Dialectica models of additive-free linear logic

Jules Hedges
Queen Mary University of London
j.hedges@qmul.ac.uk

Abstract—We define a transformation of models of additive-free propositional linear logic, closely related to de Paiva’s dialectica categories, but with a modified exponential based on Oliva’s dialectica interpretation of higher-order linear logic. As a result of this modification we obtain an elegant factorisation of dialectica models as linear-nonlinear adjunctions. From the point of view of game semantics, the transformation can be intuitively seen as prepending games with a round of simultaneous bidding. Two theorems suggest the transformation behaves like a closure operator on models: it has a monad-like structure, and it preserves completeness.

I. INTRODUCTION

This paper presents a construction which transforms models of additive-free propositional linear logic. The construction is based on Gödel’s dialectica interpretation and is closely related to de Paiva’s dialectica categories.

Gödel’s dialectica interpretation (?) is a syntactic transformation that translates a formula into a logically simpler formula which has free higher-type variables but no quantifiers. The purpose was to reduce the consistency of Heyting arithmetic to the consistency of a quantifier-free calculus called T, however the dialectica interpretation (and related translations, collectively called functional interpretations) now have a variety of uses in proof theory, see for example (?).

The paper (?) introduces the dialectica categories G(C), a family of models of linear logic based on the dialectica interpretation, whose objects are ‘generalised relations’. Historically these models were the first to interpret all of the linear connectives without identifying multiplicatives and additives.

The special case G(Set), whose objects are ordinary relations, is discussed in detail and related to game semantics in (?).

A slightly different family of dialectica categories, more important for our purposes, was introduced in (?). These generalise relations in a more intuitive way, to morphisms X × Y → R where R is a monoidal closed poset, or lineale, internal to a category. By taking R to be the two-element poset 0 < 1 we again obtain G(Set). A better-known family of models, the Chu spaces (?), is shown in (?) to be closely related.

Based on de Paiva’s models, (?) gave a syntactic dialectica interpretation to first-order classical affine logic, and (?) to higher-order classical linear logic. Oliva makes a small but crucial modification to the dialectica interpretation of the exponential: whereas de Paiva and Shirahata use

\[ |! \varphi |_x = |\varphi|_{fx} \]

Oliva uses

\[ |! \varphi |_x = ! |\varphi|_{fx} \]

(the precise meaning of the notation will be explained in later sections). This change allows Oliva to prove a completeness theorem analogous to Gödel’s completeness theorem for the dialectica interpretation. The proof-theoretic meaning is that the dialectica interpretation does not eliminate exponentials, as it eliminates quantifiers and additives. The appearance of the exponential on the right hand sides suggests a modification to the dialectica categories, for which we need the monoidal closed poset R to also have an ‘internal exponential’. Thus this work is the result of a back-and-forth of ideas between the worlds of syntax and semantics.

Following this idea, we allow R to be any model of additive-free linear logic (or multiplicative-exponential linear logic, MELL), and build a model D(R) of full propositional linear logic (LL). From a technical point of view this is only a small step from de Paiva’s lineales. However the change of viewpoint is larger: whereas de Paiva’s constructions are seen as building models from scratch, the new construction is more naturally seen as a method for transforming models. Likely this change in viewpoint results from the fact that nontrivial interpretations of the exponentials are much harder to find than interpretations of multiplicatives. From the point of view of game semantics the D construction has a particularly intuitive explanation, namely it modifies games by prepending a round of ‘simultaneous bidding’. This point of view is elaborated in section IV where the action of D on the Abramsky-Jagadeesan model (?) is given explicitly, resulting in ‘Abramsky-Jagadeesan games with bidding’.

By modifying the interpretation of the exponential in this way we obtain:

- An elegant factorisation of dialectica models as linear-nonlinear adjunctions (section III)
- A monad-like structure on the dialectica interpretation of MELL (section VI)
- A completeness theorem for the dialectica interpretation of MELL requiring no characterising principles (section VII)

Together, the second and third results intuitively suggest that the dialectica interpretation of MELL can be thought of as a ‘closure operator’ which improves the completeness of models. As a metaphor we can imagine models of MELL being subsets of some topological space X, where we interpret larger subsets as ‘more complete’ models (that is, models which validate fewer unprovable formulas). The operator D is then a closure operator on X, which generally makes sets larger, but is idempotent.

An analysis of the formulas validated by D(R) for full propositional linear logic, which requires a cut-elimination
II. CATEGORIES OF ZERO-SUM GAMES

We consider a very general notion of finite two-player zero-sum game, and two different notions of morphism of such games, corresponding to linear and intuitionistic logic respectively. We will say linear game and intuitionistic game even though these notions differ only in their morphisms. Intuitively we can think of linear games as simultaneous games, and intuitionistic games as sequential games in which Eloise moves first, however the data specifying the games is the same for both cases.

We have a pair of players Eloise and Abelard, who choose moves from finite sets \( X, Y \), not both empty (the requirement of finiteness is not used, it is only a conceptual simplification). The outcomes of the game are elements of a poset \( (R, \leq) \), which may be large (i.e. a proper class). Intuitively, \( a < b \) means the outcome \( b \) is more desirable for Eloise, and \( a \) is more desirable for Abelard. We generalise further and allow \( R \) to be a (large) category, and write \( a < b \) whenever \( \text{hom}_R(a, b) \) is nonempty and \( a = b \) whenever \( \text{hom}_R(a, b) \) and \( \text{hom}_R(b, a) \) are both nonempty. Formally, this is the result of applying the canonical functor from large categories to large posets, which first identifies morphisms with the same domain and codomain to obtain a preorder, and then identifies isomorphism classes to obtain a partial order. The proof-theoretic reason for doing this is that the dialectica interpretation is proof-agnostic in the following sense. Consider the dialectica interpretation of an intuitionistic implication \( \varphi \to \psi \). A witness for this is a pair of functions \( f, g \) such that for all \( x, v \) it is provable that \( |\varphi|_g(x, v) \to |\psi|_f(x, v) \). We don’t care which proof is used, so long as one exists.

Outcomes of plays of the game are given by a (class) function \( X \times Y \to R \), which is a functor from the finite discrete category \( X \times Y \) to \( R \). If \( G \) is a game then the outcome of the play in which Eloise plays \( x \) and Abelard plays \( y \) will be denoted \( G^x_y \). The game will sometimes be called \( G^X_Y \), and \( X \) will be called the positive set and \( Y \) the negative set. Both the sets from which moves are chosen and the outcomes of plays will be compactly specified by the notation

\[
G^X_Y : (x \atop y) \mapsto \ldots
\]

where the right hand side is an expression involving \( x \) and \( y \), resulting in an object of \( R \).

We consider morphisms of intuitionistic games first. These are closely related to the ‘intuitionistic dialectica categories’ \( \mathcal{D}(\mathcal{C}) \) of (\( \ddagger \)). Given games \( G^X_Y \) and \( H^U_V \), we must specify what it means to play \( H \) relative to \( G \). The dynamics of an intuitionistic game is that Eloise moves before Abelard, and for a relative game we use a certain interleaving of moves:

1. Abelard chooses \( x \in X \)
2. Eloise chooses \( u \in U \)
3. Abelard chooses \( v \in V \)
4. Eloise chooses \( y \in Y \)

Thus a strategy for Eloise consists of functions \( f : X \to U \) and \( g : X \times V \to Y \). Such a strategy is winning iff for all \( x \in X \) and \( v \in V \) we have

\[
G^x_{g(x,v)} \leq H^f_{v}
\]

That is, for all \( x \) and \( v \) there must be an \( R \)-morphism from \( G^x_{g(x,v)} \) to \( H^f_v \).

For the purposes of categorical logic, there is a better definition which remembers the structure of \( R \): a morphism consists of functions \( f, g \) together with an \( X \times V \)-indexed family of \( R \)-morphisms from \( G^x_{g(x,v)} \) to \( H^f_v \). However in this paper we will continue to use the forgetful definition for simplicity.

Now let \( G^X_Y \) and \( H^U_V \) be linear games. The appropriate notion of morphism for such games is to allow Eloise’s move \( u \in U \) to depend only on Abelard’s move \( x \in X \), and Eloise’s move \( y \in Y \) to depend only on Abelard’s move \( v \in V \). Thus a strategy for Eloise consists of functions \( f : X \to U \) and \( g : V \to Y \), and a strategy is winning iff for all \( x \in X \) and \( v \in V \) we have

\[
G^x_{g(x,v)} \leq H^f_v
\]

This ‘game’ cannot be physically played because of the cross-dependence of moves, unless we consider the strategies themselves as moves. These dynamics are discussed from a game-semantic point of view in (\( \ddagger \)).

Given the category \( R \) we have thus defined two more categories: \( \mathcal{D}_1(R) \) is the category of intuitionistic games with outcomes in \( R \), and \( \mathcal{D}_2(R) \) is the category of linear games with outcomes in \( R \).

Lemma 1. \( \mathcal{D}_1(R) \) and \( \mathcal{D}_2(R) \) are categories.

Proof: This is a special case of the proofs in (\( \ddagger \)) that the dialectica categories \( \mathcal{D}(\mathcal{C}) \) and \( \mathcal{G}(\mathcal{C}) \) are categories.

III. THE DIALECTICA TRANSFORMATION OF A MODEL

We begin with a general definition of a model of MELL and a model of LL. A model of multiplicative linear logic (MELL) is given by a \(*\)-autonomous category \( R \), i.e. a symmetric monoidal closed category \( (R, \otimes, \otimes, 1) \) with a functor \( \bot : R \to R \) and natural isomorphisms \( 1 \otimes - \simeq - \otimes 1 \) and

\[
\text{hom}_R(X \otimes Y, Z^\bot) \cong \text{hom}_R(X, (Y \otimes Z)^\bot)
\]

For the interpretation of exponentials we follow (\( \ddagger \)). A categorical model of MELL is given by a \(*\)-autonomous category \( R \) together with another category \( S \) with finite products and an adjunction

\[
S \xleftarrow{\bot} R \xrightarrow{1} S
\]

or, more briefly,

\[
L \bot M : R \to S
\]
Here $L$ (called linearisation) and $M$ (called multiplication) are lax symmetric monoidal functors, that is, there are natural transformations

\[ M X \times MY \to M(X \otimes Y) \quad \top \to M1 \]
\[ LP \otimes LQ \to L(P \times Q) \quad 1 \to L\top \]

and the unit and counit of the adjunction must also respect the monoidal and cartesian monoidal structures (i.e. the adjunction must be a symmetric monoidal adjunction). Such a setup is called a linear-nonlinear adjunction. Given this adjunction, the denotation of the exponential will be a new pair of categories and a comonad on $D$. Given such a model of $MELL$, the dialectica transformation of this model will be a new pair of categories and a linear-nonlinear adjunction.

Given such a model of $MELL$, the dialectica transformation of this model will be a new pair of categories and a linear-nonlinear adjunction $D(L)$ and Chu spaces (called $D(R)$ and $D(S)$). The interpretation of each connective in $D(L)$ is given in figure 1.

The proofs of lemmas 2 and 3 are skipped because they are tedious and very similar to the soundness proofs for dialectica categories (2) and Chu spaces (2).

**Lemma 2.** $D(L)$ is a $*$-autonomous category with monoidal product $\otimes$, unit $1$, internal hom $\to$, involution $^\dagger$ and dualising object $\bot$.

Since the dialectica interpretation eliminates linear additives, we get finite products for free.

**Lemma 3.** $D(L)$ has initial object $0$, terminal object $\top$, products given by $\otimes$ and coproducts given by $\oplus$.

Now we give the dialectica transformations $D_f(M)$ and $D_m(L)$ of multiplication and linearisation functors. The operation $D_f$ is a straightforward lifting operation. The subscript $f$ stands for functor since this construction will be used in section 5 to give the action of $D$ on morphisms (or functors) of models. Suppose the multiplication functor is $M : R \to S$. The functor $D_f(M) : D(L) \to D(S)$ acts on objects $G_f^X$ of $D(L)$ by

\[ D_f(M)(G_f^X) \colon \left( \begin{array}{c} x \\ y \end{array} \right) \mapsto M(G_f^x)^{\dagger} \]

For the action of $D_f(M)$ on morphisms, suppose we have a morphism of $D(L)$ from $G_f^X$ to $H_f^Y$ given by $f : X \to U$ and $g : V \to Y$, so for every $x$ and $v$ we have $G_v^x \leq H_v^fx$. We need to find an element of $\text{hom}_{D(L)}(D_f(M)(G_f^X), D_f(M)(H_f^Y))$

This morphism is given by $f : X \to U$ and $g_f : X \times V \to Y$ given by $g_f(x, v) = gv$. For all $x \in X$ and $v \in V$ we then have $G_v^x \leq H_v^fx$ since $(f, g)$ is a morphism of $D(L)$, therefore $M(G_v^x) \leq M(H_v^fx)$ since $M$ is a functor.

Now suppose the linearisation functor is $L : S \to R$. The functor $D_m(L) : D(S) \to D(R)$ is a symmetry-breaking operation that gives Abelian second-turn advantage in a game. The subscript $m$ stands for minorisation (this name is discussed below). The action on objects $G_m^X$ is given by

\[ D_m(L)(G_m^X) \colon \left( \begin{array}{c} x \\ y \end{array} \right) \mapsto L(G_m^x) \]

Now suppose we have a morphism of $D(L)$ from $G_m^X$ to $H_m^Y$ given by $f : X \to U$ and $g : V \to Y$. For all $x \in X$ and $h : U \to V$ we need to find an element of $\text{hom}_{D(L)}(D_m(L)(G_m^X), D_m(L)(H_m^Y))$

This is given by $f : X \to U$ and $g' : (U \to V) \to (X \to Y)$ given by $g'x = g(x, h(fx))$. Let $x \in X$ and $h : U \to V$. 

Since \((f, g)\) is a morphism of \(\mathcal{D}(S)\) we have
\[
\mathcal{G}^x_{g(x, h(fx))} \leq \mathcal{H}^{fx}_{h(fx)}
\]
which, by the definition of \(g^x\), is
\[
\mathcal{G}^x_{g^x h_x} \leq \mathcal{H}^{fx}_{h(fx)}
\]
Since \(L\) is a functor we have
\[
L(\mathcal{G}^x_{g^x h_x}) \leq L(\mathcal{H}^{fx}_{h(fx)})
\]
Then by the definition of the action of \(L\) on objects we have
\[
L(\mathcal{G})^x_{g^x h} \leq L(\mathcal{H})^{fx}_{h}
\]
as required.

**Lemma 4.** \(\mathcal{D}_m(L) \dashv \mathcal{D}_f(M) : \mathcal{D}_l(R) \to \mathcal{D}_l(S)\) is a linear-nonlinear adjunction.

**Proof:** The equation for the adjunction is
\[
\text{hom}_{\mathcal{D}_l}(\mathcal{D}(L), \mathcal{G}^x_{\mathcal{H}^x}) \cong \text{hom}_{\mathcal{D}_l}(\mathcal{G}, \mathcal{D}(M)(\mathcal{H}))
\]
which is
\[
\text{hom}_{\mathcal{D}_l}(\mathcal{D}(L), \mathcal{G}^x_{\mathcal{H}^x} \to \mathcal{H}^x) \cong \text{hom}_{\mathcal{D}_l}(\mathcal{G}, \mathcal{D}(M)(\mathcal{H}))
\]
These are the sets of morphisms \(f : X \to U\) and \(g : X \times V \to Y\) satisfying, respectively, \(L(\mathcal{G}^x_{f(x, y)}) \leq \mathcal{H}^x_{f(x, y)}\) and \(\mathcal{G}^x_{f(x, y)} \leq M(\mathcal{H}^x_{f(x, y)})\) for all \(x\) and \(y\). These are isomorphic since \(L \dashv M\).

The other conditions for a linear-nonlinear adjunction are omitted for space.

Notice that \(\mathcal{D}_m(L)\) factors again as \(\mathcal{D}_f(L) \circ \mathcal{M}\) where \(\mathcal{M}\) is minorisation, an endofunctor on \(\mathcal{D}_l(S)\) given by
\[
\mathcal{M}(\mathcal{G})^x_{X \to Y} : (x \mapsto f) \mapsto \mathcal{G}^x_{f_x}
\]
The minorisation operation, which modifies a simultaneous game by giving the second player second-turn advantage, was first defined in (7). This three-part factorisation \(! = L \circ \mathcal{M} \circ M\) where \(M \dashv L \circ \mathcal{M}\) seems to be a feature of dialectica-like models. However, if we see games in \(\mathcal{D}_l(S)\) as sequential and \(\mathcal{D}_l(R)\) as simultaneous then \(\mathcal{M}\) is not an operation that makes intuitive sense. In this case, if we begin with a model in which \(R = S\) and \(L\) is the identity functor then \(\mathcal{D}_m(L)\) is a forgetful functor that forgets the sequential structure of a game.

Now we can derive the interpretation of \(!\) as the composition \(\mathcal{D}_m(L) \circ \mathcal{D}_f(M)\). Given a game \(\mathcal{G}^x_{X \to Y}\), its exponential is
\[
(!\mathcal{G})^x_{X \to Y} : (x \mapsto f) \mapsto !\mathcal{G}^x_{f_x}
\]
where the exponential on the right hand side is \(! = L \circ M\).

The interpretation of \(?\) is the dualisation
\[
(?\mathcal{G})_Y^x \to X : (f \mapsto x) \mapsto ?\mathcal{G}^x_{f_y}
\]

The lemmas in this section add up to a soundness theorem.

**Theorem 1.** If \(\mathcal{R}\) is a sound model of MELL then \(\mathcal{D}(\mathcal{R})\) is a sound model of LL.
played. This is precisely a winning strategy for a compound game in which an Abramsky-Jagadeesan game is preceded by a simultaneous bidding round.

Linear negation in \( \mathcal{D}_1(AJ) \) is still given by interchange of players. This is because the general linear negation of \( \mathcal{D}_1(R) \) is to interchange \( X \) and \( Y \) and then apply the linear negation of \( R \), which in this case is also interchange of players.

For a tensor product \( G^x_X \otimes H^y_U \) we have a bidding round which cannot be physically played, in which Eloise simultaneously bids \( x \in X \) and \( u \in U \), and Abelard bids \( y \in Y \) and \( v \in V \) with \( y \) allowed to depend on \( u \) and \( v \) on \( x \). After bidding, the games \( G^x_y \) and \( H^y_v \) are played in parallel.

Finally, although Abramsky-Jagadeesan games do not have a specified interpretation of the exponential, if we use the identity functor then \( \mathcal{D}_1(AJ) \) has a nontrivial exponential given by Abelard taking second-turn advantage in the bidding round.

We can see \( \mathcal{D}_1(AJ) \) as a variant of the dialectica interpretation in which \textit{witnesses} are drawn from the category of finite sets, and \textit{truth values} are drawn from the category of Abramsky-Jagadeesan games. This can be contrasted with (2), which gives a variant of the modified realizability interpretation of Peano arithmetic, in which \textit{witnesses} are drawn from (a variant of) the category of Hyland-Ong games, and \textit{truth values} are the ordinary truth values of Peano arithmetic. Although the dialectica and modified realizability interpretations are closely related (2), these two constructions are rather different: from a game semantic point of view, Blot’s modified realizability can be seen as games with only a bidding round without the following round, but in which the strategies for bidding must be strategies for Hyland-Ong games. On the other hand, \( \mathcal{D}_1(AJ) \) allows arbitrary set-theoretic strategies for the bidding round, although we could obtain a hybrid construction by replacing the category \( \text{FinSet} \) with another category.

V. MORPHISMS OF LINEAR-NONLINEAR ADJUNCTIONS

We have defined a construction \( \mathcal{D} \) which transforms a model of MELL into a model of LL. In order to talk about further structure of \( \mathcal{D} \) we need a notion of morphism of models. There is a certain obvious choice of definition, namely a pair of functors which commute with all of the algebraic structure of a linear-nonlinear adjunction. The class of models of MELL and the class of models of LL with respect to this notion of morphism form categories, with respect to which \( \mathcal{D} \) is a functor. However if we weaken our notion of morphism slightly we can moreover obtain a monad-like structure on \( \mathcal{D} \).

**Definition 1** (Morphism of models). Let \( L \dashv M : R \to S \) and \( L' \dashv M' : R' \to S' \) be models of MELL. A strong morphism of models consists of a pair of functors \( F : R \to R' \) and \( G : S \to S' \) such that we have commuting squares
together with the algebraic properties

- \( F x \otimes F y = F(x \otimes y) \) for \( x,y \in R \) (that is, \( F \) is a monoidal functor)
- \( F(x^\perp) = (Fx)^\perp \) for \( x \in R \)
- \( F1 = 1 \) where \( 1 \in R \) and \( 1 \in R' \) are the monoidal units
- \( G \) commutes with finite products in \( S \) and \( S' \)

If \( L \dashv M : R \to S \) and \( L' \dashv M' : R' \to S' \) are moreover models of LL (that is, if \( R \) and \( R' \) have finite products) then \( F \) should additionally commute with finite products of \( R \) and \( R' \).

Next we define a weak morphism of models. The condition on the functors \( L,M \) is weakened to

where the right square must commute but the left only requires a natural transformation. We also only require \( F \) to be lax monoidal, that is, \( F x \otimes F y \leq F(x \otimes y) \) for \( x,y \in R \). All other conditions remain the same.

We have thus defined four categories. The categories of models and strong morphisms will be called MELL-Mod and LL-Mod. The categories of models and weak morphisms will be called MELL-Mod\(_w\) and LL-Mod\(_w\). In the remainder of this paper we will be concerned mainly with weak morphisms of models. This will be motivated in the next section, in which it is proved that the dialectica interpretation has a monad-like structure on the category of weak morphisms, but this result does not hold for strong morphisms.

Suppose we have a weak morphism from \( L \dashv M : R \to S \) to \( L' \dashv M' : R' \to S' \) given by \( F : R \to R' \) and \( G : S \to S' \). The action of \( \mathcal{D} \) on this morphism is given by \( \mathcal{D}_1(F) : \mathcal{D}_1(R) \to \mathcal{D}_1(R') \) defined by

\[
\mathcal{D}_1(F)(G)^X : \left( \frac{x}{y} \right) \mapsto F(G^x_y)
\]

and similarly for \( \mathcal{D}_1(G) : \mathcal{D}_1(S) \to \mathcal{D}_1(S') \).

**Lemma 5.** \( \mathcal{D} \) is a well-defined functor

\[ \text{MELL-Mod}_w \to \text{LL-Mod}_w \]

**Proof:** We need to prove that
is a morphism of $\text{LL} \cdot \text{Mod}_u$, given that $F,G$ is a morphism of $\text{MELL} \cdot \text{Mod}_u$. For reasons of space only the cases for exponentials and $\otimes$ are included.

For exponentials we need to prove

$$\begin{align*}
\mathcal{D}_i(S) & \xrightarrow{\mathcal{D}_i(l)} \mathcal{D}_i(R) \\
\mathcal{D}_i(S') & \xleftarrow{\mathcal{D}_i(l')^R} \mathcal{D}_i(S)
\end{align*}$$

We are given a natural transformation $\eta : L' \circ G \Rightarrow F \circ L$, so we have a family of morphisms $\eta_x \in \text{hom}_R((L' \circ G)(x),(F \circ L)(x))$ for $x \in S$. Let $G^\lambda_x \in \text{Ob}(\mathcal{D}_i(S))$. Then we have

$$(\mathcal{D}_i(m')(L') \circ \mathcal{D}_f(G))(\mathcal{D}_i(G)) \mathcal{D}_i(l') \mathcal{D}_i(x) = (L' \circ G)(G^\lambda_x)$$

$$(\mathcal{D}_f(F) \circ \mathcal{D}_m(L))(\mathcal{D}_i(G)) \mathcal{D}_i(l') \mathcal{D}_i(x) = (F \circ L)(G^\lambda_x)$$

For a natural transformation $\mathcal{D}(\eta) : \mathcal{D}_m(L') \circ \mathcal{D}_f(G) \Rightarrow \mathcal{D}_f(F) \circ \mathcal{D}_m(L)$ we need, for each $G^\lambda_x \in \text{Ob}(\mathcal{D}_i(S))$, a morphism in

$\text{hom}_{\mathcal{D}_i(R)}((\mathcal{D}_i(m')(L') \circ \mathcal{D}_f(G))(\mathcal{D}_i(G)), (\mathcal{D}_f(F) \circ \mathcal{D}_m(L))(\mathcal{D}_i(G)))$

The identity functions give such a morphism, because

$$\eta_{G^\lambda_x} : (L' \circ G)(G^\lambda_x) \Rightarrow (F \circ L)(G^\lambda_x)$$

For the right hand square let $G^\lambda_x \in \text{Ob}(\mathcal{D}_i(R))$. We have

$$(\mathcal{D}_f(G) \circ \mathcal{D}_f(M))(\mathcal{D}_i(G)) \mathcal{D}_i(l') \mathcal{D}_i(x) = (G \circ M)(G^\lambda_x)$$

$$(\mathcal{D}_f(F) \circ \mathcal{D}_m(L))(\mathcal{D}_i(G)) \mathcal{D}_i(l') \mathcal{D}_i(x) = (M' \circ F)(G^\lambda_x)$$

These are equal because $G \circ M = M' \circ F$.

To prove that $\mathcal{D}_f(F)$ is lax monoidal, consider games $G^\lambda_Y, H^\lambda_Y \in \mathcal{D}_i(R)$. We have

$$(\mathcal{D}_f(F)(\mathcal{G}_Y) \circ \mathcal{D}_f(F)(\mathcal{H}_Y))^{(x \times U)_{(U \times Y) \times (X \times V)}} :$$

$$\begin{align*}
(x,u) & \mapsto F(G^\lambda_{fu} \otimes H^\lambda_{gy}) \\
(f,g) & \mapsto F(G^\lambda_{fu} \otimes H^\lambda_{gy})
\end{align*}$$

and

$$(\mathcal{D}_f(F)(\mathcal{G} \otimes \mathcal{H}))^{(x \times U)_{(U \times Y) \times (X \times V)}} :$$

$$\begin{align*}
(x,u) & \mapsto F(G^\lambda_{fu} \otimes H^\lambda_{gy}) \\
(f,g) & \mapsto F(G^\lambda_{fu} \otimes H^\lambda_{gy})
\end{align*}$$

The identity maps give a relative winning strategy

$$\mathcal{D}_f(F)(\mathcal{G}) \otimes \mathcal{D}_f(F)(\mathcal{H}) \Rightarrow \mathcal{D}(F)(\mathcal{G} \otimes \mathcal{H})$$

because we have

$$F(G^\lambda_{fu} \otimes H^\lambda_{gy}) \leq F(G^\lambda_{fu} \otimes H^\lambda_{gy})$$

The other cases of the proof are very similar: they all involve computing a pair of games and constructing relative winning strategies.

VI. THE DIALECTICA INTERPRETATION IS A MONAD

We have defined the dialectica transformation as a functor $\mathcal{D} : \text{MELL} \cdot \text{Mod}_u \Rightarrow \text{LL} \cdot \text{Mod}_u$. Since every model of $\text{LL}$ is also a model of $\text{MELL}$ by forgetting the products in $R$, we also have a forgetful functor $U : \text{LL} \cdot \text{Mod}_u \Rightarrow \text{MELL} \cdot \text{Mod}_u$. The composition $U \circ \mathcal{D}$ is therefore an endofunctor on $\text{MELL} \cdot \text{Mod}_u$. It turns out that $U \circ \mathcal{D}$ is a ‘near monad’ in the sense that the required natural transformations exist but one of the monad laws fails. In the remainder of this section we will abuse notation by writing $\mathcal{D}$ for $U \circ \mathcal{D}$.

The unit

$$\eta_R : R \Rightarrow \mathcal{D}_i(R)$$

takes an object $x \in R$ to the game with one play and outcome $x$,

$$\eta_R(x)_\{x\} : \{x\} \Rightarrow x$$

It takes a morphism in $\text{hom}_R(x,y)$ to the unique strategy with the correct type (with both functions taking $*$ to $*$), which is a winning strategy.

Next we consider the multiplication

$$\mu_R : \mathcal{D}^2_i(R) \Rightarrow \mathcal{D}_i(R)$$

Consider a game $G^\lambda_x \in \mathcal{D}_i^2(R)$. For a particular play $x,y$ the game has an outcome $G^\lambda_{xy} \in \mathcal{D}_i(R)$, which is itself a game. Suppose this game is defined by

$$(G^\lambda_{xy})_{U^\alpha_y} : (u, v) \mapsto (G^\lambda_{xy})_{U^\alpha_y}$$

where $U^\alpha_y$ and $V^\alpha_x$ are families of sets indexed by $x$ and $y$. The action of $\mu_R$ is to take this game to

$$\mu_R(G^\lambda)_{\sum_{x,y} U^\alpha_y} : (x,f,y,g) \mapsto (G^\lambda_{xy})_{U^\alpha_y}$$

The witness sets are dependent types in the category of finite sets, so an element $(x,f) \in \sum_x \prod_y U^\alpha_y$ consists of a point $x \in X$ and a function $f$ with the property that $fy \in U^\alpha_y$ for every $y \in Y$. This will sometimes be written using the dependent type theory notation

$$f : (y : Y) \Rightarrow U^\alpha_y$$

Suppose we have another game $H^\lambda_Y \in \mathcal{D}_i^2(R)$ defined by

$$(H^\lambda_{xy})_{Q^\alpha_y} : (q, p) \mapsto (H^\lambda_{xy})_{Q^\alpha_y}$$
A morphism of $\mathcal{D}_1^2(R)$ from $\mathcal{G}$ to $\mathcal{H}$ consists of functions $f : X \to Z$ and $g : W \to Y$ such that for every $x \in X$ and $w \in W$ we have functions

$$\alpha_{x,w} : U^x_{gw} \to P^x_{fw}$$
$$\beta_{x,w} : Q^x_{gw} \to V^x_{fw}$$

We need to produce a morphism from $\mu_R(\mathcal{G})$ to $\mu_R(\mathcal{H})$, that is, we need functions

$$F : \prod_{x} \prod_{y} U^x_y \to \prod_{x} \prod_{w} P^x_w$$
$$G : \prod_{w} \prod_{z} Q^w_z \to \prod_{x} \prod_{y} V^x_y$$

These are defined by

$$F(x,h) = (f,\lambda w. \alpha_{x,w}(h(gw)))$$
$$G(w,h') = (gw,\lambda x. \beta_{x,w}(h'(fx)))$$

Since a weak morphism consists of a pair of functors $R \to R'$ and $S \to S'$ we define both functors in the same way (with a minor change needed to the action on the morphisms of $S$). The operations $\eta$ and $\mu$ are illustrated in figure 2

**Lemma 6.** $\eta_R : R \to \mathcal{D}_1(R)$ and $\mu_R : \mathcal{D}_1^2(R) \to \mathcal{D}_1(R)$ are well-defined natural transformations.

**Proof:** For reasons of space only some cases of this proof for $\mu_R$ are included. Although tedious, this proof is important because it fails to hold for the category of strong morphisms, and it is the cases that fail that motivate the definition of weak morphism.

Suppose we have games $\mathcal{G}_X, \mathcal{H}_Z \in \mathcal{D}_1^2(R)$ given by

$$(\mathcal{G}_X^w)^{u w} : (u \quad w) \mapsto (\mathcal{G}_X^w)^u_w$$

and

$$(\mathcal{H}_Z^w)^{p q} : (p \quad q) \mapsto (\mathcal{H}_Z^w)^p_q$$

We need to construct a relative winning strategy

$$\mu_R \mathcal{G} \otimes \mu_R \mathcal{H} \to \mu_R (\mathcal{G} \otimes \mathcal{H})$$

The game $\mu_R \mathcal{G} \otimes \mu_R \mathcal{H}$ has positive set

$$\sum_{x : X} \sum_{y : Y} U^x_y \times \sum_{w : W} P^x_w$$

and negative set

$$\left( \sum_{x : X} \sum_{y : Y} P^x_w \to \sum_{y : Y} \sum_{x : X} U^x_y \to \sum_{x : X} \sum_{y : Y} Q^x_y \right)$$

The game $\mu_R (\mathcal{G} \otimes \mathcal{H})$ has positive set

$$\sum_{x : X} \sum_{y : Y} U^x_y \times \sum_{w : W} P^x_w$$

and negative set

$$\sum_{(x,w) : X \times W} (P^u_{fw} \times Q^v_{gw}) \to \left( U^x_{fw} \times Q^v_{gw} \right)$$

To define a function

$$\sum_{x : X} \sum_{y : Y} U^x_y \times \sum_{w : W} P^x_w$$

suppose we are given $((x, \alpha), (w, \beta))$ where $\alpha : (y : Y) \to U^x_y$ and $\beta : (z : Z) \to P^x_w$. We need to define

$$F : ((f, g) : (W \to Y) \times (X \to Z)) \to U^x_{fw} \times P^u_{gx}$$

which can be given by

$$F(f, g) = (\alpha(fw), \beta(gx))$$

In the other direction we need to define a function

$$\sum_{(f,g) : (W \to Y) \times (X \to Z)} \prod_{x : X} \prod_{w : W} (P^u_{fw} \to V^x_{fw}) \times (U^x_{fw} \times Q^v_{gw}) \to \sum_{x : X} \sum_{y : Y} U^x_y \times \sum_{w : W} \sum_{z : Z} Q^x_z$$

Consider the left projection of this function. As input we are given the data

$$f : W \to Y$$
$$g : X \to Z$$
$$F : ((x, w) : X \times W) \to ((P^u_{fx} \to V^x_{fw}) \times (U^x_{fw} \times Q^v_{gw}))$$
$$w : W$$
$$h : (z : Z) \to P^u_{fw}$$

We must produce $y : Y$ and $h' : (x : X) \to V(x, y)$. We take $y = fw$ and

$$h' x = \pi_L(F(x, w))(h(gx))$$

The right projection is symmetric.
Notice that there is no relative winning strategy
\[ \mu_R(\mathcal{G} \otimes \mathcal{H}) \to \mu_R \mathcal{G} \otimes \mu_R \mathcal{H} \]
so we cannot extend $\mu_R$ to a natural transformation for the category of strong morphisms.

For the exponentials we need to show
\[
\begin{align*}
\mathcal{D}_1(S) & \xrightarrow{\mathcal{D}_m(L)} \mathcal{D}_1(R) \xrightarrow{\mathcal{D}_f(M)} \mathcal{D}_1(S) \\
\mathcal{D}_1^2(S) & \xrightarrow{\mathcal{D}_m^2(L)} \mathcal{D}_1^2(R) \xrightarrow{\mathcal{D}_f^2(M)} \mathcal{D}_1^2(S)
\end{align*}
\]

For the square on the right consider the game $\mathcal{G}_y^X \in \mathcal{D}_1^2(S)$
given by
\[
(\mathcal{G}_y^X)_{y}^{v} : \left( \frac{v}{v} \right) \mapsto (\mathcal{G}_y^X)_{y}^{v}
\]
If we explicitly find $(\mathcal{D}_f(M) \circ \mu_R)(\mathcal{G})$ and $(\mu_S \circ \mathcal{D}_m^2(M))(\mathcal{G})$
we indeed find that they are both equal to
\[
\left( \frac{x}{g}, f \right) \mapsto M((\mathcal{G}_y^X)_{g}^{f})
\]

For the square on the left suppose instead that $\mathcal{G} \in \mathcal{D}_1^2(S)$.
We have
\[
(\mathcal{D}_m(L) \circ \mu_S)(\mathcal{G}) \sum_x \sum_y U^x_y \rightarrow \sum_x \sum_y U^x_y \rightarrow \sum_x \sum_y V^x_y
\]
and
\[
(\mu_R \circ \mathcal{D}_m(L))(\mathcal{G}) \sum_x \sum_y U^x_y \rightarrow \sum_x \sum_y V^x_y
\]

There is a map
\[
\sum_x \sum_y U^x_y \rightarrow \sum_x \sum_y V^x_y
\]
given by
\[
(x, y) \mapsto (x, \lambda f X \rightarrow Y \cdot g(f(x))
\]

There is also a map
\[
\sum_x \sum_y U^x_0 \rightarrow \sum_x \sum_y V^x_x \rightarrow \sum_x \sum_y V^x_y
\]

As input we are given the data
\[
\begin{align*}
f : X & \rightarrow Y \\
g : (x, x) & \rightarrow U^x_0 \rightarrow V^x_0 \\
x : X \\
h : (y, y) & \rightarrow U^y_0
\end{align*}
\]
We must produce $y : Y$ and $h' : (x' : X) \rightarrow V^y_{x'}$. We take $y = f x$ and
\[ h' x' = g x(h(f x)) \]

This case also fails for strong morphisms because there is no general function
\[
\sum_x \sum_y U^x_y \rightarrow \sum_x \sum_y V^x_y \rightarrow \sum_f \sum_x (U^x_0 \rightarrow V^x_f)
\]

The first monad law holds, but only up to natural isomorphism.

**Theorem 2.** There are natural isomorphisms
\[ \mu_R \circ \eta_{\mathcal{D}_1(R)} \cong \mu_R \circ \mathcal{D}_f(\mu_R) \cong id_{\mathcal{D}_1(R)} : \mathcal{D}_1(R) \rightarrow \mathcal{D}_1(R) \]

**Proof:** Consider an object $\mathcal{G}_y^X$ of $\mathcal{D}_1(R)$. We can directly compute:
\[
(\mu_R \circ \eta_{\mathcal{D}_1(R)})(\mathcal{G}) \sum_x \sum_y V^x_y \rightarrow (\mathcal{G}_y^X)_{y}^{f}
\]
and
\[
(\mu_R \circ \mathcal{D}_f(\mu_R))(\mathcal{G}) \sum_x \sum_y V^x_y \rightarrow (\mathcal{G}_y^X)_{y}^{f}
\]

These are both naturally isomorphic to $\mathcal{G}_y^X$.

The second monad law
\[ \mu_R \circ \mathcal{D}_f(\mu_R) \cong \mu_R \circ \mathcal{D}_m(\mu_R) : \mathcal{D}_1^2(R) \rightarrow \mathcal{D}_1^2(R) \]
fails. Consider an object of $\mathcal{D}_1^2(R)$ given by
\[
\mathcal{G} : \left( \frac{x, X}{y, Y} \rightarrow \left( \begin{array}{c} u : U^x_y \\ v : V^x_y \end{array} \right) \right) \rightarrow \left( \begin{array}{c} q : (P_{x,y})^u \\ p : (P_{x,y})^v \end{array} \right) \rightarrow ((\mathcal{G}_y^X)_y)^{p q}
\]
We can directly compute
\[
\mathcal{D}_f(\mu_R)(\mathcal{G}) : \left( \frac{x, X}{y, Y} \rightarrow \left( \begin{array}{c} u : U^x_y \\ v : V^x_y \end{array} \right) \right) \rightarrow \left( \begin{array}{c} q : (Q_{x,y})^u \\ p : (Q_{x,y})^v \end{array} \right) \rightarrow ((\mathcal{G}_y^X)_y)^{p q}
\]

Therefore
\[
\mu_R \circ \mathcal{D}_f(\mu_R)(\mathcal{G}) \sum_x \sum_y V^x_y \rightarrow (\mathcal{G}_y^X)_y^{p q}
\]

We also get
\[
\mu_{\mathcal{D}_1(R)}(\mathcal{G}) \rightarrow (\mathcal{G}_y^X)_y^{p q}
\]

Then $(\mu_R \circ \mathcal{D}_m(\mu_R))(\mathcal{G})$ involves nested dependent types:
\[
(\mu_R \circ \mathcal{D}_m(\mu_R))(\mathcal{G}) \sum_{x, y, \sum_x \sum_y U^x_y} \sum_{y, g, \sum_y V^x_y} \rightarrow (\mathcal{G}_y^X)_y^{p q}
\]

There is a natural transformation
\[
\sum_x \sum_y \sum_u (P_{x,y}^u)^{p q} \rightarrow \sum_x \sum_y \sum_u (P_{x,y}^u)^{f g}
\]
defined by
\[
(x, f, \alpha) \mapsto (x, f, \lambda(y, g) \cdot \alpha g(x))
\]
However there is none in the opposite direction. Similarly there
is a natural transformation

$$\sum_y \prod_x \prod_v \prod_u (Q^u y)^v \rightarrow \sum_y \prod_x \prod_v \prod_u \prod_{g\circ \eta_y} (Q^u y)^v \rightarrow$$

but none in the opposite direction. As a result, there is no
morphism of $\mathcal{D}_1(R)$ in either direction between the games
$(\mu_R \circ \mathcal{D}_f(\mu_R))(G)$ and $(\mu_R \circ \mu_{\mathcal{D}_1(R)})(G)$.

The natural transformations $\eta$ and $\mu$ give a high-level explanation of some aspects of the dialectica interpretation, by
giving ‘embeddings’ of the models $\mathcal{R}$ and $\mathcal{D}^2(\mathcal{R})$ into $\mathcal{D}(\mathcal{R})$. Intuitively, the first tells us that the dialectica interpretation
loses no information, and the second tells us that applying the
dialectica interpretation twice gains no more information
than applying it once. However the logical significance of the
failure of the second monad law, and the failure of $\mu$ to exist
for strong morhisms, still requires explanation.

VII. RELATIVE COMPLETENESS FOR MELL

Definition 2 (Complete model). Let $\mathcal{R}$ be a model of $\mathsf{LL}$. A
mapping from atoms to objects of $R$ is called a valuation in
$\mathcal{R}$. Given a valuation $v$, we can extend it inductively to an
interpretation of formulas in $\mathcal{R}$, denoted $[\cdot]_v$.

$\mathcal{R}$ is called a complete model of $\mathsf{LL}$ if for all formulas $\varphi, \psi$, if
$[\varphi]_v \leq [\psi]_v$ for all valuations $v$ then the sequent $\varphi \vdash \psi$ is
derivable. Completeness for MELL is defined similarly.

Lemma 7. Let $\mathcal{R}$ be a model of MELL and let $v$ be a valuation
in $\mathcal{R}$, with resulting semantic functor $[\cdot]_v : \mathsf{MELL} \rightarrow \mathcal{R}$.

Define a valuation for $\mathcal{D}(\mathcal{R})$ by letting the interpretation of $p$
be

$$[p]_v = (\cdot)_v \mapsto [p]_v$$

Proof: Let $\varphi$ be a formula of MELL with interpretation $[\varphi]_v$ in $\mathcal{D}(\mathcal{R})$. We must prove:

1) There exists $x \in X$ such that for all $y \in Y$, $[\varphi]_v \leq [\varphi]^x_y$.
2) There exists $y \in Y$ such that for all $x \in X$, $[\varphi]^x_y \leq [\varphi]_v$.

These are proved simultaneously by induction on the formula
$\varphi$. In the base case we have that $\varphi = p$ is an atom, and the
point $*$ witnesses both (1) and (2).

In the negation case for (1) the inductive hypothesis for
(2) gives $y \in Y$ such that for all $x \in X$, $[\varphi]^x_y \leq \varphi$. Then
$\varphi^x \leq ([\varphi]^x_y)^x = [\varphi]^x_y$. The case for (2) is symmetric.

For (1) of $\otimes$, the inductive hypothesis gives $x$ and $u$ such that for
all $y$ and $v$ we have $\varphi \leq [\varphi]^x_y$ and $\psi \leq [\psi]^x_y$.
Let $f : U \rightarrow Y$ and $g : X \rightarrow V$. Then we have $\varphi \leq [\varphi]^x_y$ and
$\psi \leq [\psi]^x_y$. Then we get $\varphi \otimes \psi \leq [\varphi]^x_y \otimes [\psi]^x_y$ as required.

For (2) of $\otimes$, the inductive hypothesis gives $y$ and $v$ such that for
all $x$ and $u$ we have $[\varphi]^x_y \leq \varphi$ and $[\psi]^x_y \leq \psi$. Define
$f : U \rightarrow Y$ by $f u = y$ and $g : X \rightarrow V$ by $g x = v$. Let $x$
and $u$ be arbitrary. Then we have $[\varphi]^x_y \otimes [\psi]^x_y$ as required.

For (1) of $\unit$ the inductive hypothesis gives $x$ such that for
all $y$, $[\varphi]^x_y \leq \varphi$. Let $f : X \rightarrow Y$ be arbitrary. Then $\varphi \leq [\varphi]^x_y$, so $\varphi \leq [\varphi]^x_y$. This completes
the proof.

Theorem 3. If $\mathcal{R}$ is a complete model of MELL then $\mathcal{D}(\mathcal{R})$
is a complete model of MELL.

Proof: Let $\varphi$ and $\psi$ be formulas of MELL such that for
all valuations in $\mathcal{D}(\mathcal{R})$ there is a morphism in $\mathcal{D}_1(\mathcal{R})$ from
$[\varphi]$ to $[\psi]$. Let $v$ be a valuation in $\mathcal{R}$ and let $[\cdot]_v$ be the
resulting valuation defined in the previous lemma. Then

$$\eta_R([\varphi]_v) = [\varphi]_v \leq [\psi]_v = \eta_R([\psi]_v)$$

Since the witnesses for $\eta_R$ are trivial we have $[\varphi]_v \leq [\psi]_v$. Since $v$ was arbitrary and $\mathcal{R}$ is complete we get that the
sequent $\varphi \vdash \psi$ is derivable in MELL, as required.

Note that unlike the completeness proof in (7), on which
this proof is based, we require no characterising principles.
The reason for this seems to be that the quantifiers of the
inductive hypothesis

$$\varphi \iff (\exists x) [\varphi]^x_y$$

have been ‘externalised’ in the inductive hypothesis of lemma
[7]. To the author’s knowledge, this is the first variant of
the dialectica interpretation with no characterising principles,
or to say it another way, the first variant of the dialectica
interpretation that is complete for a logic with full cut-
elimination. Of course from some perspectives the loss of
characterising principles is a weakness. If we re-introduce
additives or quantifiers this result fails, as is discussed in
the next section.

VIII. FURTHER DIRECTIONS

There are several aspects of this work that suggest further
investigation:

1) The most obvious question is: if $\mathcal{R}$ is a complete
model of MELL, as in the hypothesis of the previous
theorem, how complete is $\mathcal{D}(\mathcal{R})$ for full propositional
linear logic? By analogy to the dialectica interpretation
in higher types, the author conjectures that $\mathcal{D}(\mathcal{R})$ is
complete for the logic obtained by extending $\mathsf{LL}$ with
certain formulas of the form

$$(\varphi \otimes \psi) \not\rightarrow (\varphi \not\rightarrow \varphi) \oplus (\psi \not\rightarrow \psi)$$
These formulas are propositional analogues respectively of independence of premise (IP) and trump advantage (TA), which are characterising principles in (?). Both are intuitively justified by the resource interpretation: for the first, if we have either \( \varphi \) or \( \psi \) together with \( \chi \) then we either have \( \varphi \) together with \( \chi \) or \( \psi \) together with \( \chi \); for the second, if we an infinite supply of things that are either \( \varphi \) or \( \psi \) then by the infinite pigeonhole principle we have either an infinite supply of \( \varphi \) or an infinite supply of \( \psi \). The author is currently working on a proof of this classification, using a cut-elimination argument, which will appear in a future paper. The central idea of the proof is to extend linear logic with a new kind of formula, called simultaneous additives, which have a similar relationship to ordinary additives as Henkin quantifiers have to ordinary higher-type quantifiers. The dialectica interpretation of such a simultaneous additive is given precisely by the natural transformation \( \mu \). In the presence of these new formulas, and certain characterising principles, we can prove completeness. The most difficult part of the proof is using cut-elimination to conclude that this new logic is conservative over a certain extension of ordinary linear logic.

2) The appearance of dependent types in \( \mathcal{D}(R) \) suggests a hybrid approach where rather than using sets and set-theoretic functions we use the objects and morphisms of another locally cartesian closed category \( C \). By taking \( C \) to be a dependent type theory (with types as objects and terms as morphisms) we can obtain a purely syntactic transformation similar to the original dialectica interpretation. A particularly interesting idea is to find a variant which refutes the characterising principles mentioned in point (1), which would lead to a complete model. We conjecture that this is impossible, and no dialectica-like model can be complete for full propositional linear logic.

3) Similarly, we can replace the category of sets with a dependently-typed programming language such as Agda (?). This would give a simple way to embed linear reasoning in a programming language. Since most of the proofs in this paper amount to constructing inhabitants of dependent types, they are natural candidates for formalisation in Agda. The author intends to verify the proofs in this paper in this way.

4) By changing the interpretation of exponentials we can make ‘semantic’ equivalents of other functional interpretations besides the dialectica interpretation. See (?) for the relationship between different functional interpretations and the semantics of the linear exponential. However the dialectica interpretation is the only functional interpretation whose exponential does not use higher-type quantifiers, so other functional interpretations will require starting with a model of full linear logic in higher types. This might be an interesting exercise, to translate the heavily syntactic literature on functional interpretations into the language of semantics, but it is unlikely to produce any interesting models.