arbitrary. Let $\mathcal{E} = \langle A, e\rangle$ be an F-coalgebra, and assume that $a \models_{\mathcal{E}} (\mathsf{inr} \circ \mathsf{F}g)e$. Let w be a reason for this. Consider the following sets

$$X \;=\; \{\langle b, g(c)\rangle : b \models_{\mathcal{E}} g(c)\} \quad \text{and} \quad Y \;=\; \{\langle b, f(c)\rangle : b \models_{\mathcal{E}} f(c)\}$$

Both of these sets are subsets of $Sat_{\mathcal{E}}$. By Lemma 5.1, we may assume that our reason w belongs to $\mathsf{F}(X)$. Let $\leq$ be a well-order of C. Define $k : \mathcal{L}_{\mathsf{F}} \to \mathcal{L}_{\mathsf{F}}$ by

$$k(\varphi) \quad = \quad \begin{cases} f(c) & \text{if } c \text{ is least so that } g(c) = \varphi \\ \mathit{true} & \text{if } \varphi \notin g[\mathcal{L}_{\mathsf{F}}] \end{cases}$$

Then $k \circ g = f$. Let $\overline{k} : X \to Y$ be given by $\overline{k}(\langle b, \varphi\rangle) = \langle b, k(\varphi)\rangle$, so that $\pi_1^X = \pi_1^Y \circ \overline{k}$ and $\pi_2^Y \circ \overline{k} = k \circ \pi_2^X$. Let $\overline{w} = (\mathsf{F}\overline{k})w$. Our claim is that $\overline{w}$ is a reason showing that $b \models_{\mathcal{E}} (\mathsf{inr} \circ \mathsf{F}f)e$. For this, note first that $(\mathsf{F}\pi_1^{Sat})w = e_a$, and $(\mathsf{F}\pi_2^{Sat})w = (\mathsf{F}g)e$. Then we see that

$$\begin{array}{lllllll} (\mathsf{F}\pi_1^{Sat})\overline{w} & = & (\mathsf{F}\pi_1^Y)\overline{w} & = & (\mathsf{F}\pi_1^X)w & = & e_a \\ (\mathsf{F}\pi_2^{Sat})\overline{w} & = & \mathsf{F}(\pi_2^Y \circ \overline{k})w & = & \mathsf{F}k((\mathsf{F}g)e) & = & (\mathsf{F}f)e. \end{array}$$

This completes the verification. $\dashv$

6 Canonical Formulas

This section generalizes the *canonical formulas* φ^b_α from our work on modal logic. Our overall goal is to prove that if b and c belong to the universe of some coalgebra $\mathcal{E}$ and satisfy the same canonical formulas, then $s_b = s_c$. It follows that if $\overline{\mathcal{F}}$ is the final coalgebra and $b \neq c$, then b and c must differ on some formula of $\mathcal{L}_{\mathsf{F}}$. This is an important result because it says that our logic is at least reasonably expressive: elements of $\overline{\mathcal{F}}$ which should be different in fact do differ on some formula.

Definition Let $\mathcal{E} = \langle A, e\rangle$ be an F-coalgebra. We define maps $j_\alpha : A \to \mathcal{L}_{\mathsf{F}}$ by recursion on α.

$$\begin{array}{lcl} j_0 & = & \mathit{true} \\ j_{\alpha+1} & = & \mathsf{inr} \circ (\mathsf{F} j_\alpha) \circ e \\ j_\lambda & = & \bigwedge\{j_\beta \mid \beta < \lambda\} \end{array}$$

In the first clause, we mean the constant function *true* on A. In the second clause, recall that $\mathsf{inr} : \mathsf{F}\mathcal{L}_{\mathsf{F}} \to \mathcal{L}_{\mathsf{F}}$. In the third clause, we have taken the map $\bigwedge : \mathcal{P}\mathcal{L}_{\mathsf{F}} \to \mathcal{L}_{\mathsf{F}}$ and raised its type to: $\bigwedge : \mathcal{P}(A \to \mathcal{L}_{\mathsf{F}}) \to (A \to \mathcal{L}_{\mathsf{F}})$, defined pointwise. A trivial induction shows that indeed $j_\alpha : A \to \mathcal{L}_{\mathsf{F}}$. We define φ^b_α to be the formula $j_\alpha(b)$.

Example 6.1 First, recall the functor $\mathsf{F}(a) = \mathcal{P}(a) \times \mathcal{P}(\mathsf{AtProp})$. Recall also that F-coalgebras are essentially Kripke models over AtProp and that we have a translation $\varphi \mapsto \varphi^t$ of the $\triangle$-fragment $\mathcal{L}(\triangle)$ of infinitary modal logic into $\mathcal{L}_{\mathsf{F}}$. In Section 2.1 we discussed the canonical formulas φ^a_α for modal logic. An easy induction shows that $(\varphi^a_\alpha)^t = j_\alpha(a)$ for all $a \in A$.

Example 6.2 Example 4.3 contains φ^s_1, φ^t_1, and φ^u_1, and φ^t_2 for Example 3.4.

In the next few results, we fix a functor F and an F-coalgebra $\mathcal{E} = \langle A, e\rangle$. We simplify our notation a bit by dropping $\mathcal{E}$ from the satisfaction relation and *Sat* from the projections.

Proposition 6.3 *For all $b \in A$ and all α, $b \models \varphi^b_\alpha$.*

Proof By induction on α. The case $\alpha = 0$ is trivial, as is the case for the limit ordinals. We check the successor case. Consider the function j_α. We can turn j_α into a map $f : A \to \mathit{Sat}$ defined by $f(b) = \langle b, j_\alpha b\rangle$. Then we also have $\mathsf{id}_A = \pi^{\mathit{Sat}}_1 \circ f$, and $j_\alpha = \pi^{\mathit{Sat}}_2 \circ f$. Let $t : A \to \mathsf{F}(\mathit{Sat})$ be $(\mathsf{F}f) \circ e$. Then $(\mathsf{F}\pi^{\mathit{Sat}}_1) \circ t = e$, and

$$(\mathsf{F}\pi^{\mathit{Sat}}_2) \circ t \quad = \quad (\mathsf{F}\pi^{\mathit{Sat}}_2 \circ \mathsf{F}f) \circ e \quad = \quad (\mathsf{F}j_\alpha) \circ e.$$

It follows that for all $b \in A$, t_b is a reason why $b \models \varphi^b_{\alpha+1}$. ⊣

Lemma 6.4 *Suppose that $b \models \varphi^c_\alpha$. Then $\varphi^b_\alpha = \varphi^c_\alpha$.*

Proof By induction on α. The case of 0 is trivial, since φ^b_0 is always *true*. The limit step is almost as easy. Assume the lemma true for α. Consider j_α as a set of ordered pairs.

Suppose that $b \models \varphi^c_{\alpha+1} = (\mathsf{inr} \circ \mathsf{F}j_\alpha)e_c$. There is a reason w for this. By Proposition 6.3 we also have some reason v for $b \models (\mathsf{inr} \circ \mathsf{F}j_\alpha)e_b$. And these have the following properties

$$\begin{array}{lcl@{\qquad}lcl} (\mathsf{F}\pi^{\mathit{Sat}}_1)v & = & e_b & (\mathsf{F}\pi^{\mathit{Sat}}_2)v & = & (\mathsf{F}j_\alpha)e_b \\ (\mathsf{F}\pi^{\mathit{Sat}}_1)w & = & e_b & (\mathsf{F}\pi^{\mathit{Sat}}_2)w & = & (\mathsf{F}j_\alpha)e_c \end{array}$$

By Lemma 5.1 and Proposition 6.3, we may assume that $w \in \mathsf{F}(j_\alpha)$. The same is true for v. Since j_α is a function, $\pi_2^{j_\alpha} = j_\alpha \circ \pi_1^{j_\alpha}$. By functoriality, $\mathsf{F}\pi_2^{j_\alpha} = (\mathsf{F}j_\alpha) \circ (\mathsf{F}\pi_1^{j_\alpha})$. It follows that

$$(\mathsf{F}\pi_2^{j_\alpha})w \quad = \quad (\mathsf{F}j_\alpha)(\mathsf{F}\pi_1^{j_\alpha})w \quad = \quad (\mathsf{F}j_\alpha)(\mathsf{F}\pi_1^{j_\alpha})v \quad = \quad (\mathsf{F}\pi_2^{j_\alpha})v.$$

This tells us that $\varphi^b_{\alpha+1} = (\mathsf{inr} \circ \mathsf{F}j_\alpha)e_b = (\mathsf{inr} \circ \mathsf{F}j_\alpha)e_c = \varphi^c_{\alpha+1}$. $\dashv$

In the lemma below and in the sequel, we consider the maps j_α as members of the function space $A \to \mathcal{L}_\mathsf{F}$, ordered pointwise.

Lemma 6.5 *If $\alpha \leq \beta$, then $j_\alpha \leq j_\beta$.*

Proof By induction on α. We need only consider the case of a successor ordinal $\alpha + 1$. For this, we use induction on β. For the successor step of $\beta + 1$, note that the induction hypothesis implies that $j_\alpha \leq j_\beta$, and so by the Monotonicity Lemma, $\mathsf{inr} \circ \mathsf{F}j_\alpha \leq \mathsf{inr} \circ \mathsf{F}j_\beta$ as elements of $\mathsf{F}(A) \to \mathcal{L}_\mathsf{F}$. Thus $j_{\alpha+1} = \mathsf{inr} \circ \mathsf{F}j_\alpha \circ e = \mathsf{inr} \circ \mathsf{F}j_\beta \circ e = j_{\beta+1}$. $\dashv$

At this point we need to recall Proposition 3.7 concerning functors which preserve weak pullbacks. Also, consider the relation $\equiv_\mathsf{F}$ defined on A by

$$b \equiv_\mathsf{F} c \quad \text{iff} \quad \text{for all } \varphi \in \mathcal{L}_\mathsf{F},\ b \models \varphi \ \text{ iff } \ c \models \varphi$$

Theorem 6.6 *Let $\mathcal{E}$ be an F-coalgebra, let $b, c \in \mathcal{E}$, and let $s : \mathcal{E} \to \overline{\mathcal{F}}$ be the final coalgebra map. If $b \equiv_\mathsf{F} c$, then $s_b = s_c$.*

Proof By Lemma 3.5, there is a coalgebra $\mathcal{E}_0 = \langle A_0, e_0 \rangle$ with The fact that $\mathcal{E}_0$ is a subcoalgebra of $\mathcal{E}$ means that the canonical formulas for elements of A_0 are the same whether we take them to be relative to $\mathcal{E}$ or to $\mathcal{E}_0$. And the same is true for the coalgebra maps: the unique $s_0 : \mathcal{E}_0 \to \overline{\mathcal{F}}$ is the restriction of s to A_0. In this way, we may assume that A is a set.

Let α be so large so that for all $b, c \in A$, $j_\alpha(b) = j_\alpha(c)$ implies that for all $\beta > \alpha$, $j_\beta(b) = j_\beta(c)$. Such an ordinal exists using Lemma 6.5 and Replacement, since we assume that A is a set. If $j_\alpha(b) = j_\alpha(c)$, then

$$(\mathsf{inr} \circ \mathsf{F}j_\alpha)(e_b) \quad = \quad j_{\alpha+1}(b) \quad = \quad j_{\alpha+1}(c) \quad = \quad (\mathsf{inr} \circ \mathsf{F}j_\alpha)(e_c).$$

Note that inr is one-to-one, so that $(\mathsf{F}j_\alpha)e_b = (\mathsf{F}j_\alpha)e_c$. We apply this fact with Proposition 3.7 (taking $t = j_\alpha$). Thus, if $j_\alpha(b) = j_\alpha(c)$, then $s_b = s_c$.

If $b \equiv_\mathsf{F} c$, then by Proposition 6.3 and Lemma 6.4, $j_\alpha(b) = j_\alpha(c)$. Putting things together: if $b \equiv_\mathsf{F} c$, then $s_b = s_c$. $\dashv$

Corollary 6.7 *If $b, c \in \overline{\mathcal{F}}$ and $b \equiv_\mathsf{F} c$, then $b = c$.*

Maps $\theta_\mathcal{E}$ Let $\mathcal{E} = \langle A, e \rangle$ be an F-coalgebra, and let $s : \mathcal{E} \to \overline{\mathcal{F}}$ be the final coalgebra map. Consider the map $\theta = \theta_\mathcal{E} : A \to \mathcal{L}$ given by

$$\theta^a \quad = \quad j_\alpha(a),$$

where α is least so that $j_\alpha(a)$ characterizes s_a in $\overline{\mathcal{F}}$. The main property is that for all F-coalgebras $\mathcal{F}$ and all $b \in \mathcal{F}$, if $b \models \theta^a$, then there is a bisimulation of $\mathcal{E}$ and $\mathcal{F}$ which relates a to b.

6.1 A Characterization Result for F_*

As we have seen, two different elements $b \neq c$ of $\overline{\mathcal{F}}$ must differ on some formula in $\mathcal{L}_{\mathsf{F}}$. This is weaker than a characterization result, because a characterization result would give, for each $b \in \overline{\mathcal{F}}$ a single formula φ_b of which b is the only satisfier. We are interested in getting such a characterization result for $\mathcal{L}_{\mathsf{F}}$, and we do this in Section 7. Our work in that section will bring in another condition on functors and will be different from what we have seen. In this section, we present a weaker result: every formula in the least fixed point F_* has a characterizing formula.

We shall need a principle of definition by recursion on least fixed points. In our context, least fixed points are initial algebras, and so we get the following principle:

Recursion Principle Let $\langle A, g\rangle$ be an F-algebra; that is, a class A and a map $g^{\#} : \mathsf{F}(A) \to A$. Then there is a unique map $g^{\#} : \mathsf{F}_* \to A$ so that $g^{\#} = g \circ \mathsf{F}g^{\#}$.

In this form, this result is due to Aczel [A]. It states that $(\mathsf{F}_*, \mathsf{id})$ is initial in the category of F-algebras. The assumptions needed on F for it are standardness and set-based-ness. The map $g^{\#}$ is defined by a recursion that parallels the usual definition of F_* as a union of sets F_α for ordinals α. Indeed, we set g_0 to be the empty map, $g_{\alpha+1} = g \circ g_\alpha$, and for limit λ, $g_\lambda = \bigcup_{\beta<\lambda} g_\beta$. Then we set $g^{\#} = \bigcup_\alpha g_\alpha$.

Now we apply this recursion principle. Consider $\mathsf{inr} : \mathsf{F}\mathcal{L}_{\mathsf{F}} \to \mathcal{L}_{\mathsf{F}}$ as an algebra. We therefore get $\mathsf{inr}^{\#} : \mathsf{F}_* \to \mathcal{L}_{\mathsf{F}}$ so that $\mathsf{inr}^{\#} = \mathsf{inr} \circ \mathsf{F}\mathsf{inr}^{\#}$.

Proposition 6.8 *For all $a \in \mathsf{F}_*$ and all $b \in \overline{\mathcal{F}}$, if $b \models_{\overline{\mathcal{F}}} \mathsf{inr}^{\#}a$, then $b = a$.*

Proof By induction on F_*. Let C be the set of $a \in \overline{\mathcal{F}}$ for which the result holds. We show $\mathsf{F}(C) \subseteq C$ so that $\mathsf{F}_* \subseteq C$. Let $a \in \mathsf{F}(C)$. Suppose that $b \models_{\overline{\mathcal{F}}} \mathsf{inr}^{\#}a = (\mathsf{inr} \circ \mathsf{F}\mathsf{inr}^{\#})a$. We claim that there is a reason w for $b \models \mathsf{inr}^{\#}a$ in D, where $D = \{\langle c, \mathsf{inr}^{\#}c\rangle : c \in C\}$. To see this, consider the following diagram:

$$\begin{tikzcd}
D \arrow[r, "\pi_1^D"] \arrow[d, "i"'] & C \arrow[d, "\mathsf{inr}_C^{\#}"] \\
Sat \arrow[r, "\pi_2^{Sat}"'] & \mathcal{L}_{\mathsf{F}}
\end{tikzcd}$$

The map i is the inclusion of D in Sat, and $\mathsf{inr}_C^{\#}$ is the restriction of $\mathsf{inr}^{\#}$ to C. We check that we have a pullback. Suppose that $\langle d, \psi\rangle \in Sat$ and $\psi = \mathsf{inr}^{\#}c$ for some $c \in C$. Then by definition of C, $d = c$. So $\langle d, \psi\rangle \in D$, and $\pi_1^D(\langle d, \psi\rangle) = c$. Thus, this square is a pullback. The image of the square under F is therefore a weak pullback. Let $w \in \mathsf{F}(Sat)$ be such that $(\mathsf{F}\pi_1^{Sat})w = a$ and $(\mathsf{F}\pi_2^{Sat})w = (\mathsf{F}\mathsf{inr}_C^{\#})a$. Recall that a belongs to $\mathsf{F}(C)$. By the weak pullback condition, there is some $\overline{w} \in \mathsf{F}(D)$ such that $(\mathsf{F}\pi_1^D)\overline{w} = b$ and $(\mathsf{F}i)\overline{w} = w$. Since $\mathsf{F}i$ is an inclusion, $w = \overline{w}$. Thus, $(\mathsf{F}\pi_2^D)\overline{w} = (\mathsf{F}\mathsf{inr}^{\#})a$.

Now $\pi_2^D = \mathsf{inr}_C^{\#} \circ \pi_1^D$, so $\mathsf{F}\pi_2^D = \mathsf{F}\mathsf{inr}_C^{\#} \circ \mathsf{F}\pi_1^D$. Thus

$$(\mathsf{F}\mathsf{inr}_C^{\#})a \quad = \quad (\mathsf{F}\pi_2^D)\overline{w} \quad = \quad (\mathsf{F}\mathsf{inr}_C^{\#} \circ \mathsf{F}\pi_1^D)\overline{w} \quad = \quad (\mathsf{F}\mathsf{inr}_C^{\#})b.$$

But $\mathsf{F}\mathsf{inr}^{\#}$ and $\mathsf{F}\mathsf{inr}_C^{\#}$ are one-to-one, since $\mathsf{F}\mathsf{inr}^{\#}$ comes from an inductive extension of a one-to-one function, and $\mathsf{F}\mathsf{inr}_C^{\#}$ is a restriction of it. (In terms of our discussion above, each inr_α is one-to-one.) Hence $b = a$. This proves that $a \in C$. $\dashv$

Concerning the connection of $\mathsf{inr}^{\#}a$ with the formulas $j_\alpha(a)$, it is easy to check that for all α, $j_\alpha(a) \leq \mathsf{inr}^{\#}a$. It follows that $\theta^a_{\overline{\mathcal{F}}} \leq \mathsf{inr}^{\#}a$. But θ^a is maximal (that is, it implies any consistent formula stronger than itself), and so we have $\theta^a \equiv \mathsf{inr}^{\#}a$.

7 Characterization by single formulas

Consider the functor $\mathsf{F}(a) = \mathcal{P}(a) \times \mathcal{P}(\mathsf{AtProp})$ coming from our discussion of modal logic. It follows immediately from Theorem 2.4 and Proposition 4.7 that we have a characterization result: for each element $a \in \overline{\mathcal{F}}$ there is some sufficiently large α such that if $b \models_{\overline{\mathcal{F}}} j^a_\alpha$, then $b = a$. The purpose of this section is to extend this result to as wide a context as possible.

We might note that the characterization result does not hold for all functors whatsoever. In fact, Example 4.4 shows that the characterization result will not hold for the identity functor I. That it, the greatest fixed point is the class V of all sets. The characterization result stated in Theorem 7.1 is false for I. (Note also that any one-point set $\{*\}$ gives a final coalgebra for I, and therefore the results of the last section are trivial for I.) So some kind of extra condition on F is required.

There are a number of conditions in the literature which are intended to say that F behaves on morphisms by a kind of "substitution." These conditions have names like *uniform on maps* (Aczel [A]) and *smoothness* (Barwise and Moss [BM1]). The relation between these two conditions was considered in Moss and Danner [MD]. In fact, the exact relation between these kinds of conditions is complicated by the fact that [BM1] works in a set theory with urelements; these are the basis of the definitions surrounding smoothness. We will work with yet another definition, one due to Turi and Rutten [TR] and also called *uniformity on maps*. We choose to work with this primarily because Theorem 7.1 worked out. It would be interesting to formulate a version of smoothness for set theory without urelements and to use it to get a characterization result; a second proof of Theorem 7.1 might lead to missing insights. In any case, we are going to work with a slight generalization of Turi and Rutten's condition which we call *uniformity* (simpliciter). The definition involves some machinery, and we turn to that development.

The functor W For any set or class C, $W(C)$ is the least fixed point of the operator $a \mapsto C + \mathcal{P}(a)$. So $W(C) = C + \mathcal{P}(W(C))$. We also have injections $\mathsf{inl}_{WC} : C \to W(C)$ and $\mathsf{inr}_{W(C)} : \mathcal{P}(W(C)) \to W(C)$. We think of $W(C)$ as the class of sets obtainable from iterations of the power set operation beginning with C as a set of atoms. The recursion principle from the last section holds for $W(C)$. To state it again, for any $f : C \to B$ and $g : \mathcal{P}(B) \to B$ (so that $[f, g] : C + \mathcal{P}(B) \to B$), there is a unique $[f, g]^{\#} : WC \to B$. This map $[f, g]^{\#}$ has the properties that $f = [f, g]^{\#} \circ \mathsf{inl}_{WC}$, and $g \circ \mathcal{P}[f, g]^{\#} = [f, g]^{\#} \circ \mathsf{inr}_{W(C)}$.

The recursion principle allows us to turn W into a functor. Indeed, given $f : C \to D$, we define $Wf : W(C) \to W(D)$ to be $[\mathsf{inl}_{W(D)} \circ f, \mathsf{inr}_{W(D)}]^{\#}$. It also allows us to define maps $\epsilon_C : WC \to V = [\mathsf{incl}_C, \bigcup]^{\#}$, where $\mathsf{incl}_C : C \to V$ is the inclusion, and $\bigcup : \mathcal{P}(V) \to V$ is the unary union operation on sets.

We also need a slight generalization of this notion. Let Q be a set, and let $q : Q \to V$ be a function. There is a functor $W_{Q,q}$ taking a set or class C to $W(C + Q)$, and a function $f : C \to D$ to $W_{Q,q}f : W_{Q,q}(C) \to W_{Q,q}(D)$, where

$$W_{Q,q}f \quad = \quad [(\mathsf{inl}_{W_{Q,q}(D)} \circ (f + \mathsf{id}_Q)) + \mathsf{inr}_{W_{Q,q}(D)}]^{\#}.$$

Here $f+\mathsf{id}_Q : C+Q \to D+Q$ is the natural map. Furthermore, we have $\epsilon_{Q,q,C} : W_{Q,q}(C) \to V$ given by

$$\epsilon_{Q,q,C} \quad = \quad [[\mathsf{incl}_C, q] + \bigcup]^{\#}.$$

Definition [Turi and Rutten [TR], Turi [T]] F is *uniform* if there are a set Q, a map $q : Q \to V$ and a natural transformation $\rho : F \to \mathcal{P}W_{Q,q}$ such that $\mathcal{P}\epsilon_{Q,q,V} \circ \rho_{\mathsf{F}V} = \mathsf{incl}_{\mathsf{F}V}$, where $\mathsf{incl}_{\mathsf{F}(V)}$ is the inclusion of $\mathsf{F}(V)$ in V.

Our formulation of this concept is slightly more general than the original formulation because we allow Q and q to appear. This extension is needed, and the theorem quoted above still goes through: assuming *AFA*, F^* is a final coalgebra.

Nearly all of the examples of functors from earlier in the paper are uniform in this sense. We check this for the case of $\mathsf{F}(b) = C \times b$, where C is a fixed set. We take $Q = C$ and $q = \mathsf{id}_C$. We define $\rho_A : C \times A \to W(A+C)$ to be

$$\langle c, a\rangle \quad \mapsto \quad \mathsf{inr}_{W(A+C)}(\{\mathsf{inr}_{W(A+C)}(\{\mathsf{inr}_{A+C}(c)\}), \mathsf{inr}_{W(A+C)}(\{\mathsf{inr}_{A+C}(c), \mathsf{inl}_{A+C}(a)\}\}).$$

This is just a translation of the usual pair $\langle c, a\rangle = \{\{c\}, \{c, a\}\}$ into the formal apparatus of $W(A+C)$. It is routine to check that ρ defined in this way gives a natural transformation. It is also clear from this example why one wants ρ_A to map into the well-founded sets over A and C. Further, is also evident that one wants ρ_A to map into $W(A+Q)$ for some set Q and not just into $W(A)$.

We mentioned above that the uniform functors contain nearly all of the functors that we mentioned in Section 3. The exception is the identity functor; it is not uniform. (To see this, suppose ρ exists, and consider $\rho_V(\emptyset)$. On the one hand $(\mathcal{P}\epsilon_{Q,q,V} \circ \rho_V)\emptyset = \emptyset$. On the other, ρ_V would map into $\mathcal{P}W(V)$, and $\mathcal{P}\epsilon_{Q,q,V}[\mathcal{P}W(V)] = V - \{\emptyset\}$.) The uniform functors are closed under composition.

Assuming *AFA*, Turi and Rutten [TR] showed that the greatest fixed point F^* of a uniform functor gives a final F-coalgebra $(\mathsf{F}^*, \mathsf{id})$. The inspiration for this theorem was Aczel's Special Final Coalgebra Theorem of [A]: assuming *AFA*, the greatest fixed point of a functor F which is uniform on sets also gives a final F-coalgebra. (*AFA* is the Anti-Foundation Axiom. This result depends critically on *AFA* in that it is false if one assumes instead the Foundation Axiom.) Because the coalgebra map of F^* is the identity, and because F^* is a fixed point of F, it offers a slight simplification of notation relative to any other final coalgebra. So in this section we are going to use F^* as a final coalgebra. However, we could formulate things differently, and indeed our work in this section does not use *AFA* in any critical way. Here is the main result of this section:

Theorem 7.1 *Let F be uniform. Consider the coalgebra $(\mathsf{F}^*, \mathsf{id})$, and the canonical formulas $j_\alpha(a)$ for $\mathcal{L}_{\mathsf{F}}$. For all $a \in \mathsf{F}^*$ there is some α such that $j_\alpha(a)$ characterizes a in F^*. That is, if $b \models_{\mathsf{F}^*} j_\alpha(a)$, then $b = a$.*

In this section, we write $\mathcal{M}$ for $\mathcal{L}_{\mathcal{P}}$ and $\mathcal{L}$ for $\mathcal{L}_{\mathsf{F}}$. Note that $\mathcal{M} = \mathcal{P}\mathcal{M} + \mathcal{P}\mathcal{M}$, and $\mathcal{M}$ is the initial $(\mathcal{P}+\mathcal{P})$-algebra. Similarly, $\mathcal{L} = \mathcal{P}\mathcal{L} + \mathsf{F}\mathcal{L}$, and $\mathcal{L}$ is the initial $(\mathcal{P}+\mathsf{F})$-algebra. We also will use subscripts to distinguish the two injections $\mathsf{inr}_{\mathcal{L}} : \mathsf{F}\mathcal{L} \to \mathcal{L}$ and $\mathsf{inr}_{\mathcal{M}} : \mathcal{P}\mathcal{M} \to \mathcal{M}$; we distinguish the injections $\bigwedge_{\mathcal{L}}$ and $\bigwedge_{\mathcal{M}}$ similarly. However, to save on notation, we drop as many Q's as possible, and abbreviate $W(\mathsf{F}^* + Q)$ to $W\mathsf{F}^*$, and $\epsilon_{Q,q}$ to ϵ.

The main idea We define a translation $i : \mathcal{L} \to \mathcal{M}$. For example, suppose as above that $\mathsf{F}(a) = C \times a$. The map i will be defined by recursion on $\mathcal{L}$, so suppose we have i defined on φ. Consider a formula of $\mathcal{L}$ such as $\mathsf{inr}_{\mathcal{L}}(\langle c, \varphi \rangle)$. Then we will translate this into $\mathcal{M}$ by

$$i(\mathsf{inr}_{\mathcal{L}}(\langle c, \varphi \rangle) \quad = \quad \mathsf{inr}_{\mathcal{M}}(\{\mathsf{inr}_{\mathcal{M}}(\{\theta(c)\}), \mathsf{inr}_{\mathcal{M}}(\{\theta(c), i(\varphi)\})\}).$$

Here $\theta(c)$ is the formula of $\mathcal{M}$ which characterizes c in the class $\mathcal{P}^*$ of all sets. We will need to define and study the translation, of course, and this leads to the machinery below. We begin with a list of maps:

$$\begin{array}{ll}
\mathsf{incl} : \mathsf{F}^* \to V & \mathsf{incl}(a) = a \\
\epsilon = \epsilon_{W\mathsf{F}^*} : W\mathsf{F}^* \to V & \epsilon \;=\; [(\mathsf{incl} + q), \bigcup]^{\#} \\
g : W\mathcal{M} \to \mathcal{M} & g \;=\; [[\mathsf{id}_{\mathcal{M}}, \theta \circ q]\,,\, \mathsf{inr}_{\mathcal{M}}]^{\#} \\
h : \mathsf{F}\mathcal{M} \to \mathcal{M} & h \;=\; g \circ \mathsf{inr}_{W\mathcal{M}} \circ \rho_{\mathcal{M}} \\
i : \mathcal{L} \to \mathcal{M} & i \;=\; [\bigwedge_{\mathcal{M}}, h]^{\#}
\end{array}$$

In order to distinguish the canonical maps for $\mathcal{L}$ and $\mathcal{P}$, we write j_α for the map determined by F and k_α for the map determined by $\mathcal{P}$. So $j_\alpha : \mathsf{F}^* \to \mathcal{L}$ and $j_{\alpha+1} = \mathsf{inr}_{\mathcal{L}} \circ \mathsf{F} j_\alpha$; $k_\alpha : \mathcal{P}^* \to \mathcal{M}$, $k_{\alpha+1} = \mathsf{inr}_{\mathcal{M}} \circ \mathcal{P} k_\alpha$. We also want to have versions of these maps that are defined on F^*. We therefore define $u_\alpha : \mathsf{F}^* \to \mathcal{M}$ and $v_\alpha : \mathsf{F}^* \to \mathcal{M}$ by recursion so that

$$\begin{array}{lcl|lcl}
u_0 & = & \mathit{true}_{\mathcal{M}} & v_0 & = & \mathit{true}_{\mathcal{M}} \\
u_{\alpha+1} & = & h \circ \mathsf{F} u_\alpha & v_{\alpha+1} & = & [[v_\alpha, k_\alpha \circ q]\,,\, \mathsf{inr}_{\mathcal{M}}]^{\#} \circ \mathsf{inr}_{W\mathsf{F}^*} \circ \rho_{\mathsf{F}^*} \\
u_\lambda & = & \bigwedge_{\mathcal{M}}\{u_\beta \mid \beta < \lambda\} & v_\lambda & = & \bigwedge_{\mathcal{M}}\{v_\beta \mid \beta < \lambda\}
\end{array}$$

The next two results make clear the connection of u_α with j_α and v_α with k_α.

Lemma 7.2 *For all α, $i \circ j_\alpha = u_\alpha$.*

Proof By induction on α. For $\alpha = 0$, we check directly that $i(\mathit{true}_{\mathcal{L}}) = \mathit{true}_{\mathcal{M}}$. Assume that $i \circ j_\alpha = u_\alpha$. Note that $i \circ \mathsf{inr}_{\mathcal{L}} = h \circ \mathsf{F} i$. Therefore

$$\begin{array}{lcl}
i \circ j_{\alpha+1} & = & i \circ \mathsf{inr}_{\mathcal{L}} \circ \mathsf{F} j_\alpha \\
& = & h \circ \mathsf{F} i \circ \mathsf{F} j_\alpha \\
& = & h \circ \mathsf{F} u_\alpha \\
& = & u_{\alpha+1}
\end{array}$$

The limit step is trivial. ⊣

Lemma 7.3 *For all α, $v_\alpha = k_\alpha \circ \mathsf{incl}$.*

Proof Here is the successor step of the induction. Assume that $v_\alpha = k_\alpha \circ \mathsf{incl}$. Then an easy induction on $W\mathsf{F}^*$ shows that $[[v_\alpha, k_\alpha \circ q]\,,\, \mathsf{inr}_{\mathcal{M}}]^{\#} = k_\alpha \circ \epsilon$. Thus for all $a \in \mathsf{F}^*$,

$$\begin{array}{lcl}
v_{\alpha+1} & = & [[v_\alpha, k_\alpha \circ q]\,,\, \mathsf{inr}_{\mathcal{M}}]^{\#} \circ \mathsf{inr}_{W\mathsf{F}^*} \circ \rho_{\mathsf{F}^*} \\
& = & \mathsf{inr}_{\mathcal{M}} \circ \mathcal{P}[[v_\alpha, k_\alpha \circ q]\,,\, \mathsf{inr}_{\mathcal{M}}]^{\#} \circ \rho_{\mathsf{F}^*} \\
& = & \mathsf{inr}_{\mathcal{M}} \circ \mathcal{P}(k_\alpha \circ \epsilon) \circ \rho_{\mathsf{F}^*} \\
& = & \mathsf{inr}_{\mathcal{M}} \circ \mathcal{P} k_\alpha \circ \mathcal{P}\epsilon \circ \rho_{\mathsf{F}^*} \\
& = & k_{\alpha+1} \circ \mathsf{incl}
\end{array}$$

⊣

Lemma 7.4 *For all $a \in \mathsf{F}^*$ and all $\varphi \in \mathcal{L}$, if $a \models_{\mathsf{F}^*} \varphi$, then $a \models_{\mathcal{P}^*} i\varphi$.*

Proof By induction on $\mathcal{L}$. Let $S \subseteq \mathcal{L}$ be the set of formulas for which this holds. Clearly S is closed under $\bigwedge_{\mathcal{L}}$. Suppose that $\varphi \in \mathsf{inr}[\mathsf{F}(S)]$. Let $a \in \mathsf{F}^*$ be such that $a \models_{\mathsf{F}^*} \varphi$. By Lemma 5.2, there is a reason w for this which belongs to $\mathsf{F}(Y)$, where

$$Y \quad = \quad \{\langle b, \psi\rangle \in \mathsf{F}^* \times S : b \models_{\mathsf{F}^*} \psi\}.$$

Let $Z = \{\langle b, i\psi\rangle : \langle b, \psi\rangle \in Y\}$. There is an evident map $l : Y \to Z$ so that $\pi_1^Z \circ l = \pi_1^Y$ and $\pi_2^Z \circ l = i \circ \pi_2^Y$. All of the squares commute in the following diagram:

$$\begin{tikzcd}
\mathsf{F}^* \arrow[r, "\rho_{\mathsf{F}^*}"] & \mathcal{P}W(\mathsf{F}^*) \arrow[r, "1_{\mathcal{P}W\mathsf{F}^*}"] & \mathcal{P}W(\mathsf{F}^*) \arrow[r, "\mathsf{inr}_{W(\mathsf{F}^*)}"] & W(\mathsf{F}^*) \arrow[r, "\epsilon"] & V \\
\mathsf{F}(Y) \arrow[u, "\mathsf{F}\pi_1^Y"] \arrow[r, "\rho_Y"] \arrow[d, "\mathsf{F}\pi_2^Y"'] & \mathcal{P}W(Y) \arrow[u, "\mathcal{P}W\pi_1^Y"'] \arrow[r, "\mathcal{P}Wl"] \arrow[d, "\mathcal{P}W\pi_2^Y"'] & \mathcal{P}W(Z) \arrow[u, "\mathcal{P}W\pi_1^Z"'] \arrow[r, "\mathsf{inr}_{WZ}"] \arrow[d, "\mathcal{P}W\pi_2^Z"'] & W(Z) \arrow[u, "W\pi_1^Z"'] \arrow[d, "W\pi_2^Z"'] & \\
\mathsf{F}(\mathcal{L}) \arrow[r, "\rho_{\mathcal{L}}"'] & \mathcal{P}W(\mathcal{L}) \arrow[r, "\mathcal{P}Wi"'] & \mathcal{P}W(\mathcal{M}) \arrow[r, "\mathsf{inr}_{W\mathcal{M}}"'] & W(\mathcal{M}) \arrow[r, "g"'] & \mathcal{M}
\end{tikzcd}$$

Claim For all $x \in W(Z)$, $(\epsilon \circ W\pi_1^Z)x \models_{\mathcal{P}^*} (g \circ W\pi_2^Z)x$.

The proof is by induction on $W(Z)$. (Recall that $W(Z)$ is actually an abbreviation for $W(Z+Q)$.) Let S be the set of x for which the claim holds. We first show that for all $x \in Z$, $(\mathsf{inl}_{WZ} \circ \mathsf{inl}_{Z+Q})x \in S$. That is,

$$(\epsilon \circ W\pi_1^Z \circ \mathsf{inl}_{WZ} \circ \mathsf{inl}_{Z+Q})x \models_{\mathcal{P}^*} (g \circ W\pi_2^Z \circ \mathsf{inl}_{WZ} \circ \mathsf{inl}_{Z+Q})x.$$

The set on the left above is $\pi_1(x)$, and the formula on the right is $\pi_2(x)$. By definition of Z, $\pi_1(x) \models_{\mathcal{P}^*} \pi_2(x)$. Next, we show that for $x \in Q$, $(\mathsf{inl}_{WZ} \circ \mathsf{inr}_{Z+Q})x \in S$. That is,

$$(\epsilon \circ W\pi_1^Z \circ \mathsf{inl}_{WZ} \circ \mathsf{inr}_{Z+Q})x \models_{\mathcal{P}^*} (g \circ W\pi_2^Z \circ \mathsf{inl}_{WZ} \circ \mathsf{inr}_{Z+Q})x.$$

This time, the set on the left is $q(x)$, and the formula on the right is $(\theta \circ q)x$. Once again, $q(x) \models_{\mathcal{P}^*} (\theta \circ q)x$. This proves the claim for all $x \in Z$. For the induction step, let $y \in \mathcal{P}(S)$; we check that $\mathsf{inr}_{WZ}(y) \in S$. Now for all $x' \in y$, $(\epsilon \circ W\pi_1^Z)x' \models_{\mathcal{P}^*} (g \circ W\pi_2^Z)x'$. So by the semantics of $\models_{\mathcal{P}^*}$,

$$(\mathcal{P}(\epsilon \circ W\pi_1^Z))y \models_{\mathcal{P}^*} (\mathsf{inr}_{\mathcal{M}} \circ \mathcal{P}(g \circ W\pi_2^Z))y.$$

But the recursion equations for $W\pi_1^Z$ and ϵ tell us that $\mathcal{P}\epsilon \circ \mathcal{P}W\pi_1^Z = \epsilon \circ W\pi_1^Z \circ \mathsf{inr}_{WZ}$, and $\mathsf{inr}_{\mathcal{M}} \circ \mathcal{P}g \circ \mathcal{P}W\pi_2^Z = g \circ W\pi_2^Z \circ \mathsf{inr}_{WZ}$. Therefore

$$(\epsilon \circ W\pi_1^Z)\mathsf{inr}_{WZ}(y) \models_{\mathcal{P}^*} (g \circ W\pi_2^Z)\mathsf{inr}_{WZ}(y).$$

So $\mathsf{inr}_{WZ}(y) \in S$ as desired. This completes the proof of the claim.

Recall that we have $w \in \mathsf{F}(Y)$ such that $(\mathsf{F}\pi_1^Y)w = a$ and $(\mathsf{inr}_{\mathcal{L}} \circ \mathsf{F}\pi_2^Y)w = \varphi$. By the claim,

$$(\epsilon \circ W\pi_1^Z \circ \mathsf{inr}_{WZ} \circ \mathcal{P}Wl \circ \rho_Y)w \ \models_{\mathcal{P}^*} \ (g \circ W\pi_2^Z \circ \mathsf{inr}_{WZ} \circ \mathcal{P}Wl \circ \rho_Y)w.$$

But

$$\begin{array}{rcl}
(\epsilon \circ W\pi_1^Z \circ \mathsf{inr}_{WZ} \circ \mathcal{P}Wl \circ \rho_Y)w & = & (\epsilon \circ \mathsf{inr}_{W\mathsf{F}^*} \circ 1_{\mathcal{P}W\mathsf{F}^*} \circ \rho_{\mathsf{F}^*} \circ \mathsf{F}\pi_1^Y)w \\
& = & (\epsilon \circ \mathsf{inr}_{W\mathsf{F}^*} \circ \rho_{\mathsf{F}^*})a \\
& = & (\mathcal{P}\epsilon \circ \rho_{\mathsf{F}^*})a \\
& = & a \\
\\
(g \circ W\pi_2^Z \circ \mathsf{inr}_{WZ} \circ \mathcal{P}Wl \circ \rho_Y)w & = & (g \circ \mathsf{inr}_{W\mathcal{M}} \circ \mathcal{P}Wi \circ \rho_{\mathcal{L}} \circ \mathsf{F}\pi_2^Y)w \\
& = & (g \circ \mathsf{inr}_{W\mathcal{M}} \circ \rho_{\mathcal{M}} \circ \mathsf{F}i \circ \mathsf{F}\pi_2^Y)w \\
& = & (h \circ \mathsf{F}i \circ \mathsf{F}\pi_2^Y)w \\
& = & (i \circ \mathsf{inr}_{\mathcal{L}} \circ \mathsf{F}\pi_2^Y)w \\
& = & i\varphi
\end{array}$$

In these calculations, we have used the commutativity of the squares in the diagrams above, the recursion equations coming from the functor W, and uniformity. We conclude that $a \models_{\mathcal{P}^*} i\varphi$. Since $a \in \mathsf{F}^*$ is arbitrary, this shows that $\varphi \in S$. We conclude that $\mathsf{inr}[\mathsf{F}(S)] \subseteq S$. This concludes the proof. $\dashv$

Lemma 7.5 *For all α, $v_\alpha \leq u_\alpha$ as elements of $\mathsf{F}^* \to \mathcal{M}$.*

Proof As usual, we need only show the successor step. Assume that $v_\alpha \leq u_\alpha$. We first show by induction on $W\mathsf{F}^*$ that $[[v_\alpha, k_\alpha \circ q]\,,\, \mathsf{inr}_{\mathcal{M}}]^\# \leq g \circ Wu_\alpha$. Here is the proof in brief: First,

$$\begin{array}{rcl}
[[v_\alpha, k_\alpha \circ q]\,,\, \mathsf{inr}_{\mathcal{M}}]^\# \circ \mathsf{inl}_{W\mathsf{F}^*} & = & [v_\alpha, k_\alpha \circ q] \\
& \leq & [u_\alpha, \theta \circ q] \\
& = & [\mathsf{id}_{\mathcal{M}}, \theta \circ q] \circ [u_\alpha, \mathsf{id}_Q] \\
& = & (g \circ Wu_\alpha) \circ \mathsf{inl}_{W\mathsf{F}^*}.
\end{array}$$

Note that we used the maximality of each θ^a to write $k_\alpha \circ q \leq \theta \circ q$. Second,

$$\begin{array}{rcl}
[[v_\alpha, k_\alpha \circ q]\,,\, \mathsf{inr}_{\mathcal{M}}]^\# \circ \mathsf{inr}_{W\mathsf{F}^*} & = & \mathsf{inr}_{\mathcal{M}} \circ \mathcal{P}v_\alpha \\
& \leq & \mathsf{inr}_{\mathcal{M}} \circ \mathcal{P}u_\alpha \\
& = & (g \circ Wu_\alpha) \circ \mathsf{inr}_{W\mathsf{F}^*}.
\end{array}$$

Here we used the induction Monotonicity Lemma. This completes our claim.

By naturality of ρ, $\rho_{\mathcal{M}} \circ \mathsf{F}u_\alpha = \mathcal{P}Wu_\alpha \circ \rho_{\mathsf{F}^*}$. So

$$\begin{array}{rcl}
v_{\alpha+1} & = & [[v_\alpha, k_\alpha \circ q]\,,\, \mathsf{inr}_{\mathcal{M}}]^\# \circ \mathsf{inr}_{W\mathsf{F}^*} \circ \rho_{\mathsf{F}^*} \\
& \leq & g \circ Wu_\alpha \circ \mathsf{inr}_{W\mathsf{F}^*} \circ \rho_{\mathsf{F}^*} \\
& = & g \circ \mathsf{inr}_{W\mathcal{M}} \circ \mathcal{P}Wu_\alpha \circ \rho_{\mathsf{F}^*} \\
& = & g \circ \mathsf{inr}_{W\mathcal{M}} \circ \rho_{\mathcal{M}} \circ \mathsf{F}u_\alpha \\
& = & h \circ \mathsf{F}u_\alpha \\
& = & u_{\alpha+1}
\end{array}$$

This completes the induction. $\dashv$

Proof of Theorem 7.1 Let $a \in \mathsf{F}^*$. Then $a \in V = \mathcal{P}^*$, so let α be such that $k_\alpha(a)$ characterizes a among all sets. We prove that for this α, $j_\alpha(a)$ characterizes a in F^*. For if $b \models_{\mathsf{F}^*} j_\alpha(a)$, then $b \models_{\mathcal{P}^*} u_\alpha(a)$ by Lemmas 7.2 and 7.4. Also, by Lemma 7.5, $v_\alpha(a) \leq u_\alpha(a)$ in $\mathcal{M}$. So $b \models_{\mathcal{P}^*} v_\alpha(a)$. But $k_\alpha(a) = v_\alpha(a)$ by Lemma 7.3. So $b \models_{\mathcal{P}^*} k_\alpha(a)$. By the characterization result for $\mathcal{P}^*$, $a = b$. This concludes the proof. $\dashv$

We conclude this section with a generalization of well-known characterization results for several standard logics. Corollary 7.6 below is actually a corollary of the proof of Theorem 7.1. Note that the functors described in it are all uniform. In addition, the natural transformation ρ actually maps $\mathsf{F}(A)$ into $V_\omega(A + Q)$ for some finite set Q. A *finitary* formula of $\mathcal{L}_\mathsf{F}$ is one built from F alone, without using conjunctions. Then the proof of Theorem 2.4 shows that sets which are hereditarily finite are characterized among all sets by countable conjunctions of finitary formulas. Putting all of this together gives the following result.

Corollary 7.6 *Let F be a composition of one of the following functors: $a \mapsto C \times a$, $a \mapsto C + a$, or $a \mapsto C \to a$, where C is a fixed finite set; or $\mathcal{P}_{fin}$, the finite power set. Let $\mathcal{E}$ and $\mathcal{E}'$ be F-coalgebras, and let $a \in \mathcal{E}$ and $a \in \mathcal{E}'$. If a and a' satisfy the same finitary formulas of $\mathcal{L}_\mathsf{F}$, then there is an F-bisimulation of $\mathcal{E}$ and $\mathcal{E}'$ relating a to a'.*

Remark Although Theorem 7.1 gives a characterization result, there are two reasons not to be satisfied with it. First, it may not be the best possible result; it might be possible to weaken the hypothesis that F be uniform. (A possibly related point: we do not know whether a characterization result for $\mathcal{L}_\mathsf{F}$ implies that the greatest fixed point F^* is a final F-coalgebra.) Second, the overall strategy of the proof is to translate $\mathcal{L}$ into $\mathcal{M}$ and then appeal to the characterization for result for sets. It would be interesting to know whether one could work directly with F^* and work out enough combinatorial details there to get a characterization result.

8 Some Sound Principles

Now that we have a syntax and semantics for the coalgebraic logic $\mathcal{L}_\mathsf{F}$, it is natural to ask for logical systems which could give completeness results for the semantics. That is, define $\varphi \models \psi$ to mean that for every coalgebra $\mathcal{E}$ and every $a \in \mathcal{E}$, if $a \models \varphi$, then $a \models \psi$. (This is the same as $\psi \leq \varphi$.) We would like to have an inductive definition of a class $\mathit{Proves} \subseteq \mathcal{L} \times \mathcal{L}$, so that, writing $\varphi \vdash \psi$ for $\mathit{Proves}(\varphi, \psi)$, we have $\varphi \vdash \psi$ iff $\varphi \models \psi$.

Let $\vdash$ be the least relation on $\mathcal{L}$ such that

(Weakening L) If $\varphi \vdash \psi$, then for all $S \subseteq \mathcal{L}$, $\bigwedge(S \cup \{\varphi\}) \vdash \psi$.

(Weakening R) If $\varphi \vdash \bigwedge S$, then for all $\psi \in S$, $\varphi \vdash \psi$.

($\bigwedge$ R) If $\varphi \vdash \psi$ for all $\psi \in S$, then $\varphi \vdash \bigwedge S$.

(Cut) If $\varphi \vdash \psi$ and $\psi \vdash \chi$, then $\varphi \vdash \chi$.

(F) If $K \in \mathsf{F}(\mathit{Proves})$, then $\mathsf{inr} \circ \mathsf{F}\pi_1(K) \vdash \mathsf{inr} \circ \mathsf{F}\pi_2(K)$.

We call this logic the *minimal* F*-logic*.

We give an example of (F), the only interesting rule of the system. Consider the functor $\mathsf{F}(A) = \mathcal{P}(A) \times \mathcal{P}(\mathsf{AtProp})$ corresponding to $\mathcal{L}(\triangle)$. Note that $\{true\} \vdash true$ by ($\bigwedge$ R). (We omit the injections in this example.) So by (F), we get a number of additional facts. For example, $\{p, \{true\}\} \vdash \{p, true\}$, and also $\{p, q, \{true\}\} \vdash \{p, q, true\}$.

Theorem 8.1 (Soundness) *The minimal* F*-logic is sound: if* $\varphi \vdash \psi$*, then* $\varphi \models \psi$*.*

Proof By induction on the relation *Proves*. Let Pr be any sound set of pairs; we check that adding the instances of the rules of the logic preserves soundness. This is immediate for the all of the rules except (F).

Suppose that $K \in \mathsf{F}(Pr)$, that $\mathcal{E} = \langle A, e\rangle$ is an F-coalgebra, and that $a \models_{\mathcal{E}} \mathsf{inr} \circ \mathsf{F}\pi_1^{Pr}(K)$. We prove that $a \models_{\mathcal{E}} \mathsf{inr} \circ \mathsf{F}\pi_2^{Pr}(K)$. We adapt Example 3.2, taking our set A, $B = \mathcal{L}_{\mathsf{F}} = C$, $R = Sat_{\mathcal{E}}$, and $S = Pr$. The semantics of $\mathcal{L}_{\mathsf{F}}$ tells us that there is some $w \in \mathsf{F}(Sat_{\mathcal{E}})$ such that $(\mathsf{F}\pi_1)w = e_a$ and $(\mathsf{F}\pi_2)w = (\mathsf{F}\pi_1^S)K$.

When we apply F to the pullback square, we get a weak pullback. The weak pullback condition applies to w and K, and we see that there is some $p \in \mathsf{F}(S \star R)$ such that $(\mathsf{F}\pi_1^{S\star R})p = w$ and $(\mathsf{F}\pi_2^{S\star R})p = K$. Recall also that we have a map $h : S \star R \to S \circ R$. So $\mathsf{F}h : \mathsf{F}(S \star R) \to \mathsf{F}(S \circ R)$. Further, $(\mathsf{F}h)p$ has the following properties:

$$\begin{aligned}
(\mathsf{F}\pi_1^{S\circ R} \circ \mathsf{F}h)p &= (\mathsf{F}\pi_1^{A} \circ \mathsf{F}\pi_1^{S\star R})p \\
&= (\mathsf{F}\pi_1^{A})w \\
&= e_a, \\
\\
(\mathsf{F}\pi_2^{S\circ R} \circ \mathsf{F}h)p &= (\mathsf{F}\pi_2^{S} \circ \mathsf{F}\pi_2^{S\star R})p \\
&= (\mathsf{F}\pi_2^{S})K
\end{aligned}$$

We are using (3) here. Note also that $S \circ R = Pr \circ Sat_{\mathcal{E}} \subseteq Sat_{\mathcal{E}}$. The reason is that Pr is sound by our overall induction assumption. Finally, by standardness we see that for $i = 1, 2$, $\mathsf{F}\pi_i^{S\circ R}$ is a subfunction of $\mathsf{F}\pi_i^{Sat_{\mathcal{E}}}$. It follows that $(\mathsf{F}h)p$ is a reason why $a \models_{\mathcal{E}} (\mathsf{F}\pi_2^{Pr})K$. $\dashv$

We suspect that the minimal F-logic is not complete. In particular, we conjecture that no instance of the following scheme is provable in the minimal F-logic.

(D) If $a \neq b$ are members of $\overline{\mathcal{F}}$ and φ is any formula, then $\theta^a \wedge \theta^b \vdash \varphi$.

Each instance of this rule (D) is also sound: For all coalgebras $\mathcal{E}$ and all $a, c \in \mathcal{E}$, it cannot be the case that $c \models \theta^a$ and $c \models \theta^b$. This is because $c \models \theta^a$ implies that $s(c) = a$, where $s : C \to \overline{\mathcal{F}}$ is the final coalgebra map. Of course, Rule (D) is completely ad-hoc, and it would be nice to have a system that derived it from more appealing principles.

Finally, here is another sound principle:

(FP) If $a \in \overline{\mathcal{F}}$, then $\theta^a \vdash (\mathsf{inr} \circ \mathsf{F}j)\theta^a$ and $(\mathsf{inr} \circ \mathsf{F}j)\theta^a \vdash \theta^a$.

The soundness results of this section obviously call out for matching completeness results. At the present time, there are no known completeness results for any logic $\mathcal{L}_{\mathsf{F}}$. However, Baltag [Bal] has a complete system for the full modal logic $\mathcal{L}(\mathsf{AtProp})$.

9 Representation of final coalgebras as maximal elements of ordered algebras

Ever since people looked at final coalgebras as models of intensional phenomena, there was a question of getting representations of the final coalgebras in terms of some sort of entities that served as approximations. This was felt strongly in the case of sets, where the operator was the power set operation. Indeed, not long after the appearance of Aczel's book[A], a number of papers appeared on the subject of getting domain-theoretic representations of the non-wellfounded sets. For example, Mislove et al [MMO] show how to obtain a domain-theoretic representation of the set HF of hereditarily finite sets in terms of initial ordered algebras of a certain type. In other work, Barr[Bar93, Bar94] considered endofunctors on the category of sets and showed that under suitable conditions, the final coalgebras exist and are completions of the initial algebras. This result is another example of an approximation result. Finally, Paulson [P] is concerned with getting concrete final coalgebra theorems for some endofunctors. He adopts Quine's pairing operation rather than the standard Kuratowski pair, and his methods do not work for endofunctors such as the power set operation. Despite these differences, his motivation is the same as ours and the other papers cited above: to obtain concrete domains in which one can find objects representing infinite computations and data structures, and manipulate them naturally.

In this section, we use $\mathcal{L}_{\mathsf{F}}$ and its semantics to show how to represent final coalgebras in terms of maximal elements of ordered algebras.

Definition Let F be a functor, and let $\mathcal{A} = \langle A, \alpha \rangle$ be an F-algebra. If R is any relation on A, we get a relation R' on $\mathsf{F}(A)$ by

$$R' \quad = \quad \{\langle (\mathsf{F}\pi_1)r, (\mathsf{F}\pi_2)r \rangle : r \in R\}.$$

An F-*algebra order on* $\mathcal{A}$ is a partial order $\leq$ on A with the property that if $a \leq' b$, then $\alpha a \leq \alpha b$.

It is not hard to check that in the case of syntactic functors, this gives the usual definition of an ordered algebra.

Here is a statement of our main result:

Theorem 9.1 *Let* F *be a uniform functor. Then there is an* F-*algebra* $\langle P, \alpha \rangle$ *and an* F-*algebra order* $\leq$ *on* P *such that*

1. *The order* $\leq$ *is a complete semi-lattice.*
2. *The collection of elements of* P *which are* $\leq$-*maximal (but below* $\top$*) is the carrier of a final* F-*coalgebra* $\langle \overline{\mathcal{F}}, e \rangle$.

The ideology behind this result is that we have a poset P with nice continuity properties (suprema of all sets). Note that $\mathcal{L}_{\mathsf{F}}$ has a top element $\top$. Any conjunction $\bigwedge\{\theta^a, \theta^b\}$ of inconsistent information is such a $\top$ element, where a and b are distinct elements of $\overline{\mathcal{F}}$. (In all cases of interest, $\overline{\mathcal{F}}$ will have more than one element.) By a *maximal* element, we mean some $x \in P$ such that if $x < y$, then $y = \top$, where $\top = \bigvee P$. Because $\overline{\mathcal{F}}$ is embedded by α as the collection of maximal elements of P, we think of P as a domain of approximations to $\overline{\mathcal{F}}$.

In our construction for Theorem 9.1, we use formulas of $\mathcal{L}_{\mathsf{F}}$ (modulo semantic equivalence) as elements of P. This conforms to the intuition about approximations because logical formulas can always be regarded as approximations to the objects which satisfy them. When we have characterization results, then this connection is quite tight. This is the key idea behind the development which follows.

The poset P Let $\leq$ be the natural semantic preorder on $\mathcal{L}_{\mathsf{F}}$: $\varphi \leq \psi$ iff for all coalgebras $\mathcal{E}$ and all $a \in \mathcal{E}$ such that $a \models_{\mathcal{E}} \psi$, we also have $a \models_{\mathcal{E}} \varphi$. So *true* is at the bottom of the preorder.

Let $\equiv$ be the equivalence relation determined by this preorder. We define equivalence classes (actually sets) by

$$[\phi] \quad = \quad \{\psi \in \mathcal{L} : \psi \text{ is a formula of minimal rank such that } \psi \equiv \phi\}.$$

The notion of rank here is from the inductive definition of $\mathcal{L}_{\mathsf{F}}$ as the least fixed point of $\mathcal{P} + \mathsf{F}$. For our poset P_{F}, we take the set or class of these:

$$P \quad = \quad \{[\phi] : \phi \in \mathcal{L}\} \tag{6}$$

Actually, P will be a *set* provided that there is a fixed cardinal κ such that every coalgebra map $e : A \to \mathsf{F}(A)$ takes values in $H_\kappa(A)$, the set of sets of hereditary cardinality at most κ. This holds for examples such as $\mathsf{F}(a) = A \times a$ and $\mathsf{F}(a) = \mathcal{P}_{fin}(a)$. But for examples like $\mathsf{F}(a) = \mathcal{P}(a)$, P will be a proper class. Indeed, in all the examples we considered, P is a proper class iff $\overline{\mathcal{F}}$ is a proper class. For functors as in Corollary 7.6, P has the cardinality of the continuum.

The algebra map $\alpha : \mathsf{F}(P) \to P$ The basic idea is to take α to be F, or rather some map related to it defined on equivalence classes.

We have a map $\mathsf{nat} : \mathcal{L} \to P$ which assigns to each φ its equivalence class $[\varphi]$. Fix a choice function $\mathsf{ch} : P \to \mathcal{L}$ which acts as an inverse in the sense that $\mathsf{nat} \circ \mathsf{ch} = \mathsf{id}_P$ and $\mathsf{ch} \circ \mathsf{nat} = \mathsf{id}_{\mathcal{L}}$. (In the case that P is a proper class, we are using the Global Axiom of Choice here. Also, ch is not uniquely determined, but this is of no consequence.) Set

$$\alpha \quad = \quad \mathsf{nat} \circ \mathsf{inr} \circ \mathsf{Fch},$$

so that $\alpha : \mathsf{F}(P) \to P$. We might point out that α does not depend on the particular map ch used in its definition. For suppose ch' is another choice function. We apply the Monotonicity Lemma with $A = P$, $f = \mathsf{ch}$, and $g = \mathsf{ch}'$. Then as maps from $\mathsf{F}(P)$ to $\mathcal{L}$, $\mathsf{inr} \circ \mathsf{Fch} \equiv \mathsf{inr} \circ \mathsf{Fch}'$. Thus as maps from $\mathsf{F}(P)$ to P, $\mathsf{nat} \circ \mathsf{inr} \circ \mathsf{Fch} = \mathsf{nat} \circ \mathsf{inr} \circ \mathsf{Fch}'$.

Lemma 9.2 $\mathsf{nat} \circ \mathsf{inr} = \alpha \circ \mathsf{Fnat}$.

Proof Recall that $\mathsf{ch} \circ \mathsf{nat} \equiv \mathsf{id}_{\mathcal{L}}$. So by the Monotonicity Lemma, $\mathsf{inr} \circ \mathsf{F}(\mathsf{ch} \circ \mathsf{nat}) \equiv \mathsf{id}_{\mathsf{F}(\mathcal{L})}$. Thus as elements of $\mathsf{F}(\mathcal{L}) \to P$, $\alpha \circ \mathsf{Fnat} = \mathsf{nat} \circ \mathsf{inr} \circ \mathsf{F}(\mathsf{ch} \circ \mathsf{nat}) = \mathsf{nat} \circ \mathsf{inr}$. $\dashv$

The F-algebra order $\leq$ on P P inherits the semantic order $\leq$ from $\mathcal{L}$. The verification that this is an F-algebra order is essentially the proof that rule (F) is sound for the minimal coalgebraic logic. P is a complete upper-semilattice. For every set S of (equivalence classes of) formulas, the least upper bound of S, $\bigsqcup S$, is $[\bigwedge\{\mathsf{ch}\, p : p \in S\}]$.

Maps $\theta_{\mathcal{E}}$ for each F-coalgebra $\mathcal{E}$ Let $\mathcal{E}$ be an F-coalgebra. Fix an element $a \in \mathcal{E}$. By Theorem 7.1, there is an ordinal α large enough so that $j_\alpha(a) \equiv j_{\alpha+1}(a)$. In this way, we define $\theta_{\mathcal{E}} : \mathcal{E} \to \mathcal{L}_{\mathsf{F}}$. Each $\theta_{\mathcal{E}}(a)$ is a maximal element of $\mathcal{L}_{\mathsf{F}}$.

Lemma 9.3 $\mathsf{inr} \circ \mathsf{F}\theta_{\mathcal{E}} \circ e \equiv \theta_{\mathcal{E}}$.

Proof Let $a \in \mathcal{E}$. For some ordinal α, $\theta_{\mathcal{E}}(e_a) = j_\alpha(a)$. Moreover, $j_{\alpha+1}(a) \equiv j_\alpha(a)$. So $(\mathsf{inr} \circ \mathsf{F}\theta_{\mathcal{E}})e_a = j_{\alpha+1}(a) \equiv j_\alpha(a) = \theta_{\mathcal{E}}(a)$. Since this is true for all $a \in \mathcal{E}$, we are done. $\dashv$

Conclusion of the proof of Theorem 9.1. Let $\langle \overline{\mathcal{F}}, e\rangle$ be a final F-coalgebra. Then e is a bijection, and so $\langle \overline{\mathcal{F}}, e^{-1}\rangle$ is an algebra. Consider $\theta = \mathsf{nat} \circ \theta_{\overline{\mathcal{F}}}$. Then $\theta : \overline{\mathcal{F}} \to P$ takes values in the maximal elements of P by the characterization result. It is injective for the same reason. To see that it is surjective, let $[\varphi]$ be maximal. Since $\varphi \neq \top$, let $a \models_{\mathcal{E}} \varphi$. Then $\theta_{\mathcal{E}}(a) \equiv \varphi$. And if $s : \mathcal{E} \to \mathcal{F}$ is the final coalgebra map, $\theta_{\overline{\mathcal{F}}}(c_a) = \theta_{\mathcal{E}}(a) \equiv \varphi$. This shows that θ is a bijection between $\overline{\mathcal{F}}$ and the maximal elements of P. We show that θ is a morphism of algebras; i.e., that $\mathsf{nat} \circ \theta_{\overline{\mathcal{F}}} \circ e^{-1} = \alpha \circ \mathsf{F}(\mathsf{nat} \circ \theta_{\overline{\mathcal{F}}})$. For this, note that

$$\begin{array}{rcll}
\alpha \circ \mathsf{F}(\mathsf{nat} \circ \theta_{\overline{\mathcal{F}}}) & = & (\alpha \circ \mathsf{F}\mathsf{nat}) \circ \mathsf{F}\theta_{\overline{\mathcal{F}}} & \\
& = & \mathsf{nat} \circ \mathsf{inr} \circ \mathsf{F}\theta_{\overline{\mathcal{F}}} & \text{by Lemma 9.2} \\
& = & \mathsf{nat} \circ (\mathsf{inr} \circ \mathsf{F}\theta_{\overline{\mathcal{F}}} \circ e) \circ e^{-1} & \\
& = & \mathsf{nat} \circ \theta_{\overline{\mathcal{F}}} \circ e^{-1} & \text{by Lemma 9.3}
\end{array}$$

In this way, the collection of maximal elements of P is the carrier of an F-coalgebra which is isomorphic to $\overline{\mathcal{F}}$.

We have three final remarks: First, Theorem 9.1 does not use *AFA*; in fact, none of the results of this paper used *AFA*. In a sense, Theorem 9.1 can be regarded as a final coalgebra theorem in its own right. It gives a representation of the final coalgebra for uniform functors which is more concrete than taking a quotient, but less concrete, say than Paulson's final coalgebra theorem in [P].

For many purposes, it would be better to have versions of Theorem 9.1 in which we only used *consistent formulas*. That is, it is desirable to remove $\top$ from P and still have an approximation result. It is not hard to check that if we consider $P_0 = P \setminus \{[\mathit{true}]\}$, then P_0 is the carrier of an F-algebra structure. Of course, the order $\leq$ is no longer complete. One would hope that it is directed complete, and this would correspond to a compactness property of the semantics. We have some results in this direction, but these require some additional assumptions on F.

Finally, by moving to stronger languages, we can also obtain embedding results like Theorem 9.1 but which ask for stronger properties on the order relation. For example, suppose we worked not with $\mathcal{L}_{\mathsf{F}}$ but with the stronger language $\mathcal{L}_{\mathsf{F}}(\neg)$. Then $\mathcal{L}$, and hence P, would have all meets. All of the results on the semantics would go through, because the induction steps for $\neg$ would be trivial. The reason that we chose to work with $\mathcal{L}$ rather than $\mathcal{L}(\neg)$ is that $\mathcal{L}$ seems to be the weakest language closed under infinitary conjunction which allows for a characterization result such as Theorem 7.1. So any stronger language will have the same property, a fortiori. More expressive languages may be easier to work with and to study, but in this paper we have tried to get more general results.

Acknowledgement

I thank Alexandru Baltag for several discussions on the topics of this paper.

References

[A] Aczel, Peter. 1988. *Non-Well-Founded Sets.* CSLI Lecture Notes Number 14. Stanford: CSLI Publications.

[AM] Aczel, Peter and Nax Mendler. 1989. A Final Coalgebra Theorem. In D. H. Pitt et al (eds.) *Category Theory and Computer Science*, 357–365. LNCS Heidelberg: Springer Verlag.

[Bal] Alexandru Baltag. 1996. Modal charactersation for sets and Kripke models. Unpublished manuscript, Indiana University.

[Bar93] Barr, Michael. 1993. Terminal coalgebras in well-founded set theory. *Theoretical Computer Science* 114 (2), 299–315.

[Bar94] Barr, Michael. 1994. Additions and corrections to "Terminal coalgebras in well-founded set theory." *Theoretical Computer Science* 124 (1), 189–192.

[BM1] Barwise, Jon, and Lawrence S. Moss. 1996. Vicious Circles. Stanford: CSLI Publications Number 60.

[BM2] Barwise, Jon, and Lawrence S. Moss. 1996. Modal Correspondence for Models. In P. Dekker and M. Stokhof (eds.) *Proceedings of the Tenth Amsterdam Colloquium*, ILLC/Department of Philosophy, University of Amsterdam, 41–56. (This paper will also appear in the *Journal of Philosophical Logic.*)

[Ben] Johan van Bethem. 1976. *Modal Correspondence Theory.* Ph.D. Thesis, Department of Mathematics, Unviersity of Amsterdam.

[F] Fagin, Ronald. 1994. A Quantitative Analysis of Modal Logic. *Journal of Symbolic Logic* 59 (1), pp. 209–252.

[G] Gerbrandy, Jelle. 1996. Characterizations of bisimulation and bounded bisimulations. In M. de Rijke (ed.) *Workshop Proceedings Observational Equivalence and Logical Equivalence*, European Summer School on Logic, Language, and Information, Prague, 1996.

[MD] Moss, L. S. and N. Danner. 1997 On the Foundations of Corecursion. *Logic Journal of the IGPL*, Vol. 5, No. 3, pp. 231–257.

[MMO] Mislove, Michael W., Lawrence S. Moss and Frank J. Oles. 1991. Non-Well-Founded Sets Modeled as Ideal Fixed Points. Information and Computation. vol. 93, No. 1, pp. 16–54.

[P] Paulson, L. 1995. A Concrete Final Coalgebra Theorem for ZF Set Theory. In: P. Dybjer, B. Nordstrom and J. Smith (eds), *Types for Proofs and Programs '94*. Springer LNCS 995, 120-139.

[R96] Rutten, J.J.M.M. 1996. Universal Coalgebra: A Theory of Systems. CWI Report CS-R9652, CWI Computer Science Department.

[T] Turi, D. 1996. Functorial Operational Semantics and its Denotational Dual. Ph.D. thesis, CWI, Amsterdam.

[TR] D. Turi and Rutten, J.J.M.M. 1996. On the Foundations of Final Semantics: non-standard sets, metric spaces, partial orders. Revised version of a paper with the same title in *Proceedings of the REX Workshop on Semantics: Foundations and Applications*, ed. J.W. de Bakker, W.-P. de Roever, and G. Rozenberg, LNCS, Vol. 666, 477–530. Beekbergen. Springer-Verlag.