Abstract Syntax: Substitution and Binders

John Power [1,2]

School of Informatics
University of Edinburgh
King's Buildings
Edinburgh EH9 3JZ
Scotland

Abstract

We summarise Fiore et al's paper on variable substitution and binding, then axiomatise it. Generalising their use of the category $\mathbb{F}$ of finite sets to model untyped cartesian contexts, we let S be an arbitrary pseudo-monad on Cat and consider $(S1)^{op}$: this generality includes linear contexts, affine contexts, and contexts for the Logic of Bunched Implications. Given a pseudo-distributive law of S over the (partial) pseudo-monad $T_{coc}- = [(-)^{op}, Set]$ for free cocompletions, one can define a canonical substitution monoidal structure on the category $[(S1)^{op}, Set]$, generalising Fiore et al's substitution monoidal structure for cartesian contexts: this provides a natural substitution structure for the above examples. We give a concrete description of this substitution monoidal structure in full generality. We then give an axiomatic definition of a binding signature, then state and prove an initial algebra semantics theorem for binding signatures in full generality, once again extending the definitions and theorem of Fiore et al. A delicate extension of the research includes the category $Pb(Inj^{op}, Set)$ studied by Gabbay and Pitts in their quite different analysis of binders, which we compare and contrast with that of Fiore et al.

Keywords: Substitution, binding signature, initial algebra semantics, pseudo-distributive law.

1 Introduction

It is an honour and a pleasure for me to give the plenary lecture associated with the *MFPS* special session celebrating Gordon Plotkin's 60th birthday. In honour of the occasion, I shall address what I regard as one of his most interesting papers, *Abstract Syntax and Variable Binding*, co-authored by two of his proteges, Marcelo Fiore and Daniele Turi [4], and later axiomatised by me and his former student, Miki Tanaka [22,23,26,27,28,29].

In my view, the title of Fiore et al's paper is a misnomer: the paper is not primarily about variable binding, but rather about variable substitution. That is seen most clearly perhaps by comparing it with another paper, also ostensibly about variable binding, that appeared at the same conference, *LICS 99*, the other paper being Gabbay and Pitts' *A New Approach to Syntax Involving Binders* [6].

[1] This work has been done with the support of EPSRC grant GR/586372/01, A Theory of Effects for Programming Languages and on a visit to AIST, Senri-Chuo, Japan.

[2] Email: `ajp@inf.ed.ac.uk`

Fiore et al's paper is convoluted, but one can readily thread a path through it as follows. The authors first define the category $\mathbb{F}$ to be a skeleton of the category of finite sets and all functions. They then consider the functor category $[\mathbb{F}, Set]$ and define a substitution monoidal structure $\bullet$ on it. Having done that, and using both the finite product structure and the substitution monoidal structure of $[\mathbb{F}, Set]$, they define the notions of binding signature Σ and Σ-monoid, with the monoid part of the definition of Σ-monoid inherently using $\bullet$. Finally, they state and prove an initial algebra semantics theorem, the statement of the theorem again inherently involving $\bullet$. The heart of their paper is therefore a description and analysis of the substitution monoidal structure $\bullet$ on the category $[\mathbb{F}, Set]$. We summarise their paper in more detail in Section 2.

Contrast that with Gabbay and Pitts' paper. From a category-theoretic perspective, their paper is also convoluted but for somewhat different reasons. They have since modified their exposition in a way that is more convenient for us, so that is the formulation I shall use here [7,19]. Rather than study $[\mathbb{F}, Set]$, they study a category equivalent to $Pb(Inj, Set)$, where Inj is a skeleton of the category of finite sets and all injections, and $Pb(Inj, Set)$ is the full subcategory of $[Inj, Set]$ determined by those functors that preserve pullbacks. This is sometimes called the Schanuel topos, and it corresponds to the category of Fraenkel-Mostowski sets [7] and to the category Nom of nominal sets [19]. Gabbay and Pitts do not give any consideration whatsoever to a substitution monoidal structure on the category, and such a structure plays no role in their analysis. In contrast, the focus of their analysis is on the invariance of bound terms under renaming and upon the possibility of introducing a fresh name. Their analysis of binding per se is more subtle than that of Fiore et al, as the preservation of pullbacks allows them to give a sensible notion of the support of a bound term, a concept that cannot similarly be investigated in Fiore et al's setting. So Fiore et al's paper was fundamentally about substitution while Gabbay and Pitts' paper was not.

The research in both papers has undergone substantial development over the years since they were written. But that development has almost entirely been separate. Notable developments of Gabbay and Pitts' research have been application to the π-calculus [2] and the developments of nominal logic [19] (see [20] for a good recent introduction), $FreshML$ [25], and a nominal datatype package for $Isabelle/HOL$ [17,20,30] that provides effective access to the nominal research for the theorem-proving community. Meanwhile, Miki Tanaka and I have axiomatised Fiore et al's paper, extending it from variable substitution and binding to axiomatically defined substitution and binding [23,27,28,29], including examples such as linear substitution and binders [26], the mixed substitution and binders of the Logic of Bunched Implications [18,24], infinitary contexts [1], and typed substitution and binders, either cartesian [3,16] or otherwise.

The axiomatisation works as follows. One first chooses a pseudo-monad S on Cat to generate contexts. For example, Fiore et al's choice of cartesian contexts corresponds to the pseudo-monad T_{fp} on Cat for which the category Ps-T_{fp}-Alg of pseudo-algebras is the category of small categories with finite products. One then observes that the construction sending a small category $\mathcal{C}$ to the presheaf category $[\mathcal{C}^{op}, Set]$ may be characterised as the free colimit completion of $\mathcal{C}$. So, except for size, it amounts to giving another pseudo-monad T_{coc} on Cat, a pseudo-monad for cocomplete categories. Applying T_{fp} to the category 1, then applying T_{coc}, yields $[\mathbb{F}, Set]$ as studied by Fiore et al. Axiomatically, all the structure used by Fiore et al follows from their choice of the pseudo-monad $S = T_{fp}$ on Cat together with the canonical pseudo-distributive law, modulo size, of S over T_{coc} [28,29]: such structure suffices to yield a pseudo-monad structure on $T_{coc}S$ and hence a substitution monoidal structure on $T_{coc}S(1)$, with the definitions of binding signature Σ and Σ-monoid and the statement and proof of an initial algebra semantics theorem axiomatising that of Fiore et al flowing from there. The analysis can also be enriched and extended to incorporate types [23]. We summarise the axiomatisation in Section 3.

The work of Max Kelly on clubs deserves a mention here [11,12]. In the late 1960's, Kelly, in a concerted attempt to provide unified category-theoretic coherence results, developed a notion he called a *club*. The idea was to give a category-theoretic account of substitution. In that regard, he did not go as far as Fiore et al did, in the precise sense that he did not formulate or prove an initial algebra semantics theorem. But their general setting is a variant of his: where they considered $[\mathbb{F}, Set]$, or equivalently the category of discrete fibrations over $\mathbb{F}^{op}$, he considered $Cat/\mathbb{F}^{op}$: the former is a full subcategory of the latter, and Kelly's substitution monoidal structure restricts to Fiore et al's substitution monoidal structure. But, as remarked above, Fiore et al developed their account further.

A priori, it is not clear how to compare Fiore et al's approach to variable substitution with Gabbay and Pitt's study of invariance of bound terms under renaming. The two groups used different categories. And a canonical substitution monoidal structure was central to the first while not being studied at all by the second. But we can do the following: take the axiomatisation of substitution as above, modify it to allow for preservation of pullbacks of monomorphisms, and consider the special case where S is the pseudo-monad T_{sm1} for small symmetric monoidal categories for which the unit is the terminal object [28,29]. It follows that $(S1)^{op}$ is Inj, and, replacing T_{coc} by the (partial) pseudo-monad T_{poecoc} for free cocompletions that respect pushouts of epimorphisms, $T_{poecoc}S(1)$ is $Pb(Inj, Set)$ as studied by Gabbay and Pitts.

Thus we have a canonical substitution monoidal structure on Gabbay and Pitts' category of nominal sets, allowing Fiore et al's analysis to carry over. The condition asserting preservation of pullbacks of monomorphisms also makes sense in Fiore et al's setting, allowing them in principle to develop Gabbay and Pitts' notion of support. Our hope is that this new axiomatisation will provide a mathematical foundation on which one can make precise comparisons between the two approaches to abstract syntax. We outline this new axiomatisation in Section 4.

We should finally mention that although Gabbay and Pitts' initial research on the category of nominal sets did not include the substitution structures it might support, there has been subsequent relevant non-category-theoretic research by Gabbay and colleagues [5,8] that may relate to the ideas here.

2 An Overview of Fiore et al's Account of Variable Substitution and Binding

In this section, we briefly overview the main thread of argument in Fiore et al's paper [4].

Let $\mathbb{F}$ denote a skeleton of the category of finite sets and all functions between them. Fiore et al considered the functor category $[\mathbb{F}, Set]$. The idea was that for any object X of $[\mathbb{F}, Set]$, the set $X(n)$ is to be understood as a set of terms, modulo α-conversion, containing at most n variables. Fiore et al then, by fiat, described a substitution monoidal structure on $[\mathbb{F}, Set]$ as follows.

Theorem 2.1 *Given X and Y in $[\mathbb{F}, Set]$, and given m in $\mathbb{F}$, define $(X \bullet Y)m$ to be the coequaliser*

$$(X \bullet Y)m = (\coprod_{n \in N} Xn \times (Ym)^n)/\sim$$

where the equivalence relation $\sim$ is induced by $(t; u_1, \ldots, u_n) \sim (t'; u'_1, \ldots, u'_{n'})$ if and only if there exists an arrow $\rho : n \to n'$ such that $X(\rho)(t) = t'$ and $u_i = u'_{\rho i}$.

Then, the family of sets $(X \bullet Y)m$ extends canonically via the universality of its definition to give a monoidal structure on the category $[\mathbb{F}, Set]$.

Having described a substitution monoidal structure, Fiore et al defined a notion of binding signature. In the absence of binders, a signature would consist of a set of operations O together with a function $ar : O \longrightarrow \mathbb{N}$ sending each operation to an arity given by a natural number. But if one wants to allow for binders, one needs more sophistication in the arities as one wants not only a natural number but an account of the number of variables to be bound in each argument. So an arity should consist of a finite sequence of natural numbers. Fiore et al accordingly made the following definition.

Definition 2.2 A *binding signature* consists of a set O of operations together with a function $ar : O \longrightarrow \mathbb{N}^*$.

Example 2.3 Consider the untyped λ-calculus

$$M ::= x \mid \lambda x.M \mid MM$$

It has two operators, one for lambda and one for application, with arities $\langle 1 \rangle$, and $\langle 0, 0 \rangle$ respectively: λ-abstraction has one argument and binds one variable, and application has two arguments and binds no variables.

Given a binding signature Σ, Fiore et al generated what they called a signature endofunctor on $[\mathbb{F}, Set]$ as follows.

Definition 2.4 Given a binding signature Σ, the *signature endofunctor* generated by Σ sends an object X of $[\mathbb{F}, Set]$ to

$$\coprod_{\{o \epsilon O | ar(o) = (n_i)_{1 \le i \le k}\}} (\delta^{n_1} X) \times \cdots \times (\delta^{n_k} X)$$

where δX is defined to be $X(1 + -)$.

Example 2.3 shows how this is to be understood in practice: $\delta X = X(1 + -)$ gives a mathematical formulation of the idea of binding over one variable. More generally, the composite $\delta^n X$, which is therefore $X(n + -)$, allows the formulation of the idea of binding over n variables. In order to construct λ-terms, λ-abstraction is a unary operation that requires the use of one additional variable, hence its arity being $\langle 1 \rangle$, while application is a binary operation, with neither argument requiring any additional variables, hence its arity being $\langle 0, 0 \rangle$.

Fiore et al's central and motivating theorem was a characterisation of $T_\Sigma(1)$, where T_Σ is the free monad on $[\mathbb{F}, Set]$ generated by the binding signature Σ. They called the theorem the initial algebra semantics theorem as $T_\Sigma(1)$ describes the terms generated by Σ. The supporting results and definitions for the initial algebra semantics theorem, modulo a tiny correction, may be expressed as follows.

Theorem 2.5 *For any binding signature Σ, there is a canonical strength of the induced endofunctor, also denoted by Σ, over $\bullet$*

$$\Sigma X \bullet Y \longrightarrow \Sigma(X \bullet Y)$$

for pointed objects Y.

Corollary 2.6 *For any binding signature Σ, if T_Σ is the free monad generated by Σ on $[\mathbb{F}, Set]$, it follows that T_Σ has a canonical strength over pointed objects with respect to $\bullet$.*

Recall that a strength over pointed objects for any monad on a monoidal closed category yields a canonical monoid structure on the free algebra on 1 [27]. Thus we immediately have the following.

Corollary 2.7 *For any binding signature Σ, the object $T_\Sigma(1)$ of $[\mathbb{F}, Set]$ has a canonical monoid structure on it with respect to the substitution monoidal structure $\bullet$.*

Fiore et al inevitably needed a coherence condition to relate the monoid structure of $T_\Sigma(1)$ to its Σ-structure. That motivated the following definition.

Definition 2.8 Let F be a strong (over pointed objects) endofunctor on a monoidal closed category $(C, \cdot, I)$. An *F-monoid* (X, μ, ι, h) consists of a monoid (X, μ, ι) in C and an F-algebra (X, h) such that the diagram

$$\begin{tikzcd}
F(X) \cdot X \arrow[r, "t_{X,X}"] \arrow[d, "h \cdot X"'] & F(X \cdot X) \arrow[r, "F\mu"] & FX \arrow[d, "h"] \\
X \cdot X \arrow[rr, "\mu"'] & & X
\end{tikzcd}$$

commutes.

F-monoids form a category with maps given by maps in C that preserve both the F-algebra structure and the monoid structure.

Theorem 2.9 (Initial Algebra Semantics) *For any binding signature Σ, the object $T_\Sigma(1)$ of the category $[\mathbb{F}, Set]$ together with its canonical Σ-algebra structure and monoid structure with respect to $\bullet$, form the initial Σ-monoid.*

Proof. By Corollaries 2.6 and 2.7, the object $T_\Sigma(1)$ necessarily exists and has a canonical monoid structure on it. For general reasons that apply by virtue of freeness, the Σ-algebra and monoid structures on $T_\Sigma(1)$ form a Σ-monoid. The fact that $T_\Sigma(1)$ is the free Σ-algebra on 1 together with the fact that a monoid map must preserve the unit of a monoid, determine a unique map from $T_\Sigma(1)$ to any Σ-monoid. The assertion that it is a monoid map is then equivalent to the coherence condition in the definition of Σ-monoid. □

This is one of the two equivalent versions of the central and final result of Fiore et al's paper, exhibiting initial algebra semantics for a binding signature.

3 An Analysis of Fiore et al's Account of Variable Substitution and Binding

In Section 2, we summarised the main line of development of Fiore et al's paper. In this section, we summarise the work of Miki Tanaka and myself axiomatising Fiore et al's paper, extending the ideas from cartesian contexts to axiomatically defined contexts, including linear contexts, affine contexts, and the contexts of the Logic of Bunched Implications. The work has been published, notably in [28,29], so we do not spell out the further examples here, but merely explain how the axiomatic development extends that of Fiore et al.

For precise definitions of pseudo-monad, pseudo-algebra, etcetera, see [27]. The concepts are a little more complicated than those associated with 2-monads, which in turn are a little more complicated than those associated with ordinary monads, but only a little. Our two leading examples are as follows.

Example 3.1 Let T_{fp} denote the pseudo-monad on Cat for small categories with finite products. The 2-category $Ps\text{-}T_{fp}\text{-}Alg$ has objects given by small categories with finite products, maps given by functors that preserve finite products in the usual sense, i.e., up to coherent isomorphism, and 2-cells given by all natural transformations. So $Ps\text{-}T_{fp}\text{-}Alg$ is the 2-category FP and the category $T_{fp}(\mathcal{C})$ is the free category with finite products on C. Taking $\mathcal{C} = 1$, the category $T_{fp}(\mathcal{C})$ is given, up to equivalence, by $\mathbb{F}^{op}$.

Example 3.2 For size reasons, there is no interesting pseudo-monad on Cat for cocomplete categories: small cocomplete categories are necessarily preorders, and the free large cocomplete category on a small category does not lie in Cat. But there are well-studied techniques to deal with that concern [27], allowing us safely to ignore it here. Assuming we do that, there is a pseudo-monad T_{coc} for cocomplete categories. For any small category $\mathcal{C}$, the category $T_{coc}(\mathcal{C})$ is given by the presheaf category $[\mathcal{C}^{op}, Set]$.

For a precise definition of a pseudo-distributive law between pseudo-monads, see [27]. Again, they are a little more complicated than ordinary distributive laws, but only a little, the only substantial complexity residing in coherence detail, i.e., a statement of exactly which two-dimensional diagrams must agree with each other. Brushing that aside, our leading example is as follows [27,28,29].

Example 3.3 There is a canonical pseudo-distributive law of T_{fp} over T_{coc}. Observe that $T_{coc}T_{fp}(1)$ is equivalent to $[\mathbb{F}, Set]$.

An immediate consequence of the definition of pseudo-distributive law is the following [27,28,29].

Theorem 3.4 *Given a pseudo-distributive law $\delta : ST \longrightarrow TS$ of pseudo-monads on Cat, the pseudo-functor TS supports a canonical pseudo-monad structure and the category $TS(1)$ has both canonical pseudo-S-algebra and pseudo-T-algebra structures on it.*

For an arbitrary pseudo-monad T on Cat, let $t_{C,D}$ be the unCurrying of the composite

$$D \longrightarrow [C, C \times D] \xrightarrow{T} [TC, T(C \times D)]$$

We call t the pseudo-strength induced by the pseudo-monad T.

Theorem 3.5 *Given a pseudo-monad T on Cat, the category $T1$ has a canonical monoidal structure with multiplication $\bullet$ defined by using the pseudo-strength induced by T as follows:*

$$\bullet : T1 \times T1 \xrightarrow{t_{1,T1}} T(1 \times T1) \xrightarrow{\cong} T^2 1 \xrightarrow{\mu_1} T1$$

and with unit given by

$$\eta_1 : 1 \longrightarrow T1$$

The associativity and unit isomorphisms are generated by those for the multiplication and unit of T together with those of the pseudo-strength. Moreover, the multiplication $\bullet : T1 \times T1 \to T1$ is a pseudo-map of T-algebras in its first variable, i.e., there is a coherent isomorphism

$$\begin{tikzcd}
T^2 1 \times T1 \arrow[r, "t_{T1,T1}"] \arrow[d, "\mu \times id"'] & T(T1 \times T1) \arrow[r, "T\bullet"] \arrow[d, phantom, "\cong"] & T^2 1 \arrow[d, "\mu"] \\
T1 \times T1 \arrow[rr, "\bullet"'] & & T1
\end{tikzcd}$$

Example 3.6 Consider the pseudo-monad $T_{coc}T_{fp}$ on Cat. By Example 3.3, the category $T_{coc}T_{fp}(1)$ is equivalent to $[\mathbb{F}, Set]$. So, by Theorem 3.5, $[\mathbb{F}, Set]$ acquires a canonical monoidal structure. By the last line of the theorem, for every object Y of $[\mathbb{F}, Set]$, the functor $- \bullet Y : [\mathbb{F}, Set] \longrightarrow [\mathbb{F}, Set]$ is a pseudo-map of $T_{coc}T_{fp}$-algebras, and so preserves both colimits and finite products. Since every functor $X : F \longrightarrow Set$ is a colimit of representables, and every object of $\mathbb{F}^{op}$ is a finite product of copies of the generating object 1, which in turn is the unit of the tensor $\bullet$, it follows that we can calculate $X \bullet Y$ as a canonical coequaliser of the form

$$(X \bullet Y)m = (\coprod_{n \in N} Xn \times (Ym)^n)/\sim$$

yielding exactly Fiore et al's construction of a substitution monoidal structure.

With a little care, the argument in Example 3.6 can be extended to full axiomatic generality as follows [28,29].

For any pseudo-monad S on Cat, if (A, a) is (part of) an arbitrary pseudo-S-algebra, e.g., $S1$ or $[(S1)^{op}, Set]$, and α is an object of the category Sk for any small category k, in particular for any natural number, we may define a functor $\overline{\alpha}_A : A^k \to A$ as follows:

$$A^k \cong A^k \times 1 \xrightarrow{S \times \alpha} (SA)^{Sk} \times Sk \xrightarrow{ev_\alpha} SA \xrightarrow{a} A$$

This construction is a routine extension of the idea that every algebra for a (finitary) monad on Set supports a semantics for every operation of the Lawvere theory corresponding to the monad, which is exploited in the modelling of computational effects in [21].

Theorem 3.7 *Given a pseudo-monad S on Cat and a pseudo-distributive law of S over T_{coc}, and given X, Y in $[(S1)^{op}, Set]$, one can calculate the value of $X \bullet Y$ at $c \in S1$ as*

$$(X \bullet Y)c = \int^{c' \in S1} Xc' \times (\overline{c'}_{[(S1)^{op}, Set]}(Y))c \qquad (1)$$

We now turn from substitution to binding. An arity $(n_1, \cdots, n_k)$ in Fiore et al's definition of binding signature, Definition 2.2, yields two pieces of data: for each i, each n_i tells you how many times to apply $X(1+-)$, and k tells you how many such $X(n_i + -)$ need to be multiplied. But in more complex settings, more specificity is needed as a finite sequence of natural numbers does not specify which sort of binder is to be used, and in what combination are the binders to be used: Fiore et al used cartesian binders and took a product; Tanaka used linear binders and took a tensor product; but in Bunched Implications, one has a choice of binders and a choice of product or tensor. These considerations lead to the following general definition:

Definition 3.8 For a pseudo-monad S on Cat, a *binding signature* $\Sigma = (O, a)$ is a set of operations O together with an arity function $a : O \to Ar_S$ where an element $(k, \alpha, (n_i, \beta_i)_{1 \leq i \leq k})$ of Ar_S consists of a natural number k, an object α of the category Sk, and, for $1 \leq i \leq k$, a natural number n_i and an object β_i of the category $S(n_i + 1)$.

The k and the n_i's here agree with those of both Fiore et al, while the α and β_i's tell us which sorts of binders are to be used and how they are to be combined.

With the definition of binding signature in hand, we can induce a signature endofunctor, as Fiore et al did, then speak of algebras for the endofunctor. We shall overload the use of 1: it will refer to the unique object 1 of the category 1, and also to its image in $S1$ via the unit of S. When we write $\mathbf{1}$, we shall mean a list $1, \ldots, 1$ of length determined by the context in which we write. We further abbreviate an expression of the form $f(x_1, \ldots, x_n)$ to $f(\boldsymbol{x})$ and one of the form $f(x_1, \ldots, x_n, -)$ to $f(\boldsymbol{x}, -)$. Using these notational abbreviations, we define the induced signature endofunctor as follows.

Definition 3.9 Given a binding signature Σ, the *signature endofunctor* on $[(S1)^{op}, Set]$

sends X to

$$\Sigma X = \coprod_{\substack{o \in O \\ a(o)=(k,\alpha,(n_i,\beta_i))}} \overline{\alpha}_{[(S1)^{op},Set]}(X(\overline{\beta_1}_{S1}(\mathbf{1},-)),\ldots,X(\overline{\beta_k}_{S1}(\mathbf{1},-)))$$

The functor constructed in Definition 3.9 agrees with the Fiore et al's Definition 2.4 as we shall show in Example 3.10. Following Fiore et al, we overload notation by denoting both the signature and the functor it generates by Σ.

Example 3.10 Let S be T_{fp}, i.e., consider Fiore et al's cartesian binders. Our k is their k. Our α is the object $\mathbf{1}$ of $T_{fp}k$, which generates the functor

$$[\mathbb{F}, Set]^k \longrightarrow [\mathbb{F}, Set]$$

defining the k-fold product. Our n_i is their n_i. And our β_i is the object $\mathbf{1}$ of $T_{fp}(n_i+1)$, which generates the functor

$$\overline{\beta_i}_{\mathbb{F}} : \mathbb{F}^{n_i+1} \longrightarrow \mathbb{F}$$

that sends $(a_1, \cdots, a_{n_i}, b)$ to $a_1 + \ldots + a_{n_i} + b$. So every one of Fiore et al's binding signatures generates one of our binding signatures, and the endofunctor we define on $[\mathbb{F}, Set]$ agrees, when restricted to one of their binding signatures, with their construction.

For a specific example, recall Example 2.3, that of the untyped λ-calculus

$$M ::= x \mid \lambda x.M \mid \mathsf{app}(M, M)$$

with its two operators λ and app, with arities, in Fiore et al's terms, given by $\langle 1 \rangle$, and $\langle 0, 0 \rangle$, respectively. Let 2 be defined to have elements x and y. Then, in our terms, for the first operator λ, the arity is given by $k = 1$, $\alpha \in T_{fp}(1)$ is 1, n_1 is 1, and β_1 is the element $x \times y$ of $T_{fp}(2)$. And for the application app, $k = 2$, α is the element $x \times y$ of $T_{fp}(2)$, $n_0 = n_1 = 0$, and both β_i's are given by 1 seen as an element of $T_{fp}(1)$.

We should mention that not only are our binding signatures a priori more general than those of Fiore et al, but there seems to be one of our binding signatures for which there is none of Fiore et al's signatures with an equivalent category of algebras.

Example 3.11 Consider the signature in our sense consisting of one arity, with $k = 1$, with α being the generating object 1 of $T_{fp}1$, and with β_1 given by the pair $y \times y$ in the notation of Example 3.10. An algebra would consist of a presheaf X together with a natural transformation

$$X(2 \times -) \longrightarrow X(-)$$

which does not appear to be constructable as an algebra for any signature in the sense of Fiore et al: note that $y \times y$ generates $X(2 \times -)$ rather than $X(2 + -)$.

The signature in Example 3.11 does not seem to have computational significance. That does not unduly perturb us: our main theorem about signatures is a positive one, asserting that any signature yields initial algebra semantics, so including uninteresting examples within that result does not bother us.

From this point, one can *verbatim* extend Fiore et al's sequence of definitions and results leading to their initial algebra semantics theorem, with the sole exception of the need to replace $\mathbb{F}$ by $(S1)^{op}$. In generalising the proofs, the only difficult point is in generalising the proof of Theorem 2.5, i..e, the construction of the strength of Σ over $\bullet$ with respect to pointed objects: a proof appears in [29]. Summarising, we have the following.

Theorem 3.12 (Initial Algebra Semantics) *Given any pseudo-monad S on Cat together with a pseudo-distributive law of S over the (partial) pseudo-monad $T_{coc}- = [(-)^{op}, Set]$, and given any binding signature Σ, the object $T_\Sigma(1)$ of the category $[(S1)^{op}, Set]$ together with its canonical Σ-algebra structure and monoid structure with respect to $\bullet$, form the initial Σ-monoid.*

Examples include Fiore et al's Theorem 2.9 and examples for linear contexts [26] and for affine contexts and the contexts of Bunched Implications [29]. Moreover, the analysis extends to include both types and enrichment, thereby allowing for recursion [23].

4 Extending to Include Gabbay and Pitts

As mentioned in the introduction, Gabbay and Pitts, in [6,7], studied the invariance of bound terms under renaming rather than studying substitution, with their account of binding being more subtle than that of Fiore et al. Specifically, their restriction from $[Inj, Set]$ to the full subcategory of those functors that preserve pullbacks allowed them to speak sensibly of the *support* of a bound term. One can analyse the idea in terms of the axiomatisation of Section 3.

Example 4.1 Let $S = T_{sym1}$ be the pseudo-monad on Cat for small symmetric monoidal categories with a terminal object. So $(S1)^{op} = Inj$. It is shown in [29] that there is a canonical pseudo-distributive law of T_{sym1} over T_{coc}, and so Theorem 3.12 applies. So, given any binding signature Σ, one can consider the functor $T_\Sigma(1)$ from Inj to Set. Observe that this functor preserves pullbacks: an element of $T_\Sigma(1)n$ amounts to a term in an at most n variables, and the pullback condition asserts that to give terms s with variables among $x_1, \cdots, x_n, y_1, \cdots, y_m$ and t with variables among $y_1, \cdots, y_m, z_1, \cdots, z_k$, with the x's and z's distinct, such that s and t agree when seen as terms with variables among the union, is to give a term with variables among $y_1, \cdots y_m$. The pullback condition therefore allows Gabbay and Pitts axiomatically to speak of the *support* of a term, this amounting to the smallest set of variables that are required to express the term.

Example 4.1 suggests that it might be possible to find a common mathematical foundation for Fiore et al and Gabbay and Pitts' ideas if one could modify the axiomatisation of Section 3 in a way that replaces $[(S1)^{op}, Set]$ by $Pb((S1)^{op}, Set)$ for those examples in which the small category $S1$ has pushouts. Such a modification does seem possible. But there is a problem as follows:

Example 4.2 Let $S = T_{fp}$, i.e., consider Fiore et al's setting. Given a binding signature Σ, it is generally not the case that $T_\Sigma(1)$ preserves pullbacks. For the category $\mathbb{F}$ has a terminal object 1, and the pullback of the map from 2 to 1 along itself in $\mathbb{F}$ is 4. Preservation of that pullback would imply that to give a term with at most four variables is to give a pair of terms each with at most two variables subject to a coherence condition, which is false even in very simple examples.

The difficulty is that $\mathbb{F}$ has too many pullbacks, so asking for preservation of all pullbacks of $\mathbb{F}$ precludes essentially all interesting examples of binding signatures. So one needs to consider some pullbacks but not all pullbacks. A natural solution is to demand the preservation of pullbacks of pairs of monomorphisms: that agrees with Gabbay and Pitts, as all maps in Inj are monomorphisms, and in $[\mathbb{F}, Set]$, it is true of Fiore et al's T_Σ. In $[\mathbb{P}, Set]$, i.e., for Tanaka's study of linearity in [26], it yields all functors, which is fine because, in linear λ-calculus for example, there is no relation between the set of terms that use one free variable precisely once and the set of terms that use two free variables precisely once each. The situation for bunches is more complex but similar.

So we propose to modify the analysis of Section 3 as follows.

Definition 4.3 Let PO_e denote the 2-category of small categories with pushouts of epimorphisms and functors that preserve such pushouts, hence preserving all epimorphisms. Given an object $\mathcal{C}$ of PO_e, let $Pb_m(\mathcal{C}^{op}, Set)$ denote the full subcategory of $[\mathcal{C}^{op}, Set]$ determined by those functors that preserve pullbacks of monomorphisms.

Proposition 4.4 *The 2-category PO_e is a cartesian pseudo-closed 2-category [10] and thus supports a notion of PO_e-enriched category, cf [23].*

Proposition 4.5 *One can lift the various pseudo-monads for contexts of [28,29], notably T_{fp}, from Cat to PO_e, and the liftings have canonical PO_e-enrichments or equivalently pseudo-strengths.*

Observe that the construction T_{poecoc} sending a small category $\mathcal{C}$ with pushouts of epimorphisms to the free cocompletion that respects the pushouts of epimorphisms in $\mathcal{C}$, namely $Pb_m(\mathcal{C}, Set)$ [14], extends canonically to a PO_e-enriched pseudo-monad on PO_e.

Theorem 4.6 *The category $T_{poecoc}S(1)$ with pushouts of epimorphisms has a canonical substitution monoidal structure on it satisfying the coherence condition of Theorem 3.5.*

Theorem 4.7 *The definitions of binding signature Σ and Σ-monoid, and the statement and proof of Theorem 3.12, i.e., the initial algebra semantics theorem, extend to the setting of PO_e.*

The details of the above rely upon somewhat delicate but straightforward 2-categorical definitions that are not yet in the literature. For explicit descriptions of the various constructs, sums and filtered colimits are pointwise in $Pb_m(\mathcal{C}, Set)$ for any small category $\mathcal{C}$ with pushouts of epimorphisms, which is quite special to the particular class of limits under consideration. So explicit descriptions of the above constructs are not difficult. The results now put us in a position to make precise comparison between Fiore et al's account of substitution and Gabbay and Pitts' account of invariance of binding under renaming. Such comparison may or may not relate to the recent non-category theoretic research on substitution of Gabbay and colleagues [5,8].

5 Further Work

The most obvious item of further work has been the focus of the paper, so I shall not repeat it here.

But beyond that, this paper has been directed specifically towards giving a category theoretic account of substitution and binding. Although the notions of substitution and binder are syntactic, in this paper, we have not addressed the question of giving a general syntax at all. So an obvious issue for further work is to provide such a syntax that includes Fiore et al's cartesian binders, Tanaka's linear binders, and the mixed variable binders of the Logic of Bunched Implication, preferably consistently with the developments of Gabbay and Pitts' work. It seems most unlikely that there is a natural syntax to be found that corresponds to the full generality of this paper. Nevertheless, a general syntax surely should exist with category theoretic models given herein. The notion of a pseudo-commutative monad [10] may be relevant.

Hofmann also studied logical principles on binding structures in [9], which was a third paper about binding presented at *LICS 99*. Accordingly, one would hope to incorporate logical principles such as induction over higher-order terms into the axiomatics here.

References

[1] S. J. Ambler, R. L. Crole, and Alberto Momigliano. A Definitional Approach to Primitive Recursion over Higher Order Abstract Syntax. In *Proc. MERLIN 2003, ACM Digital Library*, 2003.

[2] James Cheney. Toward a General Theory of Names, Binding and Scope. In *Proc. MERLIN 2005, ACM Digital Library*, 2005.

[3] M. Fiore, Semantic Analysis of Normalisation by Evaluation for Typed Lambda Calculus. In *Proc. PPDP 02*, ACM Press, pages 26–37, 2002.

[4] M. Fiore, G. Plotkin, and D. Turi. Abstract Syntax and Variable Binding. In *Proc. LICS 99*, pages 193–202. IEEE Press, 1999.

[5] M. Gabbay and A. Mathijssen. Capture-avoiding Substitution as a Nominal Algebra. In *Proc. ICTAC 2006, LNCS 4281*, pages 198–212, 2006.

[6] M. Gabbay and A. M. Pitts. A New Approach to Abstract Syntax Involving Binders. In *Proc. LICS 99*, pages 214–224. IEEE Press, 1999.

[7] M. Gabbay and A. M. Pitts. A New Approach to Abstract Syntax with Variable Binding. *Formal Aspects of Computing* 13:341–363, 2002.

[8] M. Gabbay, S. Rota Buló, and A. Marin. Substitution as a property of Fraenkel-Mostowski sets. Submitted.

[9] M. Hofmann, Semantical Analysis of Higher-Order Abstract Syntax. In *Proc. LICS 99*, IEEE Press, pages 204–213, 1999.

[10] M. Hyland and A. J. Power. Pseudo-Commutative Monads and Pseudo-Closed 2-categories. *J. Pure and Applied Algebra* 175:141–185, 2002.

[11] G. M. Kelly. An abstract approach to coherence. In *Coherence in categories, LNM 281*, pages 106–147, 1972.

[12] G. M. Kelly. On Clubs and Doctrines. In *Category Seminar Sydney 1972/73, LNM 420*, pages 181–256, 1974.

[13] G. M. Kelly. A Unified Treatment of Transfinite Constructions for free algebras, free monoids, colimits, associated sheaves, and so on. *Bull. Austral. Math. Soc*, 22:1–83, 1980.

[14] G. M. Kelly. *Basic Concepts of Enriched Category Theory, London Math. Soc. Lecture Notes Series 64*, Cambridge University Press, 1982.

[15] M. Lenisa, A. J. Power, and H. Watanabe. Category Theory for Operational Semantics. *Theoretical Computer Science* 327:135–154, 2004.

[16] M. Miculan and I. Scagnetto, A Framework for Typed HOAS and Semantics. In *Proc. PPDP 2003, ACM Press*, pages 184–194, 2003.

[17] M. Norrish. Recursive Function Definition for Types with Binders. In *Theorem Proving in Higher Order Logics 2004, LNCS 3223*, pages 241–256, 2004.

[18] P. O'Hearn, *On Bunched Typing*, *Journal of Functional Programming*, 13:747-796, Cambridge University Press, 2003.

[19] A. M. Pitts. Nominal logic: a First Order Theory of Names and Binding. *Information and Computation* 183:165–193, 2003.

[20] A. M. Pitts. Alpha-Structural Recursion and Induction *Journal of the ACM* 53:459–506, 2006.

[21] G. D. Plotkin and A. J. Power. Algebraic Operations and Generic Effects. In *Proc. MFCSIT 2000, Applied Categorical Structures* 11:69–94, 2003.

[22] A. J. Power. A Unified Category-Theoretic Approach to Variable Binding. In *Proc. MERLIN 2003, ACM Digital Library*, 2003.

[23] A. J. Power and M. Tanaka. Category Theoretic Semantics for Typed Binding Signatures with Recursion. Submitted.

[24] D. Pym. *The Semantics and Proof Theory of the Logic of Bunched Implications, Applied Logic Series.* Kluwer, 2002.

[25] M. R. Shinwell, A. M. Pitts and M. J. Gabbay. Fresh ML: Programming with Binders Made Simple. In *Proc. ICFP 2003, ACM Press*, pages 263–274, 2003.

[26] M. Tanaka. Abstract Syntax and Variable Binding for Linear Binders. In *Proc. MFCS 2000, LNCS 1893*, pages 670–679, 2000.

[27] M. Tanaka. Pseudo-Distributive Laws and a Unified Framework for Variable Binding. Edinburgh Ph.D. thesis, 2004.

[28] M. Tanaka and A. J. Power. Pseudo-Distributive Laws and Axiomatics for Variable Binding. *Higher-Order and Symbolic Computation* 19:305–337, 2006.

[29] M. Tanaka and A. J. Power. A Unified Category-Theoretic Semantics for Binding Signatures in Substructural Logics. *Journal of Logic and Computation*, 16:5–25 , 2006.

[30] C. Urban and S. Berghofer. A Recursion Combinator for Nominal Datatypes Implemented in Isabelle/HOL. In *Proc. IJCAR 2006, LNCS 4130*, pages 498–512, 2006.