A note on proof interpretations and Dialectica categories

Thomas Powell

written early 2010

Introduction

In most applications of functional interpretations, the interpretation is a means to an end, a syntactic translation that extracts witnesses from proofs. As a consequence, on the whole proof theorists pay little attention to the structural properties of functional interpretations, though the use of interpretations is central to their work.

This is a short note in which we discuss and bring together several works which view functional interpretations from a more abstract perspective. Our ultimate aim is to construct a general abstract framework in which a range of interpretations can be compared and better understood.

A rich variety of functional interpretations have been developed since Gödel invented his prototype, ranging from early examples used to prove foundational theorems to more exotic modern varieties tailored specifically for the purpose of proof mining. It is natural, then, to ask whether we can isolate the key features of functional interpretations and develop a unifying framework in which they can be compared, either on a syntactic or a semantic level.

The question of unifying proof interpretations has been separately considered from each of these perspectives, by Oliva and de Paiva respectively. de Paiva used the language of categorical logic to gain a better semantic understanding of the Dialectica interpretation - constructing and studying the *Dialectica category* [7]. This yielded some interesting results, notably that the Dialectica interpretation itself behaves rather badly - and that the best that can be achieved in terms of a categorical semantics is a model of linear logic. However, an interpretation of the linear modality ! via a comonad on the category produced an elegant model of a variant of the Dialectica interpretation - the Diller-Nahm interpretation.

Over a decade later, in his work on unifying functional interpretations [5], Oliva introduced, on a syntactic level, a parametrised functional interpretation with a uniform soundness proof, from which a large family of familiar interpretations could be retrieved.

This note attempts to combine these ideas in the construction of a uniform semantic framework for functional interpretations, based on de Paiva's Dialectica category. Studying interpretations in this way yields insights into their structure that may appear hidden in a more syntactic presentation. The idea is that many different interpretations can be modelled in an abstract way via comands on the Dialectica category. While this has been observed before, by Biering in [1] for instance, we show that these comonads arise in a uniform way from monads on the underlying type theory. In this respect our approach differs from de Paiva's original construction, in particular we emphasise the proof theoretic meaning of our categorical constructions.

Variations on the Dialectica interpretation

A key feature of a functional interpretation is the way in which it interprets implication. Different choices can result in interpretations with very different structural properties. Recall that the Dialectica interpretation interprets

$$\exists u \forall x A(u,x) \to \exists v \forall y B(v,y)$$

as

$$\exists f, F \forall u, y (A(u, Fuy) \to B(fu, y))$$

This defines a fairly strong interpretation, which is why it is so useful for extracting computational information from proofs. The cost, however, is that in a wider sense it is fairly restrictive in terms of the theories it admits. As previously observed, in order to verify even the basic logical axioms it demands that formulas in the interpreted theory are decidable. While arithmetic is simple enough to ensure this property, this is not the case for more complex systems (anything involving sets, for instance, because the membership predicate $\in$ is not decidable).

However, by adjusting the way we interpret implication, this requirement can be bypassed and we obtain functional interpretations that can be applied to a wide range of theories. A familiar example of such an interpretation is Kreisel's modified realizability. Here a witness for implication is simply a tuple of functionals that takes a witness for the premise and produces a witness for the conclusion. In other words, we interpret implication as:

$$\exists f \forall u, y (\forall x A(u,x) \to B(fu, y))$$

Of course the interpreting theory is no longer a quantifier free calculus, it contains instances of the universal quantifier, hence modified realizability is a weaker interpretation with regard to the principles it will admit. In particular, it does not interpret Markov's principle - Kreisel used this fact to prove that Markov's principle is underivable is subsystems of intuitionistic analysis.

A more refined interpretation for the purpose of eliminating the need for decidability was proposed by Diller and Nahm. It stems from the observation that in a natural deduction proof, only finitely many instances of the premise of an implication are used. Given that

$$\forall x A(x) \to B$$

holds, then irrespective of decidability of formulas we can extract a finite set X such that

$$\forall x \in X A(x) \to B$$

Motivated by this, the Diller-Nahm interpretation translates $\exists u \forall x A(u,x) \to \exists v \forall y B(v,y)$ as

$$\exists f, F \forall u, y (\forall x \in Fuy A(u,x) \to B(fu, y))$$

where Fuy is a finite set. The interpreting theory in this case can be understood as a language of functionals of finite type, much like Gödel's system T, in which the underlying type theory is enriched with 'finite set types' X^* for each type X, with an associated quantifier.

In his work on unifying functional interpretations, Oliva, observing the similarities between different interpretations, defines a more general interpretation in which the verifying system of functionals is enriched with an unspecified *bounded quantifier* $\forall x \sqsubset t A(x)$, and implication is interpreted as

$$(A \to B)^I :\equiv \exists f, F \forall u, y (\forall x \sqsubset Fuy A_I(u,x) \to B_I(fu, y))$$

He identifies sufficient conditions under which the generalised quantifier produces a sound interpretation, and shows that a large family of interpretations can be obtained via this general framework. Aside from those mentioned above, instantiations of this parametrised interpretation include Stein's interpretations and the recent monotone interpretation, which will be discussed later.

When presented syntactically, one aspect of functional interpretations that remains hidden is the difference between certain variants on a structural level. Interpretations like the Diller-Nahm are far more flexible than the original Dialectica, not only because they are more widely applicable, but because key interpreting functionals are *canonical*. The interpretation of contraction is much more benign for the Diller-Nahm variant - since we allow quantification over finite sets in our verifying system, contraction is satisfied by functionals $\lambda u.(u,u)$ and $\lambda u x_1 x_2.\{x_1\} \cup \{x_2\}$, since

$$\forall u, x_1, x_2 (\forall x \in \{x_1\} \cup \{x_2\} A_\wedge(u,x) \to A_\wedge(u,x_1) \wedge A_\wedge(u,x_2))$$

is valid. Comparing this to the Dialectica interpretation above, it is evident that the Diller-Nahm should be better behaved. This contrast in structural behaviour becomes clear when we use the language of category theory.

The Dialectica Category

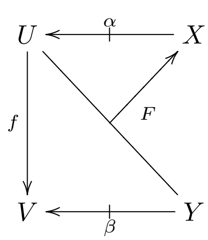
From the Curry-Howard correspondence has emerged the area of categorical logic, in which a rich semantics for type theories stems from the idea of representing propositions as objects in a category. By associating logical connectives with constructions in the category, beginning with an interpretation of atomic propositions we obtain interpretations of general propositions by induction on their structure. Deductions $\Gamma \vdash A$ in the type theory correspond to morphisms $|\Gamma| \to |A|$. In her thesis [7], de Paiva investigated the Dialectica interpretation in this manner, constructing the *Dialectica category*. The idea of the Dialectica category is to capture the interpretation's treatment of implication internally.

In this section we largely follow the more general presentation given in [4].

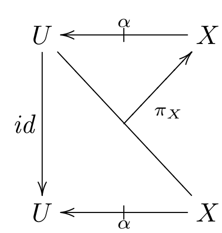
Definition 0.1 (Dialectica Category)**.** *Suppose we are given a category $\mathbb{T}$, which intuitively interprets some type theory, and a pre-ordered fibration $p \colon \mathbb{P} \to \mathbb{T}$ which interprets formulae over the type theory. We construct the Dialectica category **Dial** as follows.*

- *Objects of **Dial** consist of formulas $\alpha \in \mathbb{P}(U \times X)$ over types $U, X \in \mathbb{T}$ (which we abbreviate to $(U \overset{\alpha}{\leftarrow} X)$). Intuitively we read these formulas as $\exists u \forall x \alpha(u,x)$.*

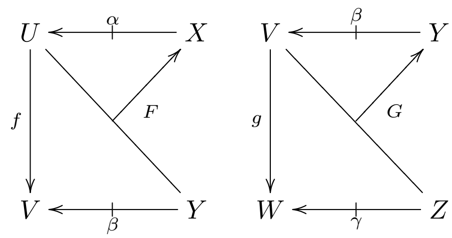
- *Morphisms $(U \overset{\alpha}{\leftarrow} X) \to (V \overset{\beta}{\leftarrow} Y)$ consists of pairs of maps $f \colon U \to V$, $F \colon U \times Y \to X$ of $\mathbb{T}$ such that $\alpha(u, F(u,y)) \leqslant \beta(f(u), y)$ in $\mathbb{P}(U \times Y)$, which we express as diagrams*



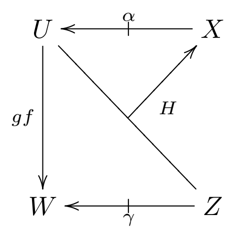
The identity map is given by



and composition of



is given by



*where $H(u,z) := F(u, G(f(u), z))$. It is easy to check that **Dial** is indeed a category.*

Of course, in order to do something useful with **Dial** we need some additional structure. The structure of **Dial** is thoroughly investigated in [7] and [4], so here we give just a brief summary of the main results. The essence of the next proposition is that provided the base category $\mathbb{T}$ and the fibration p have structure we would expect from a model of propositions over a type theory, the Dialectica category has structure we might expect from a model of a functional interpretation.

Proposition 0.2 (Structure of the Dialectica category)**.** *Suppose that $p\colon \mathbb{P} \to \mathbb{T}$ is a cartesian closed fibration, and that in addition, $\mathbb{T}$ has finite coproducts, that $\mathbb{P}(0) \cong 1$ and that the coprojections $X \to X+Y$ and $Y \to X+Y$ induce an equivalence $\mathbb{P}(X+Y) \simeq \mathbb{P}(X) \times \mathbb{P}(Y)$. Then*

- ***Dial** is a symmetric monoidal closed category. The tensor product $A \otimes B$ of $A = (U \overset{\alpha}{\leftarrow} X)$ and $B = (V \overset{\beta}{\leftarrow} Y)$ is the object $A \otimes B = (U \times V \overset{\alpha\otimes\beta}{\leftarrow} X \times Y)$ intuitively given by the pointwise relation $\alpha \otimes \beta(u,v,x,y)$ iff $\alpha(u,x) \wedge \beta(v,y)$. The unit is the object $\iota = (1 \leftarrow 1)$*

  The function space of objects $B = (V \overset{\beta}{\leftarrow} Y)$ and $C = (W \overset{\gamma}{\leftarrow} Z)$ is given by the obvious object $[B,C] = (W^V \times Y^{V \times Z} \overset{[\beta,\gamma]}{\leftarrow} V \times Z)$.

  *The adjunction **Dial**$(A \otimes B, C) \cong$ **Dial**$(A, [B,C])$ is easily verified.*

- ***Dial** has finite products, where the product of A and B in **Dial** is given by $A \times B = (U \times V \overset{\alpha\times\beta}{\leftarrow} X + Y)$, where intuitively $(u,v)\alpha \times \beta z$ iff $u\alpha z$ or $v\beta z$ (for a precise categorical definition see [7]). The terminal object is given by $(1 \leftarrow 0)$.*

Unfortunately, the Dialectica category does not have a natural cartesian closed structure - this boils down to the fact that the Dialectica interpretation requires 'definition by cases' functionals in order to model intuitionistic logic, and as might be expected, these functionals are not in the least bit natural from a categorical perspective. Assuming that objects in the category are decidable the product and tensor product collapse into one in the poset reflection and we obtain cartesian closed structure on the level of provability, but this simply reflects the fact that the Dialectica interpretation is sound. Any interesting categorical semantics will have a non-trivial notion of equality on proofs. Here we have confirmation that the Dialectica interpretation behaves in a rather peculiar way on a structural level, in that its underlying logic is essentially linear.

Proposition 0.3. *A model of the modality-free multiplicative fragment of intuitionistic linear logic consists of a symmetric monoidal closed category.*

Of course, full linear logic contains the modality ! that allows us to regain intuitionistic logic (see [3]). It's defining axioms are the following:

$$\frac{\Gamma \vdash B}{\Gamma, !A \vdash B}\ (wkn) \qquad \frac{\Gamma, !A, !A \vdash B}{\Gamma, !A \vdash B}\ (ctr) \qquad \frac{\Gamma, A \vdash B}{\Gamma, !A \vdash B}\ (der) \qquad \frac{!\Gamma \vdash A}{!\Gamma \vdash !A}\ (!)$$

The categorical semantics of ! are not obvious. However, a number of categorical models have been proposed - the one of interest here is that of Seely [8]:

Proposition 0.4. *A model of the multiplicative-exponential fragment of intuitionistic linear logic consists of a symmetric monoidal closed category with finite products, together with a comonad* ! *such that*

1. *for each object* A, $!A$ *has a comonoid structure* $(!A, \epsilon_A \colon !A \to I, \delta_A \colon !A \to !A \otimes !A)$ *with respect to the tensor product*
2. *there exist natural isomorphisms* $!A \otimes !B \cong !(A \times B)$, $I \cong !1$ *and moreover* ! *takes the comonoid structure* $(A, A \to 1, \triangle \colon A \to A \times A)$ *to the comonoid structure in (1).*

Proof. For details see [8]. Weakening and contraction are admitted by the existence of morphisms $!A \to I$ and $!A \to !A \otimes !A$ [make clear contexts etc.]. The remaining rules are given by the fact that ! is a comonad, and (!) in particular utilises the isomorphisms.

The following statement reflects the fact that we regain intuitionistic logic by taking implication to be the linear formula $!A \to B$.

Proposition 0.5. *Given a categorical model* $(\mathbb{C}, !)$ *of linear logic in the above sense, its Kleisli category* $\mathbb{C}_!$ *is cartesian closed.*

Proof. It is a general fact that products in $\mathbb{C}_!$ can be lifted from products in $\mathbb{C}$ [function spaces] The isomorphism $!A \otimes !B \cong !(A \times B)$ in $\mathbb{C}$ induces the following series of equivalences:

$$\mathbb{C}_!(A \times B, C) \cong \mathbb{C}(!(A \times B), C) \cong \mathbb{C}(!A \otimes !B, C)$$

$$\cong \mathbb{C}(!A, [!B, C]_{\mathbb{C}}) \cong \mathbb{C}_!(A, [!B, C]_{\mathbb{C}}) \cong \mathbb{C}_!(A, [B, C]_{\mathbb{C}_!})$$

which verifies that the function space $[B, -]$ in $\mathbb{C}_!$ is right adjoint to $(-) \times B$.

Having shown that **Dial** models modality free linear logic, de Paiva went about constructing a comonad in the above sense and obtained a Dialectica model of full linear logic, and hence of intuitionistic logic.

The key feature of her comonad was that it could be related back to the world of functional interpretations. Intuitively, the comonad sent the relation $\alpha(u, x)$ to the relation $!\alpha(u, \mathbf{x}) := \forall x \in \mathbf{x}\alpha(u, x)$, where $\mathbf{x}$ is some finite sequence of elements of type X.

Her model of intuitionistic implication $!A \to B$ became

$$\forall x \in Fuy\alpha(u, x) \to \beta(fu, y)$$

which is precisely that of the Diller-Nahm interpretation, and so the Kleisli category $\mathbf{Dial}_!$ provides a categorical semantics of the Diller-Nahm interpretation at the level of proofs.

Thus de Paiva demonstrated that the procedure of enriching Gödel's system T with quantification over finite sets in order to obtain the Diller-Nahm variant of the Dialectica interpretation is analogous to constructing a certain comonad on her Dialectica category. Moreover, she demonstrated that a meaningful categorical semantics exists for the Diller-Nahm interpretation, which indicates that some interpretations have good structural properties not present in the Dialectica interpretation.

In what follows we extend de Paiva's idea to deal with a range of familiar functional interpretations by implementing a range of comonads. Our construction of comonads follows a more general route - in particular it takes us closer to the underlying link between the Dialectica category and functional interpretations. Our idea is to formulate an abstract notion of a *bounded set* in the base category $\mathbb{T}$ and show that this structure gives rise, uniformly, to a comonad on **Dial** that corresponds to Oliva's generalised quantifier $\forall x \sqsubset t$.

Comonads in the Dialectica Category

The Dialectica interpretation maps formulas α to formulas of form $\exists u \forall x \alpha_D(u,x)$ where $\alpha_D(u,x)$ is quantifier free. Under Oliva's parametrised interpretation formulas are sent to $\exists u \forall x \alpha_{\sqsubset}(u,x)$ where, due to the interpretation of implication, $\alpha_{\sqsubset}(u,x)$ may contain instances of the generalised bounded quantifier $\forall x \sqsubset t$. Our aim here is to enrich the base category so that we can construct formulas over what are intuitively 'bounded set types', and add enough structure to the fibration $p\colon \mathbb{P} \to \mathbb{T}$ so that we can model formulas $\forall x \sqsubset tA(x)$ as objects in the Dialectica category.

We make the observation that the generalised quantifier $\forall x \sqsubset t$ acts only as an abbreviation in [6], actual instantiations vary significantly on a syntactic level. It has a clear semantic meaning, though, as membership of some bounded set, and our semantics identifies those terms that may differ on a syntactic level but or which the abbreviation $\forall x \sqsubset t$ means essentially the same thing.

Oliva states three conditions that must be satisfied by the abbreviation $\forall x \sqsubset tA(x)$ for it to behave like a bounded quantifier. The first states that for any $\sqsubset$-bounded formulas α, there must exist a term b_1 such that

$$\forall x^\rho \sqsubset b_1 x_0 \alpha(x) \to \alpha(x_0)$$

Suppose that $T\rho$ intuitively represents a collection of bounded sets of type ρ, so we can tentatively write $\forall x^\rho \sqsubset b_1 x_0^{T\rho}$, then the meaning of b_1 is a map $\rho \to T\rho$ that sends each x to a canonical bounded set $b_1 x$ containing x.

The second condition is that there must exist a term b_2 such that

$$\forall x^\rho \sqsubset b_2 x_0 x_1 \alpha(x) \to \forall x \sqsubset x_i \alpha(x)$$

for $i = 0, 1$. This time, we can view b_2 as a map $T\rho \times T\rho \to T\rho$ that essentially acts as a union operation. The final condition is the existence of a term b_3 such that

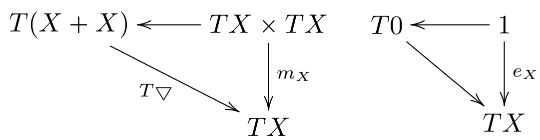
$$\forall x^\rho \sqsubset b_3 st\alpha(x) \to \forall z^\tau \sqsubset t \forall x \sqsubset sz\alpha(x)$$

In other words, if z is bounded by t then any element bounded by sz is also bounded by $b_3 st$. b_3 has type $(\tau \to T\rho) \to (T\tau \to T\rho)$. If we assume that T has the property of functoriality this condition reduces to the requirement of a map $b_3'\colon TT\rho \to Tp$. This all suggests that the structure we want to attach to our underlying type theory is a *monad* sending objects X to bounded set objects TX, along with a union operation $TX \times TX \to TX$.

However, in addition to the conditions above, the abbreviation $\forall x \sqsubset t$ must satisfy a handful additional assumptions in order to ensure it behaves like a quantifier in general, and in an analogous fashion we ask for a more robust categorical framework, namely a *strong* monad T - that is a monad $(T, \mu\colon TT \Rightarrow T, \eta\colon 1 \Rightarrow T)$ equipped with a natural transformation $C_{U,X} : U \times TX \to T(U \times X)$ obeying the relevant coherence conditions - and in addition natural isomorphisms $T(X+Y) \cong TX \times TY$ and $T0 \cong 1$ (induced by the coprojections $X \to X+Y$, $Y \to X+Y$). The necessity of this additional structure becomes evident in what follows. From this we obtain a union operation that coexists sensibly with the monad:

Lemma 0.6. *Suppose we are given cartesian closed category $\mathbb{T}$ with finite coproducts and a monad T on $\mathbb{T}$ equipped with isomorphisms $T(X+Y) \cong TX \times TY$ and $T0 \cong 1$. Then for any object X of $\mathbb{T}$, the object TX has a monoidal structure $(TX, e_X : 1 \to TX, m_X : TX \times TX \to TX)$ induced by the monoidal structure $(X, 0 \to X, \triangledown : X + X \to X)$.*

Proof. Define e_X, m_X via the following diagrams.



Having introduced some structure on $\mathbb{T}$ that intuitively captures the notion of a bounded set, we want to extend this to the fibration $p : \mathbb{P} \to \mathbb{T}$ and capture the notion of a bounded quantifier. Our aim is that formulas $\forall x \sqsubset \mathbf{x}\alpha(x)$ can be seen as objects of the fibre $\mathbb{P}(TX)$.

Definition 0.7. *The subfibre $\mathbb{P}_\star(TX) \subseteq \mathbb{P}(TX)$ consists of formulas that satisfy $\alpha(\mathbf{x_1} \cup \mathbf{x_2}) \to \alpha(\mathbf{x_1}) \wedge \alpha(\mathbf{x_2})$ for $\mathbf{x_1}, \mathbf{x_2} \in TX$.*

Objects of $\mathbb{P}(TX)$ are just arbitrary formulas that range over bounded sets of elements of type X. However, we argue that formulas in the subfibre $\mathbb{P}_\star(TX)$, namely those that are *closed under union*, are precisely those that intuitively have the form $\forall x \sqsubset \mathbf{x}\alpha(x)$ for some $\alpha \in \mathbb{P}(X)$. The way to capture this intuition is to demand that the map $\mathbb{P}_\star(TX) \to \mathbb{P}(X)$ induced by reindexing along the unit map η_X has a right adjoint $!_X$, i.e.

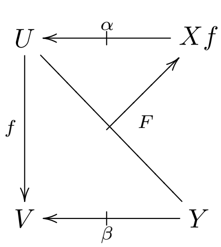
$$\mathbb{P}_\star(TX) \longrightarrow \mathbb{P}(TX) \xrightarrow{\eta_X} \mathbb{P}(X) \quad (\text{with a right adjoint } \mathbb{P}(X) \to \mathbb{P}_\star(TX),\ \perp)$$

The idea is that $\forall x \sqsubset \mathbf{x}\alpha(x) := !\alpha(\mathbf{x})$. Let $!_{U,X} : \mathbb{P}(U \times X) \to \mathbb{P}_\star(U \times TX)$ denote the U-indexed extension of $!_X$

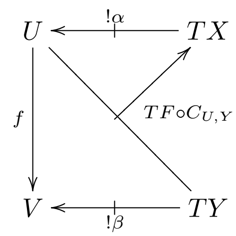
Proposition 0.8 (Main Result)**.** *The map $! : \textbf{Dial} \to \textbf{Dial}$ sending objects $\alpha \in \mathbb{P}(U \times X)$ to $!_{U,X}(\alpha) \in \mathbb{P}_\star(U \times TX)$ extends to a comonad on **Dial** such that*

1. *for each object A, $!A$ has a comonoid structure $(!A, \epsilon_A : !A \to I, \delta_A : !A \to !A \otimes !A)$ with respect to the tensor product*
2. *there exist natural isomorphisms $!A \otimes !B \cong !(A \times B)$, $I \cong !1$ and moreover $!$ takes the comonoid structure $(A, A \to 1, \triangle : A \to A \times A)$ to the comonoid structure in (1).*

Proof. First we show that ! extends to a functor. It acts on morphisms by sending



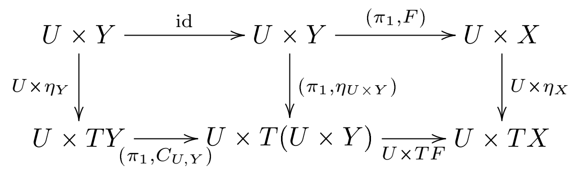
to



and it remains to show that this is well defined, i.e.

$$(\pi_1, TF \circ C_{U,Y})^*(!\alpha) \leqslant (f \times \mathrm{id})^*(!\beta)$$

The following diagram commutes,



the left hand side by the properties of strength and the right hand side by naturality of η.

Notice that

$$((U \times \eta_X) \circ (\pi_1, F))^*(!\alpha) \simeq (\pi_1, F)^* \circ (\pi_1, \eta_X)^*(!\alpha)$$

We have $(\pi_1, \eta_X)^*(!\alpha) \leqslant \alpha$ by definition of !, and hence

$$(\pi_1, F)^* \circ (\pi_1, \eta_X)^*(!\alpha) \leqslant (\pi_1, F)^*(\alpha)$$

Also,

$$((U \times TF) \circ (\pi_1, C_{U,Y}) \circ (U \times \eta_Y))^*(!\alpha) \simeq (U \times \eta_Y)^* \circ (\pi_1, TF \circ C_{U,Y})^*(!\alpha)$$

By commutativity of the diagram we obtain

$$(U \times \eta_Y)^* \circ (\pi_1, TF \circ C_{U,Y})^*(!\alpha) \leqslant (\pi_1, F)^*(\alpha)$$

[because $\mathbb{P}_\star()$ preserved under $C_{U,Y}$ and TF] we know that

$$(U \times \eta_Y)^* \circ (\pi_1, TF \circ C_{U,Y})^*(!\alpha) \in \mathbb{P}_\star(U \times TY)$$

therefore using the adjunction

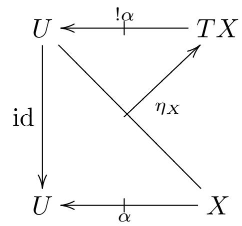
$$(\pi_1, TF \circ C_{U,Y})^*(!\alpha) \leqslant !(\pi_1, F)^*(\alpha) \leqslant !(f \times \mathrm{id})^*(\beta)$$

where the second inequality comes from the hypothesis $(\pi_1, F)^*(\alpha) \leqslant (f \times \mathrm{id})^*(\beta)$ But [naturality in first component] $!(f \times \mathrm{id})^*(\beta) \simeq (f \times \mathrm{id})^*(!\beta)$, and hence

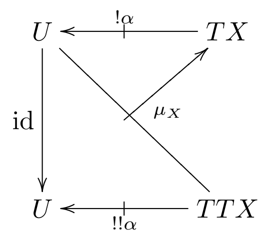
$$(\pi_1, TF \circ C_{U,Y})^*(!\alpha) \leqslant (f \times \mathrm{id})^*(!\beta)$$

which confirms that ! is well defined on morphisms. Functorality follows by properties of strength.

To show that ! defines a comonad on **Dial** we need to exhibit natural transformations $! \Rightarrow 1$ and $!! \Rightarrow !$. For any object $\alpha \in \mathbb{P}(U \times X)$ there is a map $!\alpha \to \alpha$ given by



That this is indeed a morphism follows because by definition $(\pi_1, \eta_X)*(!\alpha) \leqslant \alpha$ is equivalent to $!\alpha \leqslant !\alpha$. The map $!\alpha \to !!\alpha$ is given by



so $(\pi_1, \mu_X)^*(!\alpha)$ lies in $\mathbb{P}_\star(U \times TTX)$. Therefore $(\pi_1, \mu_X)^*(!\alpha) \leqslant !!\alpha$ iff

$$(\pi_1, \eta_T X)^*(\pi_1, \mu_X)^*(!\alpha) \leqslant !\alpha$$

But of course $(\pi_1, \eta_T X)^*(\pi_1, \mu_X)^*$ collapses to the identity by monad laws, so our morphism is once again well defined. Naturality of these components are easily verified. The comonad laws are inherited from the monad laws of T.

Natural isomorphisms $!A \otimes !B \cong !(A \times B)$, $I \cong !1$ are induced from the natural isomorphisms $T(X + Y) \cong TX \times TY$ and $T0 \cong 1$ [reindexing preserved products, ! right adjoint so preserves products] The image of the comonoid structure $(A, A \to 1, \triangle \colon A \to A \times A)$ under ! defines a comonoid structure $(!A, \epsilon_A \colon !A \to I, \delta_A \colon !A \to !A \otimes !A)$ on $!A$ induced by the monoid structure $(TX, e_X \colon 1 \to TX, m_X \colon TX \times TX \to TX)$ $\square$

Thus the comonad ! extends **Dial** to a Seely model of linear logic.

Examples of Functional Interpretations

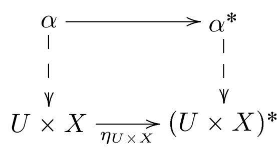
In this section we implement our construction and exhibit categorical models of several variants of the Dialectica interpretation.

The Diller-Nahm interpretation

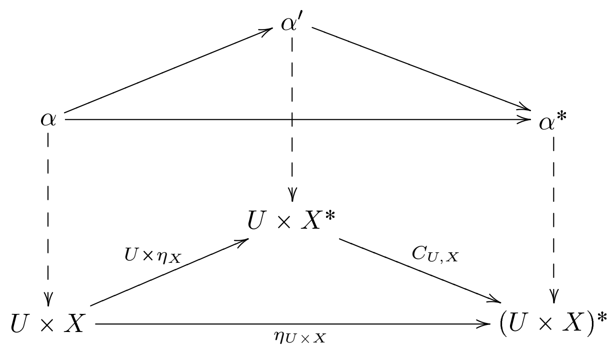
We examine de Paiva's original construction is some detail. Her comonad is based on the monad on the base category that sends types X to the free commutative monoid X^* generated by X, in other words finite strings of elements of X, identified up to reordering. $\eta_X \colon X \to X^*$ sends elements of X to singleton sequences, and $\mu_X \colon X^{**} \to X^*$ intuitively sends a sequence of element of X^* to a sequence of elements of X by 'removing brackets'. Commutativity means that we obtain natural isomorphisms $(X + Y)^* \cong X^* \times Y^*$. The induced union map is the operation of concatenation.

$\mathbb{P}_\star(U \times TX)$ consists of those formulas in $\mathbb{P}(U \times TX)$ such that $\beta(u, \mathbf{x})$ if and only if $\beta(u, \langle x \rangle)$ for all $x \in \mathbf{x}$. The map $\mathbb{P}_\star(U \times TX) \to \mathbb{P}(U \times X)$ induced by re-indexing along $U \times \eta_X$ has right adjoint, sending $\alpha(u, x)$ to $\forall x \in \mathbf{x}\alpha(u, x)$.

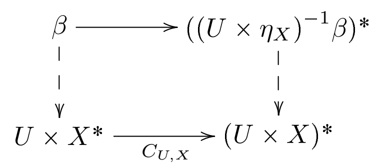
The construction given in [7] is in many ways more elegant that the one given here, in that it has a more abstract formulation. This is in part due to the fact that in de Paiva's setting the free monad extends to a fibred monad, and we get the following useful property, that



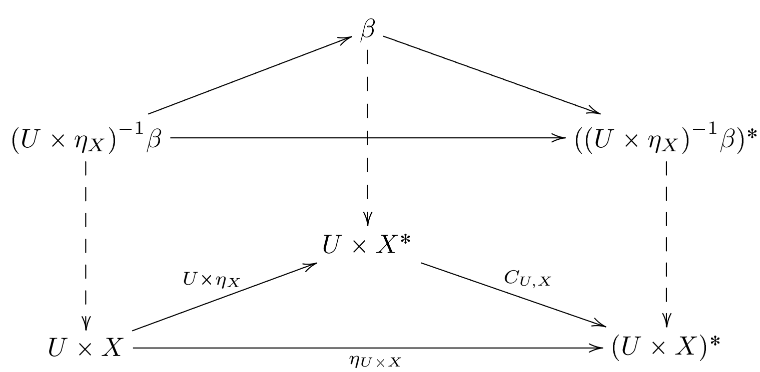
By properties of strength, we can decompose this as



If we define the preorder $\mathbb{P}_\star(U \times TX)$ to be those objects in $\mathbb{P}(U \times TX)$ that for which



then the following diagram commutes



and comparing with diagram [?] reveals that $(U \times \eta_X)^{-1}\beta \leqslant \alpha$ in $\mathbb{P}(U \times X)$ precisely when $\beta \leqslant \alpha'$ in $\mathbb{P}_\star(U \times X^*)$, so the required adjoint is given by

$$!_{U,X}\alpha := \alpha' = {C_{U,X}}^{-1}(\alpha^*)$$

Of course de Paiva's more abstract formulation corresponds precisely to what we would expect intuitively.

Kreisel's modified realizability

A model for modified realizability can be obtained, rather trivially, from the monad that sends types X to the terminal object 1 of the base category. The fact that this defined a strong monad is immediate, as is the existence of the required natural isomorphisms.

In this case the map $U \times \eta_X$ is just the projection $U \times X \xrightarrow{\pi} U$, and provided $\mathbb{T}$ has enough structure the re-indexing map

$$\mathbb{P}(U) \xrightarrow{\pi^{-1}} \mathbb{P}(U \times X)$$

has the familiar right-adjoint $\forall_X$ corresponding to quantification over the whole of X, sending the formula $\alpha(u, x)$ to $\forall x \alpha(u, x)$.

Stein's family of interpretations

Full details not typeset - heavy syntax! Family parametrised by integer n, quantifiers of type level $< n$ left untouched while quantifiers of type level $\geqslant n$ pulled out as an infinite set of witnesses indexed by elements of the pure type $(n-1)$, here presented as functionals of type $(n-1) \to \rho$. Formulas α interpreted as $\exists u \forall \underline{x} \forall \overline{x} \alpha_n(u, x)$ where $\underline{x}$ type level $\geqslant n$, $\overline{x}$ type level $< n$ and α_n contains only universal quantifiers of type level $< n$. Implication treated as

$$\exists f, F \forall \underline{u}, \underline{y} \forall \overline{u}, \overline{y} (\forall i^{(n-1)} \forall \overline{x} \alpha_n(u, Fuyi, \overline{x}) \to \beta_n(fu, y))$$

On a semantic level interpretation based on the monad T_n where $T_n(\underline{X} \times \overline{X}) = T_n(\underline{X}) \times 1$ where $T_n(\underline{X})$ identifies functionals f, g when $rng(f) = rng(g)$. Syntactic treatment of interpretation is easily understood on a semantic level - in fact key functionals needed to satisfy Oliva's conditions can be related to categorical constructions

The bounded interpretation

The bounded functional interpretation was designed explicitly for the extraction of information from proofs. It is not concerned with obtaining precise witnesses, but rather upper bounds for witnesses, and this weakening makes the interpretation extremely flexible both structurally and in terms of the range of non-logical axioms it will admit. It was originally defined to interpret $IL^{\omega}_{\leqslant^*}$, intuitionistic logic in finite types enriched with axioms defining Bezem's majorizability relations $\leqslant^*_\rho$ on all types ρ. Oliva observed that the full interpretation does not exist as a direct instantiation of his parametrised interpretation: this is due to its complex treatment of quantifiers. We do not give full details of the interpretation here (for that the reader is referred to [2] and [5]), but make the important remark remark when restricted to the *propositional* part of the theory, the interpretation does, in fact, fit directly into the general framework.

Propositions α in $IL^{\omega}_{\leqslant^*}$ are translated to $\tilde{\exists} u \tilde{\forall} x \alpha_B(u, x)$ where u and x are *self majorizing*, the *monotone* quantifiers $\tilde{\exists}$ and $\tilde{\forall}$ just quantify over self-majorizing elements, and α_B contains only bounded quantifiers. The translation follows from a *relativization* of quantifiers followed by the usual Skolemization. The key point is that implication is interpreted as

$$\tilde{\exists} f, F \tilde{\forall} u, y (\tilde{\forall} x \leqslant^* Fuy \alpha_B(u, x) \to \beta_B(fu, y))$$

and so the propositional fragment is essentially only concerned with monotone functionals. Note that this is an instantiation of the parametrised quantifier $\forall x \sqsubset t$ where t is some upper bounded set of a preorder. This motivates the abstract treatment we give below, based on a category of preorders.

Let the objects of $\mathbb{T}$ be preorders orders [bottom element?] $(A, \leqslant_A)$, and morphisms $f\colon (A, \leqslant) \to (B, \leqslant)$ be order preserving maps. We can construct products, coproducts and exponentials in our category of preorders in the usual way (in B^A, $f \leqslant_{B^A} g$ if and only if $\forall a(fa \leqslant_B ga)$).

Taking account of the additional structure on $\mathbb{T}$, we define bounded sets of a preorder $(A, \leqslant_A)$ to be sets $\{a | a \leqslant m \wedge m \in M\}$ for some finite set $M \subset A$. We will abbreviate this to $\downarrow M$.

Proposition 0.9. *The map $\mathcal{B}\colon \mathbb{T} \to \mathbb{T}$ that sends preorders A to their set of bounded sets $\mathcal{B}(A)$ is a strong monad on $\mathbb{T}$.*

Proof We need to define $\mathcal{B}$ on morphisms. Given $f\colon A \to B$ and a bounded set $\downarrow M \in \mathcal{B}(A)$, $f(\downarrow M)$ need not be a bounded set in B, so we define $\mathcal{B}(f)(\downarrow M) = \downarrow f(\downarrow M) = \downarrow f(M)$ because f is monotone. We need to verify that given, in addition, $g\colon B \to C$, that $\mathcal{B}(gf) = \mathcal{B}(g)\mathcal{B}(f)$.

Given $\downarrow M \in \mathcal{B}(A)$, $\mathcal{B}(g)\mathcal{B}(f)(\downarrow M) = \mathcal{B}(g) \downarrow f(\downarrow M) = \mathcal{B}(g)(\downarrow f(M)) = \downarrow g(\downarrow f(M)) = \downarrow gf(M) = \downarrow gf \downarrow M = \mathcal{B}(gf) \downarrow M$, which verifies functoriality.

Next, we show that $\mathcal{B}$ forms a monad on $\mathbb{T}$. We define the family of maps $\eta_A\colon A \to \mathcal{B}(A)$ as the maps that sends elements $a \in A$ to bounded sets $\downarrow \{a\}$. These components are clearly monotone, and naturality is easily verified.

We also need maps $\mu_A\colon \mathcal{B}\mathcal{B}(A) \to \mathcal{B}(A)$. $\mathcal{B}$ is idempotent in that $\mathcal{B}\mathcal{B}(A) \cong \mathcal{B}(A)$. μ_A simply sends bounded sets in $\mathcal{B}\mathcal{B}(A)$ to the union of their bounding sets [?]. Again, that the components are monotone and natural in A are easily verified.

The monad laws follow naturally from our definitions, so the only work left is to define strength maps $C_{A,B}\colon A \times \mathcal{B}(B) \to \mathcal{B}(A \times B)$. For $a \in A$ and $\downarrow N \in \mathcal{B}(B)$ we set $C_{A,B}(a, \downarrow N) :\equiv \downarrow(\{a\}, N)$. It is straightforward to show that the relevant diagrams commute and that these maps indeed define a strength on the monad.□

In addition, we clearly have isomorphisms $\mathcal{B}(A + B) \cong \mathcal{B}(A) \times \mathcal{B}(B)$ which induce the obvious union map, sending $(\downarrow M, \downarrow N)$ to the set bounded by maximal elements of $M \cup N$.

The idea, of course, is that given a suitable fibration over $\mathbb{T}$ this structure induces a comonad ! that sends formulas $\alpha(u, x)$ to $!\alpha(u, \downarrow M) := \forall x \in \downarrow M \alpha(u, x)$ where $\downarrow M \in \mathcal{B}X$. Our resulting model corresponds exactly to the bounded interpretation - contraction, for instance, is modelled by

$$\forall u, x_1, x_2(\forall x \leqslant \max x_1 x_2 \alpha(u, x) \to \alpha(u, x_1) \wedge \alpha(u, x_2))$$

which is precisely that given by the bounded interpretation.

Further remarks

To summarise: our purpose has been to Oliva's uniform treatment of functional interpretations to the world of preordered fibrations. Oliva shows that different interpretations arise as instantiations of a parametrised interpretation in which implication is interpreted as

$$\exists f, F \forall x, y(\forall x \sqsubset Fuy \alpha(u, x) \to \beta(fu, y))$$

We present a general construction of a comonad on the Dialectica category, from simple structures on the base category, that models the linear modality !, and hence produce a family of cartesian closed Kleisli categories in which implication is modelled as a map

$$!\alpha \to \beta$$

in **Dial** , where the comonads ! correspond directly to Oliva's generalised bounded quantifier $x \sqsubset t$. In this way we provide categorical models for a number of familiar interpretations, and a general *semantic* framework in which they can be compared.

Our work highlights the fact that, despite the rather peculiar features of the Dialectica interpretation, by enriching the interpreting system with some kind of bounded quantifier we obtain variants that posses excellent structural properties: by interpreting the contraction axiom in a canonical manner we gain a model of the $\rightarrow$, $\wedge$, $\perp$ fragment of logic that identifies proofs that are equivalent under normalisation, which is not the situation with the messy definition by case functionals required for the Dialectica interpretation.

A nice feature of the work begun by de Paiva is its natural link to linear logic and in particular the categorical semantics of linear logic, where the rather mysterious model of Seely is given a concrete illustration by the Dialectica category. We have already referenced the work of Biering [1], which demonstrates that there is certainly potential for the cateogorical semantics of the Dialectica interpretation to be explored further.

However, while it is always important to be able to step back and see things from an abstract perspective, the key significance of functional interpretations today lies in what they are capable of as tools in logic. While the main features of functional interpretations can be expressed in the language of category theory, many of their more interesting aspects lie outside of our framework.

References

[1] B. Biering. *Dialectica Interpretations: a Categorical Analysis.* PhD thesis, IT-University of Copenhagen, 2008.

[2] F. Ferreira and P. Oliva. Bounded functional interpretation. *Annals of Pure and Applied Logic*, 135:73–112, 2005.

[3] J.-Y. Girard. Linear logic. *Theoretical Computer Science*, 50(1):1–102, 1987.

[4] J. M. E. Hyland. Proof theory in the abstract. *Annals of Pure and Applied Logic*, 114:43–78, 2002.

[5] P. Oliva. Unifying functional interpretations. *Notre Dame Journal of Formal Logic*, 47(2):263–290, 2006.

[6] P. Oliva. An analysis of Gödel's dialectica interpretation via linear logic. *dialectica*, 62(2):269–290, 2008.

[7] V. C. V. de Paiva. The Dialectica categories. Technical Report 213, Computer Laboratory, University of Cambridge, Jan 1991.

[8] R. A. G. Seely. Linear logic, *-autonomous categories and cofree coalgebras. *Contemporary Mathematics*, 92, 1989.