THE INEXHAUSTIBLE CONTENT OF MODAL BOXES

Johan van Benthem

1 Introduction: the good old days

Dick de Jongh has been a long-time friend and colleague, ever since we met in those late sixties when the world was young. It is a pleasure to think back of the many things, both domestic and exotic, that we have done together. But in this note, I want to focus on just one special interest that ties us together, viz.

Löb's axiom $[]([]p \rightarrow p) \rightarrow []p$

in provability logic. When my predecessor Martin Löb arrived in Amsterdam, he left it to Dick to explain his seminal theorem, encoded in this axiom, to us eager graduate students. In my recollection, this marked the beginning of an active modal logic period, including a wonderful joint seminar on intuitionistic logic and provability logic by Dick and Craig Smorynski, a creative and unusual American visitor whose work we had already learnt about in the classes of Anne Troelstra. By then, I had just started working on modal frame correspondence theory – and Löb's Axiom was a nice challenge, since it does not fit the usual axioms in the *S4* or *S5* pattern. I remember coming up one day with a direct semantic argument that it corresponded precisely to the conjunction of well-foundedness of the alternative relation plus the first-order condition of transitivity:

Fact Löb's Axiom is true at point s in a frame $\boldsymbol{F} = (W, R)$ iff
(a) $\boldsymbol{F}$ is upward R-well-founded starting from s, and
(b) $\boldsymbol{F}$ is transitive at s, i.e., $\boldsymbol{F}, s \models \forall y(Rxy \rightarrow \forall z(Ryz \rightarrow Rxz))$.

The latter clause was a bit strange, as the *S4*-axiom $[]p \rightarrow [][]p$ corresponding to transitivity was postulated separately in provability logic. It took Dick just one day to come up with a beautiful syntactic derivation of the transitivity axiom from Löb's Axiom, one of the many elegant proofs that form his 'signature' in the area. [1]

[1] Later, Wim Blok – another participant in our logic milieu – found the more complex derivation of $[]p \rightarrow [][]p$ from Grzegorczyk's Axiom (the counterpart of Löb's Axiom on reflexive frames), using algebraic methods. Cf. van Benthem & Blok 1978. This was to have been one of many illustrations in our planned joint book on connections between modal logic and universal algebra, commissioned by Anne Troelstra for "Studies in Logic" as a sort of merge of our dissertations. This book unfortunately never happened, although chapter drafts are still lying around.

Despite our early encounters, Dick and I have gone separate ways. His road took him to the study of provability and interpretability in arithmetical theories, whereas my interests were in general modal logic and its connections with first- and higher-order logic. In particular, whenever I look at results in provability logic, I ask myself – to the exasperation of many colleagues, I fear: – how much of this is a general fact of logic, and what reflects special features of arithmetic such as induction and numerical coding power? In the same spirit, I often wonder whether some key result in modal logic is really more general than stated. Indeed, I once managed to drag Dick into a bit of research along these lines, on the more general model theory behind the De Jongh-Sambin Fixed-Point Theorem. We found that it holds for large non-modal classes of generalized quantifiers on well-founded orderings (see Section 3 below). There was no joint paper: just a brief report in van Benthem 1987, which used such fixed-point results in analyzing 'semantic automata' computing quantifiers in natural language. After that, I have occasionally tried to lure Dick, or his students and colleagues, into systems of general interpretability and other meta-logic of first-order model theory – without any special role for numbers and coded proof predicates. All my attempts met with a remarkable lack of success. [2]

But it is too late to change myself! My present offering is again a bunch of logical observations in a general perspective, all taking Löb's Axiom as their starting point.

2 Scattering modal axioms

Perhaps the primordial frame correspondence concerns the above *S4*-axiom.

Fact $[]p \rightarrow [][]p$ is true in a frame $\boldsymbol{F} = (W, R)$ (i.e., true at all worlds under all valuations) iff $\boldsymbol{F}$'s accessibility relation R is transitive.

Local versions of such correspondences work in each world *s* separately, but I will use global versions for convenience. Behind the Fact lies a general result by Sahlqvist, discovered independently in my dissertation. Here is just one version:

Theorem There is an algorithm computing first-order frame correspondents for modal formulas $\alpha \rightarrow \beta$ with an antecedent α constructed from atoms possibly prefixed by universal modalities, conjunction, disjunction, and existential modalities, and the consequent β any syntactically positive modal formula.

---

[2] Albert Visser has informed me, though, about much ongoing spin-off of provability logic, which strikes out for greater logical generality in a number of new directions.

The algorithm works as follows:

(a) translate the modal axiom into its canonical first-order form, prefixed with monadic set quantifiers for proposition letters: $\forall x: \forall \mathbf{P}: translation(\phi)(\mathbf{P}, x)$,
(b) pull out all existential modalities in the antecedent, and turn them into bounded universal quantifiers in the prefix,
(c) compute a first-order *minimal valuation* for the proposition letters making the remaining portion of the antecedent true,
(d) substitute this definable valuation for the proposition letters occurring in the body of the consequent – and if convenient,
(e) perform some simplifications modulo logical equivalence.

For details of this 'substitution algorithm' and a proof of its semantic correctness, cf. Blackburn, de Rijke &Venema 2000.

Example For the modal transitivity formula $[]p \rightarrow [][]p$,

(a) yields $\forall x: \forall P: \forall xy\,(Rxy \rightarrow Py) \rightarrow \forall z\,(Rxz \rightarrow \forall u\,(Rzu \rightarrow Pu))$,
(b) is vacuous, while (c) yields the minimal valuation $Ps := Rxs$ – and then
(d) substitution yields $\forall x: \forall xy(Rxy \rightarrow Rxy) \rightarrow \forall z(Rxz \rightarrow \forall u(Rzu \rightarrow Rxu))$.
The latter simplifies to the usual form $\forall x: \forall z\,(Rxz \rightarrow \forall u\,(Rzu \rightarrow Rxu))$. ♣

Now we make a simple observation. Notice that the very same procedure would also work if all these modalities were entirely *independent*, as in the following formula:

Fact $[1]p \rightarrow [2][3]p$ also has a first-order frame correspondent, computed in exactly the same fashion, viz. $\forall x: \forall z\,(R_2xz \rightarrow \forall u\,(R_3zu \rightarrow R_1xu))$.

Definition The *scattered version* of a modal formula ϕ arises by marking each modality in ϕ uniquely with an index for its own accessibility relation. ♣

Evidently, the conditions in the Sahlqvist Theorem apply to the scattered version of any implication of the sort described above. The reason is that these conditions make statements about individual occurrences: they do not require pairwise co-ordination of occurrences. This sort of condition is very frequent in logical meta-theorems, and hence many results have more general scattered versions. [3]

[3] Scattering makes sense in other formal languages, too – but we stick with the modal case here.

Now let us take this notion to a typically non-first-order modal principle like Löb's Axiom. We find that it still makes sense there, leading to the following generalized frame correspondence result for the scattered version.

Fact The modal formula $[1]([2]p \rightarrow p) \rightarrow [3]p$ is equivalent on frames $\boldsymbol{F} = (W, R_1, R_2, R_3)$ to the conjunction of the relational conditions
(a) $R_3 ; (R_2)^* \subseteq R_1$ (with $(R_2)^*$ the reflexive-transitive closure of R_2)
(b) upward well-foundedness in the following sense: no world x starts an infinite upward sequence of worlds $x R_3 y_3 R_2 y_2 R_2 y_3 ...$

I forego a proof here, but it should not be difficult to cognoscenti of the original Löb correspondence. [4] E.g., if conditions (a) and (b) hold but (c) fails, then the valuation making *p* false only on the infinite *y*-sequence will falsify the scattered Löb Axiom at the world *x*. This successful generalization might suggest that scattering keeps modal formulas more or less the same qua expressive power – but this is not true.

Fact There are first-order frame-definable modal formulas whose scattered versions are not first-order definable.

Proof Consider the first-order definable modal formula which conjoins the transitivity axiom with the so-called McKinsey Axiom (cf. van Benthem 1983):

$$([]p \rightarrow [][]p) \,\&\, ([]<>p \rightarrow <>[]p)$$

Even its partly scattered version $([1]p \rightarrow [1][1]p) \,\&\, ([2]\langle 2\rangle p \rightarrow \langle 2\rangle [2]p)$ is not first-order definable. In any frame, taking the universal relation for R_1 will verify the left conjunct, and so, substituting these, the purported total first-order equivalent would become a first-order equivalent for the McKinsey axiom: *quod non*. ♣

Scattering seems of general interest to me, for several reasons. It focuses the search for *most general versions* of modal results, it fits with *pluriform provability logics* where each box stands for a different arithmetical theory, and finally, the interplay of many modalities fits with the current trend toward *combining logics*.

---

[4] This observation arose out of an email correspondence with Chris Steinsvold from CUNY New York, who had looked at the axiom $[1]([2]p \rightarrow p) \rightarrow [1]p$. The above Fact about the scattered version of Löb's Axiom has also been found independently by Melvin Fitting.

3 Frame conditions in fixed-point logic

Now let's go back to standard correspondence. [5] Löb's Axiom seems typically beyond the syntactic range of the Sahlqvist Theorem, as its antecedent has a modal box over an implication. But still, its frame-equivalent of transitivity plus well-foundedness, though not first-order, is definable in a natural extension – viz. *LFP(FO): first-order logic with fixed-point operators* (Ebbinghaus & Flum 1995).

Fact The *well-founded part* of any binary relation R is definable as a smallest fixed-point of the monotone set operator $\boldsymbol{[]}(X) = \{y \mid \forall z(Ryz \rightarrow z \in X)\}$.

The simple proof is, e.g., in Aczel 1977. The well-founded part is written in the language of *LFP(FO)* as the smallest-fixed-point formula

$$\mu P, x \bullet \forall y\,(Rxy \rightarrow Py)$$

How can we find modal frame equivalents of this extended *LFP(FO)*-definable form as systematically as first-order frame conditions? Löb's Axiom suggests a general principle, as the *minimal valuation* step in the substitution algorithm still works. Consider the antecedent $[](\Box p \rightarrow p)$. If this modal formula holds anywhere in a model $\boldsymbol{M}, x$, then there must be a smallest predicate P for p making it true at $\boldsymbol{M}, x$ – because of a set-theoretic property guaranteeing a minimal verifying predicate:

Fact If $[](\Box p_i \rightarrow p_i)$ holds at a world x for all $i \in I$,
then $[](\Box P \rightarrow P)$ holds at x for $P = \bigcap_{i \in I} [[p_i]]$

This fact is easy to check. Here is the more general notion behind this observation.

Definition A first-order formula $\phi(P, \boldsymbol{Q})$ has the *intersection property* if, in every model $\boldsymbol{M}$, whenever $\phi(P, \boldsymbol{Q})$ holds for all predicates in some family $\{\boldsymbol{P}_i \mid i \in I\}$, it also holds for the intersection, that is: $\boldsymbol{M}, \bigcap \boldsymbol{P}_i \models \phi(P, \boldsymbol{Q})$. ♣

Now, the Löb antecedent displays a typical syntactical format which ensures that the intersection property must hold. We can specify this more generally as follows.

Definition A first-order formula is a *PIA condition* – short-hand for: 'positive antecedent implies atom' – if it has the following syntactic form:

$\forall \boldsymbol{x}(\phi(P, \boldsymbol{Q}, \boldsymbol{x}) \rightarrow P\boldsymbol{x})$ with P occurring only positively in $\phi(P, \boldsymbol{Q}, \boldsymbol{x})$. ♣

---

[5] The following section summarizes the main results of my paper 'Minimal Predicates, Fixed-Points, and Definability', ILLC Tech Report PP-2004-01.

The Löb antecedent then has the first-order *PIA* form

$$\forall y\,((Rxy \,\&\, \forall z(Ryz \rightarrow Pz)) \rightarrow Py)$$

Example Horn clauses
A simpler case of the *PIA* format is the universal Horn clause defining modal accessibility via the transitive closure of a relation *R*

$$Px \wedge \forall y \forall z\,((Py \wedge Ryz) \rightarrow Pz))$$

The minimal predicate *P* satisfying this consists of all points *R*-reachable from *x*. ♣

For the Löb antecedent itself, such an explicit description is a bit more complex.

Example Computing the minimal valuation for Löb's Axiom
Analyzing $[][]p \rightarrow p)$ a bit more closely, the minimal predicate satisfying the antecedent of Löb's Axiom at a world *x* describes the following set of worlds:

$$\{y \mid \forall z\,(Ryz \rightarrow Rxz) \text{ \& no infinite sequence of } R\text{-successors starts from } y\}.$$

Then, if we plug this description into the Löb consequent $[]p$, precisely the usual, earlier-mentioned conjunctive frame condition will result automatically. ♣

Here are a few facts about the general situation. First, by way of background, we have a preservation theorem showing that *PIA*-conditions are expressively complete for the intersection behaviour guaranteeing unique existence of minimal predicates:

Theorem The following are equivalent for all first-order formulas $\phi(P, \boldsymbol{Q})$:
(a) $\phi(P, \boldsymbol{Q})$ has the Intersection Property w.r.t. predicate *P*
(b) $\phi(P, \boldsymbol{Q})$ is definable by a conjunction of *PIA* formulas.

This result involves a pleasantly complex model-theoretic construction, but it is not relevant to our further concerns here. More to the point is that minimal predicates defined through intersections are definable in an ordinary fixed-point logic: [6]

Fact The minimal predicates for *PIA*-conditions are definable in *LFP(FO)*.

Plugging these into the proof of the Sahlqvist Theorem yields the correspondents:

---

[6] The technical *PIA* link here is that the minimal fixed-point of the monotone operation *F(P)* defined by a *P*-positive formula is the intersection of all 'pre-fixed points' *X* with $F(X) \subseteq X$.

Corollary Modal axioms with *PIA* antecedents and syntactically positive consequents have their corresponding frame conditions definable in *LFP(FO)*.

This generalized correspondence covers much more than just Löb's Axiom:

Fact The modal axiom $(\langle\rangle p \wedge [](p \rightarrow []p)) \rightarrow p$ has a *PIA* antecedent whose minimal valuation yields the *LFP(FO)*-frame-condition that, whenever *Rxy*, *x* can be reached from *y* by some finite sequence of successive *R*-steps. [7]

The complexity of the required substitutions can still vary considerably here, depending on the complexity of reaching the smallest fixed-point for the antecedent via the usual bottom-up ordinal approximation procedure. E.g., obtaining the well-founded part of a relation may take any ordinal up to the size of the model. But for Horn clauses with just atomic antecedents, the approximation procedure will stabilize uniformly in any model by stage ω, and the definitions will be simpler.

Thus, Löb's Axiom suggests a drastic extension of modal correspondence methods.

Even so, there are limits. Not every modal axiom yields to the fixed-point approach!

Fact The tense-logical axiom expressing Dedekind Continuity
is not definable by a frame condition in *LFP(FO).*

The reason is that Dedekind Continuity holds in *(IR, <)* and fails in *(Q, <),* whereas these two relational structures are equivalent w.r.t. *LFP(FO)*-sentences.

Conjecture The McKinsey Axiom $[]\langle\rangle p \rightarrow \langle\rangle[]p$, with its typically non-*PIA* antecedent $[]\langle\rangle p$, has no *LFP(FO)*-definable frame correspondent. [8]

4 Excursion: provability logic in the modal μ–calculus

A conspicuous trend in contemporary modal logic has been the *strengthening* of modal languages to remove expressive deficits of the base language with just *[], <>*. Often, this reflects a desire for optimal logic design, trying not to get stuck with peculiarities of some language just because it was the first to occur to our ancestors.

---

[7] Yet another illustration of this extended correspondence is the scattered Löb Axiom of Section 2, whose frame correspondent was a conjunction of typical *LFP(FO)*-conditions.

[8] Normally, one views the Löb' antecedent $[]([]p \rightarrow p)$ and the McKinsey antecedent $[]\langle\rangle p$ as being at the same level of complexity, beyond Sahlqvist forms. But in the present generalized analysis of minimizable predicates, the latter is much more complicated than the former!

One such extended language is the modal μ*–calculus* – the natural modal fragment of *LFP(FO)*, and at the same time, a natural extension of propositional dynamic logic. Harel, Kozen & Tiuryn 2000 has a quick tour of its syntax, semantics, and axiomatics. This formalism can define smallest fixed-points in the format

$\mu p \bullet \phi(p)$, provided that p occurs only positively in ϕ.

This adds general syntactic recursion, with no assumption on the accessibility order. In any model $\boldsymbol{M}$, the formula $\phi(p)$ defines an inclusion-monotone set transformation

$$F_\phi(X) = \{s \in \boldsymbol{M} \mid (\boldsymbol{M}, p:=X), s \models \phi\}$$

By the Tarski-Knaster Theorem, the operation F_ϕ must have a smallest fixed-point. This can be reached bottom-up by ordinal approximation stages

$$\phi^0, \quad ..., \quad \phi^\alpha, \phi^{\alpha+1}, \quad ..., \quad \phi^\lambda, \quad ...$$

with $\phi^0 = \varnothing$, $\phi^{\alpha+1} = F_\phi(\phi^\alpha)$, and $\phi^\lambda = \bigcup_{\alpha<\lambda} \phi^\alpha$

The smallest fixed-point formula $\mu p \bullet \phi(p)$ denotes the first stage where $\phi^\alpha = \phi^{\alpha+1}$. This can define, e.g., a typical transitive closure modality from dynamic logic like 'some ϕ-world is reachable in finitely many R_a-steps':

$$\langle a^* \rangle \phi = \mu p \bullet \phi \vee \langle a \rangle p.$$

Also included are greatest fixed points $\nu p \bullet \phi(p)$, definable as $\neg \mu p \bullet \neg \phi(\neg p)$. Smallest and greatest fixed-points need not coincide, and others may be in between. The μ–calculus is decidable, and its validities are axiomatized by two simple proof rules:

(i) $\mu p \bullet \phi(p) \leftrightarrow \phi(\mu p \bullet \phi(p))$

(ii) if $\vdash \phi(\alpha) \rightarrow \alpha$, then $\vdash \mu p \bullet \phi(p) \rightarrow \alpha$

This modal fixed-point logic captures many facts of interest here.

Fact The smallest fixed-point formula $\mu p \bullet [] p$ defines the well-founded part of the accessibility relation for $[]$ in any modal model.

Now, we can extend $[]$, $\langle\rangle$-based provability logic to a μ–calculus with fixed-point operators, and thereby restore some harmony between the modal language and its frame-correspondence language. In particular, dualizing the above $\langle a^* \rangle \phi$ gives dynamic logic-style modalities $[]^* \phi$ saying that ϕ is true at all worlds reachable in

the transitive closure of the accessibility relation R for single $[]$. Then we can play with explicit versions of correspondence arguments, and variations on Löb's Axiom.

Example Scattered Löb Revisited.
Basic correspondence arguments go through in this setting. [9] For instance, recall the scattered Löb Axiom in Section 2. One of the two conjuncts defining its was the condition that $R_3 ; (R_2)^* \subseteq R_1$, which corresponds to the modal axiom

$$[1]p \rightarrow [3][2^*]p$$

In the right dynamic language this is indeed derivable from a scattered Löb Axiom:

(a)	$[1]([2][2^*]p \rightarrow [2^*]p) \rightarrow [3]\,[2^*]p)$	scattered Löb axiom
(b)	$[2^*]p \leftrightarrow p \,\&\, [2][2^*]p$	fixed-point axiom for *
(c)	$p \rightarrow ([2][2^*]p \rightarrow [2^*]p)$	consequence of (b)
(d)	$[1]p \rightarrow [1]([2][2^*]p \rightarrow [2^*]p)$	consequence of (c)
(e)	$[1]p \rightarrow [3][2^*]p$	from (a), (d) ♣

But there are also other versions of Löb's Axiom now. For instance,

Fact $[]^*([]p \rightarrow p) \rightarrow \; []^*p$ defines just upward well-foundedness of R. [10]

Thus, transitivity would now need an explicit *K4*-axiom, separating the two aspects of provability logic explicitly. But this can also be done in another matter, using the μ-calculus definability of upward well-foundedness:

Fact Löb's Logic is equivalently axiomatized by the two principles
(a) $[]p \rightarrow [][]p$, (b) $\mu p\bullet\, []p$

This version seems quite illuminating to me. First, it is easy to derive Löb's Axiom from (a) and (b), using the above proof principles for smallest fixed-points. And also, unpacking these two principles for the specific case of $\mu p\bullet\, []p$, we see that (i) drops out, while (ii) becomes the induction rule that, if $\vdash []\alpha \rightarrow \alpha$, then $\vdash \alpha$.

As a final observation, we cast the close connection between provability logic and fixed-point logics yet differently.

---

[9] Many modal definability results have extensions to richer fragments of second-order logic. E.g., Sahlqvist consequents could be just any syntactically positive second-order formula.

[10] This is related to the partly scattered formula $[1]([2]p \rightarrow p) \rightarrow [1]p$, whose correspondent can be obtained by earlier methods. In particular, its '(a)-clause' now amounts to just '$R_1 ; R_2 \subseteq R_1$'.

Proposition Löb's Logic can be faithfully embedded into the μ–calculus.

Proof The translation doing this works as follows:

(a) replace every $[]$ in a formula ϕ by its transitive closure version $[]^*$

(b) for the resulting formula $(\phi)^*$, take the implication $\mu p \bullet [] p \rightarrow (\phi)^*$.

It is straightforward to check that a plain modal formula ϕ is valid on transitive upward well-founded models iff $\mu p \bullet [] p \rightarrow (\phi)^*$ is valid on all models. ♣

It should be fun to explore the consequences of this. E.g., decidability of Löb's Logic now follows from that of the μ–calculus. And the latter's uniform interpolation properties may also be significant. Or, thinking vice versa: can we extend the arithmetical interpretation of provability logic to the whole μ–calculus?

5 Fixed-points and fixed-points

But there are other fixed-point results in modal logic! This final section is merely a discussion of 'the right linkage' which has intrigued me for a long time. It may just show my ignorance, and perhaps some reader can set me right and clarify it all.

A special fixed-point theorem Here is a celebrated result on defining modal notions in provability logic, due to De Jongh and Sambin. It is the modal version of the arithmetical Fixed-Point Lemma from the proof of Gödel's Theorem.

Theorem Consider any modal formula equivalence $\phi(p, \boldsymbol{q})$ in which the proposition letter p only occurs in the scope of at least one box, while q is some sequence of other proposition letters. There exists a formula $\psi(\boldsymbol{q})$ such that $\psi(q) \leftrightarrow \phi(\psi(\boldsymbol{q}), \boldsymbol{q})$ is provable in the Löb's Logic, and moreover, any two solutions to this fixed-point equation w.r.t. ϕ are provably equivalent.

Smorynski 1984 gives a simple algorithm for computing the fixed-point $\psi(\boldsymbol{q})$. Typical outcomes are the following fixed points:

Example

equation:	solution:
$p \leftrightarrow []p$	$p = T$
$p \leftrightarrow \neg[]p$	$p = \neg[]\bot$
$p \leftrightarrow ([]p \rightarrow q)$	$p = ([]q \rightarrow q)$

More complex iteratively obtained solutions arise when the body of the modal equation has multiple occurrences of p. ♣

Now, let us look at the connection between this result and the *general* fixed-points of Sections 3, 4, definable in logical languages such as the μ–calculus or *LFP(FO)*.

What does the Fixed-Point Theorem really do? There are two aspects to the result:

(a) existence and uniqueness of the predicate defined
(b) explicit definability within the modal base language.

As to the first, existence and uniqueness of the above predicate p is just a general property of recursive definitions over a well-founded ordering – of the type shown correct in elementary set theory texts. But we also get the further information that this recursive predicate can be defined inside the original modal language, without μ- or ν-operators for modal fixed-points. Let's first look at the unique existence.

Comparing De Jongh-Sambin fixed-points and μ-calculus In general fixed-point logics, one defining formula may produce smallest and greatest fixed-points, and others in between. But we can compare the two approaches, in particular, the results of the general approximation procedure of Section 4 and the special-purpose algorithm mentioned just now. For a start, evidently, definitions $\mu p \bullet \phi(p)$ with only positive boxed occurrences of p in ϕ fall under both approaches.

Example The fixed-point for the modal equation $p \leftrightarrow []p$
In any modal model, the μ–calculus formula $\mu p \bullet []p$ defined the well-founded part of the accessibility ordering R. Thus, in well-founded models, it defines the whole universe – which explains the earlier solution T ('true'). ♣

But the De Jongh-Sambin Theorem also allows for negative occurrences of p in the defining equation. These fall outside of general fixed-point logics.

Example The fixed-point for the modal equation $p \leftrightarrow \neg []p$
Here, in general, the approximation sequence for the set operator $F_{\neg []p}$ can fail to yield any fixed point, with an approximation sequence oscillating all the way. E.g., in the model $(N, <)$, that sequence is $\varnothing, N, \varnothing, N, \ldots$ ♣

Actually, the situation in general fixed-point logic is a bit more complex. Formulas with mixed positive and negative occurrence can sometimes be admissible.

Example The mixed-occurrence formula $p \leftrightarrow (p \vee \neg []p)$
In this case, the approximation sequence will be monotonically non-decreasing, because of the initial disjunct p. So, there will be a smallest fixed-point! We can

even compute it in this particular case. The sequence stabilizes at stage 2, yielding $<>T$. There is also a greatest fixed-point, which is the whole set defined by T. ♣

Note that this formula falls outside the scope of the De Jongh-Sambin Theorem, as the first occurrence of p in $p \vee \neg[]p$ is not boxed. But then, there is no unique definability in this extended format, since the smallest and greatest fixed-points are different here. In general fixed-point logic, this example motivates an extension of the monotonic framework (cf. (Ebbinghaus & Flum 1995).

Definition *Inflationary fixed-points* for arbitrary formulas $\phi(p, \boldsymbol{q})$ without syntactic restrictions on the occurrences of p are computed using an ordinal approximation sequence which forces upward cumulation:

$\phi^{\alpha+1} = \phi^{\alpha} \cup \phi(\phi^{\alpha})$, taking unions again at limit ordinals. ♣

There is no guarantee that a set P where this stabilizes is a fixed-point for the modal formula $\phi(p, \boldsymbol{q})$. It is rather a fixed-point for the modified formula $p \vee \phi(p, \boldsymbol{q})$.

Combining the two sorts of fixed-point But comparison may also mean combination. Would *adding* general monotone fixed-points extend the scope of the De Jongh-Sambin result? The answer is no.

Fact Any p-positive formula $\mu p \bullet \phi(p)$ with $\phi(p)$ having unboxed occurrences of p is equivalent to one in which all occurrences of p occur boxed.

Proof Without loss of generality, we can take the formula to be of the form

$\mu p \bullet (p \& A) \vee B$ with only boxed occurrences of p in A, B

Let ϕ^{α} be the approximation sequence for $\phi = (p \& A) \vee B$, and let B^{α} be such a sequence executed separately for the formula B. We have the following collapse:

Claim $\phi^{\alpha} = B^{\alpha}$ for all ordinals α

This is proved by induction. The zero and limit cases are obvious. Next,

$\phi^{\alpha+1} = (\phi^{\alpha} \& A(\phi^{\alpha})) \vee B(\phi^{\alpha})$

$= (B^{\alpha} \& A(B^{\alpha})) \vee B(B^{\alpha})$

where by the fact that F_B is monotone: $B^{\alpha} \subseteq B(B^{\alpha})$, and hence $B^{\alpha} \cap A(B^{\alpha})) \subseteq B(B^{\alpha})$

$= B(B^{\alpha})$

$= B^{\alpha+1}$

Thus, the same fixed-point is computed by the boxed formula $\mu p \bullet B$. ♣

Thus, to some extent, adding a μ-calculus to provability logic has no effect. [11]

Here is another question concerning the interplay of the two formats for inductive definitions. Can we fit De Jongh-Sambin recursions into the general format of fixed-point logic? Note that the notion of well-founded order itself has an inductive character: it was the smallest fixed-point for the operator matching the modal box:

$$[](X) = \{y \mid \forall z(Ryz \rightarrow z \in X\}.$$

And on well-founded orders, this means that the whole universe is eventually computed through the monotonically increasing ordinal approximation stages

$$D^0, D^1, .., D^\alpha, ...$$

of the modal fixed-point formula $\mu r \bullet []r$. Now we cannot compute similar stages for the above fixed-point formula $\phi(p, \mathbf{q})$, since it may have both positive and negative occurrences of the proposition letter p. But we can define the related monotonic sequence of *inflationary* fixed-points, defined above. As we noted, this need not lead to a fixed-point for $\phi(p, \mathbf{q})$ per se. But this time, we do have monotone growth within the D-hierarchy, as the ϕ's stabilize inside its stages:

Fact $\quad \phi^{\alpha+1} \cap D^\alpha = \phi^\alpha \cap D^\alpha$

Thus a general fixed-point procedure for solving De Jongh-Sambin equations runs monotonically when restricted to approximation stages for a well-founded universe. This prediction pans out for the usual modal examples like the above $[]p$, $\neg[]p$, or $[]p \rightarrow q$. Their fixed-point is indeed easily found by reference to the D-stages. [12] We will not prove this here, as we will redescribe the situation now. Recall that the well-founded part of a relation was found as the smallest fixed-point $\mu r \bullet []r$.

Fact De Jongh-Sambin fixed-points can be found by the following *simultaneous inductive definition:*

$$r \quad \leftrightarrow \quad []r$$

$$p \quad \leftrightarrow \quad []r \ \& \ \phi(p, \mathbf{q})$$

---

[11] Still, it may make sense to extend the language in this way, as in Section 4. But to make that interesting, models should have an underlying accessibility relation that is no longer necessarily transitive. This is more natural anyway, e.g., when we study modal logics of finite trees.

[12] I have tried to make this whole definition *syntactically positive* by a trick of introducing two proposition letters: p_1 for p and p_2 for $\neg p$. But so far, this does not seem to do the job right.

Here we compute the approximation stages for p, r simultaneously:

$(r^{\alpha+1}, p^{\alpha+1})$	=	$([]r^{\alpha}, []r^{\alpha} \wedge \phi(p^{\alpha}))$	*successors*
$(r^{\lambda}, p^{\lambda})$	=	$(\cup_{\alpha<\lambda} r^{\alpha}, \cup_{\alpha<\lambda} p^{\alpha})$	*limits*

Note that the conjunct $[]r$ (rather than $'r'$) for p makes sure that the next stage of p is computed by reference to the new value of r. In fact we can prove the following relation between the stages, written here with some abuse of notation:

Lemma If $\beta<\alpha$, then $p^{\alpha} \wedge r^{\beta} = r^{\alpha}$

Note that this implies monotonicity: if $\beta<\alpha$, then $p^{\beta} \rightarrow p^{\alpha}$.

Proof The main induction is best done on α, with an auxiliary one on β. The inductive cases of *0* and limit ordinals α are straightforward. Now consider the successor step. We make use of the following two auxiliary facts. The first is a version of the invariance of modal formulas for generated submodels, and the second an immediate consequence of the approximation procedure for r:

(i) $\boldsymbol{M}, P, x \models \phi(p)$ iff $\boldsymbol{M}, P \cap R^{*}[x], x \models \phi(p)$

(ii) Let $R^{*}[x]$ be all points reachable from x by some finite number of R-steps. If $x \in r^{\alpha}$, then $R^{*}[x] \subseteq \cup_{\beta<\alpha} r^{\beta}$

Now we compute - again with some beneficial abuse of notation:

$x \models p^{\alpha+1} \wedge r^{\beta+1}$	iff
$x \models r^{\alpha+1} \wedge \phi(p^{\alpha}) \wedge r^{\beta+1}$	iff
$x \models \phi(p^{\alpha}) \wedge r^{\beta+1}$	iff (by (i), (ii))
$x \models \phi(p^{\alpha} \wedge r^{\beta}) \wedge r^{\beta+1}$	iff (ind. hyp.)
$x \models \phi(p^{\beta}) \wedge r^{\beta+1}$	iff
$x \models p^{\beta+1}$	♣

What is the real scope of explicit definability? Still, this analysis does not explain why the De Jongh-Sambin fixed-points are *explicitly* modally *definable*. Indeed, I do not understand the general reason. First, in my joint work with Dick in the 1980s, we found that this explicit definability is not specific to the modal language.

Theorem Explicit definability for fixed-point equations with all occurrences of p in the scope of some operator holds for all propositional languages with generalized quantifiers $\boldsymbol{Q}p$ over sets of worlds satisfying

(a) the above property (i): $\boldsymbol{Q}(P)$ is true at x iff $\boldsymbol{Q}(P \cap R_x)$ is true at x

(b) the Heredity Property $\boldsymbol{Q}p \rightarrow []\boldsymbol{Q}p$

This includes first-order quantifiers $\boldsymbol{Q}$ like "in at most five successors", or second-order ones like "in most successors of each successor". Cf. van Benthem 1987 for the simple argument, generalizing the proof of the De Jongh-Sambin Theorem.

But still, the general principle behind this extended explicit definability eludes us. [13] One important factor is the transitivity of accessibility. E.g., the Gödel fixed-point equation $p \leftrightarrow \neg[]p$ has no explicit modal solution on finite trees with the immediate successor relation. But there are other factors, too, such as the specific operators available in the language. [14] Here is yet another take. Smorynski proved the De Jongh-Sambin Theorem via a Beth theorem for Löb's Logic. Could there be some general Beth theorem in the background here for first-order languages and theories with frame conditions in suitable fragments of monadic second-order logic? [15] [16]

6 Conclusion

This note has shown how various aspects of provability logic, high-lighted by Löb's Axiom, suggest exploring a broader background in modal and classical logic, with fixed-point languages as a running thread. [17] My observations are mainly about possible connections, and work yet to be done. [18] But such as it is, this is my offering to Dick on the occasion of his retirement, which we all regret.

---

[13] Incidentally, this reduction of all inductive notions to the original base language seems a flaw, rather than a virtue, at least from the *general* perspective of fixed-point languages

[14] E.g., the preceding theorem suggests that 'explicit fixed-point theorems' might arise when we add enough 'modal operators' to the language for generalized quantifiers.

[15] Smallest and greatest fixed points for a first-order formula $\phi(P)$ coincide if $\phi(P)$ implies an explicit definition for P. The converse is true as well, by Beth's Theorem, as observed by Martin Otto, and Craig Smorynski. Such explicit first-order definitions even arise uniformly by some fixed finite approximation stage (Barwise-Moschovakis Theorem).

[16] Albert Visser and Giovavva d'Agostino have pointed out that there may also be a deeper analysis of these phenomena in terms of ideas from Hollenberg 1998, using uniform interpolation properties of the μ-calculus and its associated languages with *bisimulation quantifiers*.

[17] Another running theme could be scattered versions of fixed-point results. E.g., multiple occurrences of a fixed-point variable suggest coordination, and hence resistance to shattering.

[18] Thanks to Giovanna d'Agostino and Albert Visser for their useful and lively comments!

7 References

P. Aczel, 1977, 'An Introduction to Inductive Definitions', in J. Barwise, ed., *Handbook of Mathematical Logic*, North-Holland, 739–782, Amsterdam.

J. van Benthem, 1983, *Modal Logic and Classical Logic*, Bibliopolis, Napoli.

J. van Benthem, 1987, Towards a Computational Semantics', in P. Gärdenfors, ed., 1987, *Generalized Quantifiers: Linguistic and Logical Approaches*, Reidel, Dordrecht, 31-71.

J. van Benthem, 2004, 'Minimal Predicates, Fixed-Points, and Definability', Tech Report PP-2004-01, ILLC Amsterdam.

P. Blackburn, M. de Rijke & Y. Venema, 2000, *Modal Logic*, Cambridge University Press, Cambridge.

W. Blok & J. van Benthem, 1978, 'Transitivity Follows from Dummett's Axiom', *Theoria* 44:2, 117-118.

H-D Ebbinghaus & J. Flum, 1995, *Finite Model Theory*, Springer, Berlin.

D. Harel, D. Kozen & J. Tiuryn, 2000, *Dynamic Logic*, The MIT Press, Cambridge (Mass.).

M. Hollenberg, 1998, *Logic and Bisimulation*, Dissertation Series Vol. XXIV, Zeno Institute of Philosophy, University of Utrecht.

C. Smorynski, 1984, 'Modal Logic and Self-Reference', in D. Gabbay & F. Guenthner, eds., *Handbook of Philosophical Logic*, Vol. II, Reidel, Dordrecht, 441–495.