Dialectica Principles via Gödel Doctrines*

Davide Trotta
University of Pisa, IT

Matteo Spadetto
University of Leeds, UK

Valeria de Paiva
Topos Institute, CA

Nov 2022

Abstract

Gödel's Dialectica interpretation was conceived as a tool to obtain the consistency of Heyting arithmetic in the 40s. In recent years, several proof theoretic transformations, based on Gödel's Dialectica interpretation, have been used systematically to extract new content from classical proofs, following a suggestion of Kreisel. Thus, the interpretation has found new relevant applications in several areas of mathematics and computer science. Several authors have explained the Dialectica interpretation in categorical terms. In particular, de Paiva introduced the notion of a *Dialectica category* as an internal version of Gödel's Dialectica Interpretation in her doctoral work. This was generalised by Hyland and Hofstra, who considered the interpretation in terms of fibrations. In our previous work, we introduced an *intrinsic* presentation of the Dialectica construction via a generalisation of Hofstra's work, using the notion of Gödel fibration and its proof-irrelevant version, a Gödel doctrine. The key idea is that Gödel fibrations (and doctrines) can be thought of as fibrations generated by some basic predicates playing the role of *quantifier-free predicates*. This categorification of quantifier-free predicates is crucial not only to show that our notion of Gödel fibration is equivalent to Hofstra's Dialectica fibration in the appropriate way, but also to show how Gödel doctrines embody the main logical features of the Dialectica Interpretation. To show that, we derive the soundness of the interpretation of the implication connective, as expounded by Troelstra, in the categorical model. This requires extra logical principles, going beyond intuitionistic logic, namely (suitable versions of) the Markov Principle (MP) and the Independence of Premise (IP) principle, as well as some choice. We show how these principles are satisfied in the categorical setting, establishing a tight correspondence between the logical system and the categorical framework. This tight correspondence should come in handy not only when discussing the traditional applications of Dialectica, but also when dealing with some newer uses of the interpretation, as in modelling games or concurrency theory.

*Research supported by the project MIUR PRIN 2017FTXR IT-MaTTerS (Trotta) and by a School of Mathematics EPSRC Doctoral Studentship (Spadetto).
0000-0003-4509-594X | 0000-0002-6495-7405 | 0000-0002-1078-6970
trottadavide92@gmail.com | matteo.spadetto.42@gmail.com | valeria@topos.institute

Finally, to complete our analysis, we characterise categories obtained as a result of the Hyland, Johnstone and Pitts tripos-to-topos construction when applied to Gödel doctrines.

Keywords. Gödel doctrine; Hyperdoctrine; Dialectica doctrine; Dialectica category; Dialectica interpretation; Logical principles.

1 Introduction

Gödel's Dialectica Interpretation is a proof interpretation of intuitionistic arithmetic in a quantifier-free theory of functionals of finite type, called System T. The interpretation's original use was to show the consistency of Heyting (or intuitionistic) arithmetic. When combined with Gödel's double-negation interpretation, which reduces classical arithmetic to intuitionistic arithmetic, the Dialectica interpretation yields a reduction of the classical theory as well. This approach has since been extended and adapted to other theories, and the pattern usually follows Gödel's original example: first, one reduces a classical theory to a variant based on intuitionistic logic. Then one reduces the intuitionistic theory to a quantifier-free functional theory, like system T. Much work has been done to understand this proof interpretation using categorical means. The notion of a *Dialectica category* was introduced by de Paiva [6] as an internal version of Gödel's Dialectica Interpretation. The idea is to construct a Dialectica category $\mathfrak{Dial}(\mathcal{C})$ from a category $\mathcal{C}$ with finite limits. The main focus in de Paiva's original work is on the categorical structure of the category obtained, as this notion of Dialectica category turns out to be also a model of Girard's Linear Logic [12].

The construction was generalised by Hyland, who investigated the Dialectica construction associated to a fibred preorder [17]. Biering [3] studied the Dialectica construction for an arbitrary cloven fibration. Later Hofstra [16] wrote an exposition and interpretation of the Dialectica construction, emphasising its universal properties. His work gives centre stage to the well-known concepts of pseudo-monads, simple products, and co-products.

Taking advantage of the abstract presentation of Hofstra, in previous work [38, 39] we introduced an *intrinsic* presentation of the Dialectica construction via the notion of Gödel fibration (and its proof-irrelevant version) a Gödel doctrine. The key idea is that Gödel fibrations (and doctrines) can be thought of as fibrations generated by some basic predicates playing the role of *quantifier-free predicates*. This categorification of quantifier-free predicates is crucial not only to show that the notions of Gödel fibrations introduced in [38] and Dialectica fibrations (as presented in [16]) are mathematically equivalent in the appropriate way, but also to show how Gödel doctrines embody the main logical features of the Dialectica Interpretation [39, 40]. While in [38] we presented a reconstruction of where the categorification of concepts came from in the proof-theory, in [40] we showed that this categorification worked not only for rules as in [39], but also for the logical principles themselves, which is always more exciting for logicians.

The main purpose of the work here is to provide a self-contained and complete study of Gödel doctrines, presenting in detail the results developed in [39, 40], their connections with [38], and carrying out the analysis of these doctrines. The notion of Gödel doctrine provides an intrinsic categorical description of the Dialectica Interpretation, faithfully modelling the logical language, and generalised enough to provide a bridge between Gödel fibrations (useful for more expressive logical languages like dependent type theory) and Dialectica categories, very helpful for propositional languages and simply typed calculi. A relevant advantage of having such an intrinsic presentation (with respect to the previous presentations in terms of dialectica categories and fibrations) is that it allows us to recognize and represent all the logical features of the Dialectica interpretation (such as the logical principles) in the categorical setting.

Thus, in this paper, we first recall the Dialectica interpretation as presented by Troelstra in the Introductory Note [10] to the Dialectica article [13]. We discuss the Dialectica interpretation of the implication connective and the logical principles involved in the categorical interpretation of this connective. Then we present in detail the notion of categorical *doctrine* and its logical meaning. Moreover, in order to present our results with a notation familiar to both logicians and category theorists, we employ the *internal language* of a doctrine [32]. As far as the notion of doctrine is concerned, we follow the notation and the definitions of Maietti and Rosolini in [24, 25], where the authors introduce primary, elementary, and existential doctrines as generalisations of the original notion of Lawvere's hyperdoctrine [21].

After recalling the main categorical tools involved, we focus on the notion of Gödel doctrine introduced in [39], and we show how these doctrines embody the main logical features of the Dialectica interpretation. In particular, the *soundness* of the interpretation of the implication connective, as expounded by Spector and Troelstra [35], in the categorical models will follow as a direct consequence of this tight correspondence. In particular, recall that such an interpretation is motivated by the equivalence:

$$(\exists u.\forall x.A_D(u,x) \to \exists v.\forall y.B_D(v,y)) \Leftrightarrow$$
$$\exists f_0, f_1.\forall u,y.(A_D(u,f_1(u,y)) \to B_D(f_0(u),y))$$

where A_D and B_D are quantifier-free formulae. Showing this equivalence requires extra logical principles, going beyond intuitionistic logic, specifically the Markov Principle (MP) and the Independence of Premise (IP) principle, as well as some choice. While the traditional categorical approach takes this equivalence as the starting point for defining categorical models, as it is done in Dialectica categories [6], for example, our approach focuses instead on abstracting, in the setting of doctrines, the key logical features that allow us to conclude such an equivalence.

We show how these key logical features are satisfied in the categorical setting, establishing a tight correspondence between the logical system and the categorical framework. Our results build on the categorical presentation of existential and universal-free predicates we introduced first in the context of fibrations in [38], and then in the language of doctrines [39]. Having a categorification of such quantifier-free predicates is fundamental to properly state logical principles in categorical terms, since both (IP) and (MP) involve quantifier-free formulae.

Finally, to complete our analysis of Gödel (hyper)doctrines, we characterise categories obtained as a result of the tripos-to-topos construction of Hyland, Johnstone, and Pitts [18] applied to these doctrines. After recalling the notions and the construction of the category of predicates associated with a hyperdoctrine from [24], and the exact completion of a lex category [5], we present an explicit characterisation of the tripos-to-topos construction associated with a Gödel hyperdoctrine. Combining our results with the characterisation of exact completions presented in [26], we show that every category obtained as tripos-to-topos construction of a Gödel hyperdoctrine can be equivalently presented as the exact completion of the (lex) category of predicates associated to the hyperdoctrine itself. We then conclude with a brief discussion of future work.

2 Dialectica Interpretation

This section recalls the definitions and results we shall need from the critical analysis of the Dialectica Interpretation provided by Troelstra in [10]. Gödel's original article [13] appeared in 1958, in the journal *Dialectica* in honour of P. Bernays. It is believed that the ideas in the paper date back at least as far as 1941, when Gödel lectured at Yale on "In what sense is intuitionistic logic constructive?". Gödel was not satisfied with the philosophical aspects of his description of the work on the interpretation and, for this reason, never returned the proofs of the English translation of the paper that he worked on and off from 1965 until his death. According to Troesltra ([10], page 219), Gödel presents his work as a "contribution to a liberalized version of Hilbert's program: to justify classical systems, in particular arithmetic, in terms of notions as intuitively clear as possible." In the original paper[13], System T is only outlined. Gödel argues that since finitistic methods are not sufficient to carry out Hilbert's program, one has to admit at least some abstract notions in a consistency proof. He suggests one can replace the notion of constructive proof by the notion of 'computable functional of finite type', which he considers more definite, less abstract, and more 'nearly finitistic'.

2.1 Interpretation of implication

Gödel's Dialectica interpretation [13, 10] associates to each formula A in the language of Heyting arithmetic HA its *Dialectica interpretation* A^D, i.e. a formula of the form:

$$A^D = \exists u.\forall x.A_D$$

where A_D is a quantifier-free formula in the language of System T, which is *as constructive as possible.* The associations $(-)^D$ and $(-)_D$ are defined inductively on the structure of the formulae, and we refer to [13, 10] for a complete description.

The Dialectica interpretation's original use was to show the consistency of Heyting arithmetic. In particular, Gödel's principal result can be stated as follows:

Theorem 2.1 (Gödel 1958)**.** *Let A be a formula in the language of Heyting arithmetic* HA*. Whenever* $\mathrm{HA} \vdash A$*, then* $\mathrm{T} \vdash A^D$ *by means of an application of the rules of the introduction of quantifiers to* A_D*, that is:*

$$\mathrm{T} \vdash A_D(t, x)$$

for some (finite sequence of) closed terms t.

Thus, the consistency of HA follows, provided that System T is consistent.

Recall that the most complicated clause of the translation is the definition of the translation of the implication connective $(A \to B)^D$. This involves logical principles which are usually not acceptable from an intuitionistic point of view, namely a form of the *Principle of Independence of Premise*, a generalisation of *Markov Principle* and some choice. The interpretation is given by:

$$(A \to B)^D := \exists V, X.\forall u, y.(A_D(u, X(u, y)) \to B_D(V(u), y)).$$

Given a witness u for the hypothesis A_D one should be able to obtain a witness for the conclusion B_D, i.e. there exists a function V assigning a witness $V(u)$ of B_D to every witness u of A_D. Moreover, this assignment has to be such that from a counterexample y of the conclusion B_D we should be able to find a counterexample $X(u, y)$ to the hypothesis A_D. This transformation of counterexamples of the conclusion into counterexamples for the hypothesis is what gives Dialectica its essential bidirectional character.

We first recall the details behind the translation of $(A \to B)^D$ showing the precise points where we have to employ the non-intuitionistic principles (MP*), (IP*) and choice. The sequence of equivalences below comes from the analysis of the Dialectica interpretation in ([10]), which Troelstra attributes to Spector. Then we explain the equivalences in detail:

$$\begin{aligned}
A^D \to B^D = \quad & \exists u.\forall x.A_D(u, x) \to \exists v.\forall y.B_D(v, y) \\
\equiv \quad & \forall u.(\forall x.A_D(u, x) \to \exists v.\forall y.B_D(v, y)) \\
\equiv \quad & \forall u.\exists v.(\forall x.A_D(u, x) \to \forall y.B_D(v, y)) \\
\equiv \quad & \forall u.\exists v.\forall y.(\forall x.A_D(u, x) \to B_D(v, y)) \\
\equiv \quad & \forall u.\exists v.\forall y.\exists x.(A_D(u, x) \to B_D(v, y)) \\
=: \quad & (A \to B)^D
\end{aligned}$$

First notice that $A^D \to B^D$, that is:

$$\exists u.\forall x.A_D(u, x) \to \exists v.\forall y.B_D(v, y) \tag{1}$$

is equivalent to:

$$\forall u.(\forall x.A_D(u, x) \to \exists v.\forall y.B_D(v, y)). \tag{2}$$

If we apply a special case of the **Principle of Independence of Premise**, namely:

$$(\forall x.C(x) \to \exists v.\forall y.D(v,y)) \to \exists v.(\forall x.C(x) \to \forall y.D(v,y)) \qquad \text{(IP*)}$$

we obtain that (2) is equivalent to:

$$\forall u.\exists v.(\forall x.A_D(u,x) \to \forall y.B_D(v,y)). \qquad (3)$$

Moreover, we can see that this is equivalent to:

$$\forall u.\exists v.\forall y.(\forall x.A_D(u,x) \to B_D(v,y)). \qquad (4)$$

The next equivalence is motivated by a generalisation of **Markov's Principle**, namely:

$$\neg\forall x.C(u,x) \to \exists x.\neg C(u,x). \qquad \text{(MP*)}$$

In particular, we obtain that (4) is equivalent to:

$$\forall u.\exists v.\forall y.\exists x.(A_D(u,x) \to B_D(v,y)). \qquad (5)$$

Recall from [1] the following argument to show the equivalence (4) $\iff$ (5), assuming that the law of excluded middle holds for B_D: if B_D is true then (4) $\iff$ (5) is justified, and if B_D is false, then apply (MP*).

To conclude that $A^D \to B^D$ and $(A \to B)^D$ are equiprovable we apply the **Axiom of Choice** (or **Skolemisation**), i.e.:

$$\forall y.\exists x.C(y,x) \to \exists V.\forall y.C(y,V(y)) \qquad \text{(AC)}$$

twice, obtaining that (5) is equivalent to:

$$\exists V,X.\forall u,y.(A_D(u,X(u,y)) \to B_D(V(u),y)).$$

This analysis (from Gödel's Collected Works, page 231) highlights the key role the principles (IP*), (MP*), and (AC) play in the Dialectica interpretation of implicational formulae. Next we examine the two principles (IP*) and (MP*), following our discussion in [38].

2.2 Independence of Premise

In proof theory, the **Principle of Independence of Premise** states that:

$$(C \to \exists u.D(u)) \to \exists u.(C \to D(u)) \qquad \text{(IP)}$$

where u is not a free variable of C. While this principle is valid in classical logic (it follows from the law of the excluded middle), it does not hold in intuitionistic logic, and it is not generally accepted constructively [1]. The principle (IP) is not generally accepted constructively because turning any proof of the premise C into a proof of $\exists u.D(u)$ means turning a proof of C into a proof of $D(t)$ where t is a witness for the existential quantifier, depending on the proof of C.

In particular, the choice of the witness *depends* on the proof of the premise C, while the (IP) principle tells us that the witness can be chosen independently of the proof of the premise C.

In the Dialectica translation we only need a particular version of (IP), namely:

$$(\forall y.C(y) \to \exists u.\forall v.D(u,v)) \to \exists u.(\forall y.C(y) \to \forall v.D(u,v)) \qquad \text{(IP*)}$$

which means that we are asking (IP) to hold not for every formula, but only for those formulas of the form $\forall y.C(y)$ with C quantifier-free. We recall a useful strengthening of the (IP*) principle, namely:

$$(C \to \exists u.D(u)) \to \exists u.(C \to D(u))$$

where C is $\exists$-free, i.e. C contains neither existential quantifiers nor disjunctions (of course, it is also assumed that u is not a free variable of C).

2.3 Markov Principle

Markov Principle is a statement that originated in the Russian school of constructive mathematics. Formally, the Markov principle is usually presented as the statement:

$$\neg\neg\exists x.C(x) \to \exists x.C(x) \qquad \text{(MP)}$$

where C is a quantifier-free formula. Thus, the logical principle employed in the Dialectica interpretation, namely:

$$\neg\forall x.C(x) \to \exists x.\neg C(x) \qquad \text{(MP*)}$$

with $C(x)$ a quantifier-free formula, can be thought of as a *generalisation* of the Markov Principle above. As remarked in [1], the Markov Principle is not generally accepted in constructive mathematics because in general there is no reasonable way to choose constructively a witness x for $\neg C(x)$ from a proof that $\forall x.C(x)$ leads to a contradiction. However, in the context of Heyting Arithmetic, i.e. when x ranges over the natural numbers, one can prove that these two formulations of the Markov Principle are equivalent. More details about the computational interpretation of the Markov Principle can be found in [28].

A natural generalisation of (MP*) is given by the following principle, that we call **Modified Markov principle**: i.e. whenever $A(y)$ is a quantifier-free predicate and $B(x)$ is an existential-free predicate, it is the case that:

$$(\forall x.B(x) \to A(y)) \to \exists x.(B(x) \to A(y)) \qquad \text{(MMP)}$$

where $A(y)$ is quantifier free, $B(x)$ is existential-free and the variable x does not occur free in $A(y)$. Notice that (MP*) is obtained from (MMP) by replacing $A(y)$ with $\bot$, the falsum constant.

We now establish the notation of the *rule-version* of the previous logical principle. Recall that logical rules are weaker than logical principles, namely if a given theory satisfies a certain logical principle then, in particular, it satisfies the rule-version of such a principle, while the converse does not hold.

Notation. To denote the rule-version of the logical principles we consider, we will add R- to the name of the principle in question. For example, we will denote by (R-IP) the rule:

$$\top \vdash \theta \to \exists u.\eta(u) \ \text{ implies } \top \vdash \exists u.(\theta \to \eta(u))$$

corresponding to the principle of independence of premise (IP):

$$\top \vdash (\theta \to \exists u.\eta(u)) \to \exists u.(\theta \to \eta(u))$$

and similarly, we will use (R-MP) for the Markov rule.

Having recalled the logical notions we need, we proceed to describe these notions using categorical logic, in the next section.

3 Lawvere Doctrines

One of the most important notions of categorical logic which enables the study of logic from a pure algebraic perspective is that of a *hyperdoctrine*, introduced in a series of seminal papers by F.W. Lawvere. Hyperdoctrines synthesise the structural properties of logical systems [19, 20, 21]. Lawvere's crucial intuition was to consider logical languages and theories as fibrations to study their 2-categorical properties, so that e.g. connectives, quantifiers, and equality are determined by structural adjunctions. Theories and models can be viewed as objects and morphisms of a suitable category, the category of hyperdoctrines, which is, in particular, a 2-category. In the 2-category of hyperdoctrines 2-cells represent morphisms of models. Thus, having a 2-categorical structure allows us not only to compare theories (objects) via models (1-cells) but also to compare models (1-cells) via the 2-cells that represent morphisms of models.

Recall from [32, 33] that a **first-order hyperdoctrine** is a contravariant functor:

$$P \colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$$

from a category with finite products $\mathcal{C}$ to the category of Heyting algebras **Hey** satisfying:

1. for every product projection $A \times B \xrightarrow{\pi_A} A$ in $\mathcal{C}$, the homomorphism $P_{\pi_A} \colon P(A) \longrightarrow P(A \times B)$ of Heyting algebras, where P_{π_A} denotes the action of the functor P on the arrow π_A, has a left adjoint $\exists_{\pi_A}$ and a right adjoint $\forall_{\pi_A}$. These adjoints satisfy the Beck-Chevalley conditions (BC), i.e. for any pullback square:

$$\begin{array}{ccc} A' \times B & \xrightarrow{f \times \mathrm{id}} & A \times B \\ {\scriptstyle \pi_{A'}}\downarrow & & \downarrow{\scriptstyle \pi_A} \\ A' & \xrightarrow[f]{} & A \end{array}$$

it is the case that the squares:

$$\begin{array}{ccc} P(A \times B) & \xrightarrow{P_{f\times \mathrm{id}}} & P(A' \times B) \\ \Big\downarrow{\scriptstyle \exists_{\pi_A}} & & \Big\downarrow{\scriptstyle \exists_{\pi_{A'}}} \\ P(A) & \xrightarrow[P_f]{} & P(A') \end{array} \quad \text{and} \quad \begin{array}{ccc} P(A \times B) & \xrightarrow{P_{f\times \mathrm{id}}} & P(A' \times B) \\ \Big\downarrow{\scriptstyle \forall_{\pi_A}} & & \Big\downarrow{\scriptstyle \forall_{\pi_{A'}}} \\ P(A) & \xrightarrow[P_f]{} & P(A') \end{array}$$

commute, i.e. the equalities:

$$P_f \exists_{\pi_A} = \exists_{\pi_{A'}} P_{f\times \mathrm{id}} \quad \text{and} \quad P_f \forall_{\pi_A} = \forall_{\pi_{A'}} P_{f\times \mathrm{id}}$$

hold. The purpose of the Beck-Chevalley conditions is to guarantee that substitution commutes with quantification, appropriately, thus BC forces the equality in both diagrams.

2. For every object A of $\mathcal{C}$ there exists a predicate δ_A of $P(A \times A)$ satisfying for every α of $P(A \times A)$ that:

$$\top \leq P_{\Delta_A}(\alpha) \text{ if and only if } \delta_A \leq \alpha$$

where $A \xrightarrow{\Delta_A} A \times A$ denotes the diagonal arrow.

A first-order hyperdoctrine determines an appropriate categorical structure to represent a first-order theory and its corresponding Tarski semantics. Semantically, a first-order hyperdoctrine is essentially a generalisation of the contravariant *powerset functor* on the category of sets:

$$\mathcal{P}\colon \mathbf{Set}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$$

sending a set A into the Heyting algebra $\mathcal{P}(A)$ of its subsets (ordered by inclusion), and a set-theoretic function $A \xrightarrow{f} B$ to the inverse image functor $\mathcal{P}B \xrightarrow{\mathcal{P}f=f^{-1}} \mathcal{P}A$. In this case, the adjoints $\forall_f$ and $\exists_f$ must be evaluated, on a subset D of A, respectively as the subsets $\exists_f(D) = \{a \in B \mid \exists a \in A.\ (b = f(a) \wedge a \in D)\}$ and $\forall_f(D) = \{a \in B \mid \forall a \in A.\ (b = f(a) \Rightarrow a \in D)\}$.

From a syntactic point of view, a first-order hyperdoctrine can be seen as a generalisation of the *Lindenbaum-Tarski algebra* of well-formed formulae of a first-order theory. In particular, given a first-order theory $\textsc{Th}$ in a many-sorted first-order language $\mathcal{L}$, one can consider the functor:

$$\mathcal{L}^{\textsc{Th}}\colon \mathcal{V}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$$

whose base category $\mathcal{V}$ is the *syntactic* category of $\mathcal{L}$, i.e. the category whose objects are (α-equivalence classes of) finite lists $\overrightarrow{x} := [x_1 : X_1, \ldots, x_n : X_n]$ of typed variables and whose morphisms are lists of substitutions, while the predicates of $\mathcal{L}^{\mathrm{TH}}(\overrightarrow{x})$ are given by equivalence classes (with respect to provable reciprocal consequence $\dashv\vdash$) of well-formed formulae in the context $\overrightarrow{x}$, and order is given by the provable consequences, according to the fixed theory TH. In this case, the left adjoint to the weakening functor $\mathcal{L}^{\mathrm{TH}}_{\pi}$ is computed by existentially quantifying the variables that are not involved in the substitution induced by the projection (dually, the right adjoint is computed by quantifying universally).

3.1 Existential and universal doctrines

Recently, several generalisations of the notion of a Lawvere hyperdoctrine were considered, and we refer for example to [23, 24, 25] or to [33, 18] for higher-order versions. In this work, we consider a natural generalisation of the notion of first-order hyperdoctrine, and we call it simply a *doctrine*.

Definition 3.1. A **doctrine** is a contravariant functor:

$$P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$$

where the category $\mathcal{C}$ has finite products and **Pos** is the category of posets.

Now we recall from Maietti and Rosolini [23, 24, 36] the notions of existential and universal doctrines.

Definition 3.2. A doctrine $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ is called **existential** (respectively **universal**) if, for every A_1 and A_2 in $\mathcal{C}$ and every product projection $X \xrightarrow{\pi} A$, the functor:

$$P(A) \xrightarrow{P_\pi} P(X)$$

has a left adjoint $\exists_\pi$ (respectively a right adjoint $\forall_\pi$), and these satisfy the Beck-Chevalley condition BC: for any pullback diagram:

$$\begin{array}{ccc} X' & \xrightarrow{\pi'} & A' \\ {\scriptstyle f'}\downarrow & & \downarrow{\scriptstyle f} \\ X & \xrightarrow[\pi]{} & A \end{array}$$

where π and π' are projections, and for any β in $P(X)$ the equality:

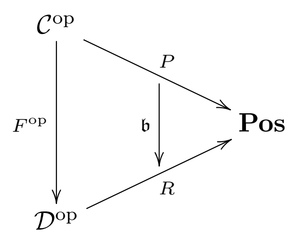
$$\exists_{\pi'} P_{f'}\beta = P_f \exists_\pi \beta \quad (\text{ resp. } \forall_{\pi'} P_{f'}\beta = P_f \forall_\pi \beta\ )$$

holds.

Observe that the inequality $\exists_{\pi'} P_{f'}\beta \leq P_f \exists_\pi \beta$ (resp. $\forall_{\pi'} P_{f'}\beta \geq P_f \forall_\pi \beta$) of BC in Definition 3.2 always holds.

We recall from [23, 24, 25] that doctrines form a 2-category Doc where:

- a **1-cell** is a pair $(F, \mathfrak{b})$:



  such that $\mathcal{C} \xrightarrow{F} \mathcal{D}$ is a finite product preserving functor between the cartesian categories $\mathcal{C}$ and $\mathcal{D}$, and:

$$P \xrightarrow{\mathfrak{b}} RF^{\text{op}}$$

  is a natural transformation;

- a **2-cell** $(F, \mathfrak{b}) \xrightarrow{\theta} (G, \mathfrak{c})$ is a natural transformation $F \xrightarrow{\theta} G$ such that for every A in $\mathcal{C}$ and every α in $P(A)$, we have:

$$\mathfrak{b}_A(\alpha) \leq R_{\theta_A}(\mathfrak{c}_A(\alpha)).$$

We denote as ExD the 2-full subcategory of Doc whose objects are existential doctrines, and whose 1-cells are those 1-cells of Doc which preserve the existential structure, namely those 1-cells commuting with left adjoins $\exists_\pi$. Similarly, we denote by UnD the 2-full subcategory of Doc whose objects are universal doctrines, and whose 1-cells are those 1-cells of Doc which preserve the universal structure.

From a logical perspective, the intuition is that a 1-cell between doctrines is a generalisation of the notion of a set-theoretic model, whereas 2-cells represent morphisms of models.

3.2 Internal language of doctrines

A milestone in the history of mathematical logic is Gödel's proof of the completeness property of Tarski's semantics (in its usual set-theoretical formulation) for classical first-order logic. It tells us that a given first-order theory proves a given first-order statement if (and only if) every set-theoretic model of the given theory happens to be also a model of the given statement.

As long as we are interested in (fragment of many-sorted) intuitionistic first-order logic, this notion of semantics can be generalised allowing interpretations in classes of categories and categorical structures that generalise to some extent the properties of the category of sets. The notion of hyperdoctrine is suitable to this purpose: different weakenings of this notion can host interpretations (hence models) of different fragments of many-sorted intuitionistic first-order logic according to a notion of semantics that formally coincides with Tarski's set-theoretic one for classical logic. In these generalised frameworks, a soundness and a completeness result continue existing: a correspondence between provability and satisfiability holds.

The characteristic category-theoretic form of proof by diagram chasing to establish properties expressible in category-theoretic terms, can in complex cases be difficult to construct and hard to follow because of the rather limited forms of expression of a purely category-theoretic language. Categorical logic enables the use of richer and more familiar forms of expression meant to establish properties of particular kinds of categories and categorical structures, depending on the fragment of intuitionistic logic that those structures model according to generalised Tarski's semantics. In fact, one can define a suitable notion of *internal language*, naming the relevant constituents of the category and then applying the corresponding categorical semantics to turn assertions of this language (according to the corresponding fragment of intuitionistic logic) into categorical statements. Such a procedure has become highly developed e.g. in the theory of toposes where the internal language of a topos coupled with the generalised semantics of the whole intuitionistic first-order logic in toposes enables one to reason about the objects and morphisms of a topos as if they were sets and functions (see [27]). The notions of internal language and generalised semantics are not just a useful tool to simplify notation, but also a powerful instrument to formally prove a categorical equivalence between doctrines (in our case) and theories.

First, we briefly recall that theories in a given (possibly many-sorted) language over a fragment of first-order logic induce doctrines. Let us assume that we are given a fragment F of first-order logic. Whenever $\mathcal{L}$ is a (possibly) many sorted F-language and $\textsc{Th}$ is an $\mathcal{L}$-theory, we can define a doctrine $\mathcal{L}^{\textsc{Th}}$ over the syntactic category $\mathcal{V}$ associated to $\mathcal{L}$, as described at the beginning of the current section. Depending on F, the doctrine $\mathcal{L}^{\textsc{Th}}$ has several categorical properties. For example, if F is regular, then the fibres of $\mathcal{L}^{\textsc{Th}}$ are inf-semilattices and $\mathcal{L}^{\textsc{Th}}$ has left-adjoints to pullbacks along product projections and diagonals (satisfying both the Back-Chevalley condition and Frobenius reciprocity, see [36] for more details).

Whenever P is a doctrine over some category $\mathcal{C}$ *such that* P *has the same categorical properties of* $\mathcal{L}^{\textsc{Th}}$, then P can host *models* of $\textsc{Th}$ according to a natural generalisation of Tarski's semantics, which is sound and complete (for it admits the syntactic model) and is formally defined as classical Tarskian semantics for (a fragment of) first-order logic. In fact, an $\mathcal{L}$-structure S in P consists of:

- an object of $\mathcal{C}$ for every $\mathcal{L}$-sort;
- an arrow of $\mathcal{C}$ (between the appropriate S-interpretations of the sorts) for any $\mathcal{L}$-function symbol;
- for every $\mathcal{L}$-predicate symbol in some context, a predicate of the P-fibre of the S-interpretation of that context.

Then (terms and) formulas are inductively interpreted in S as usual (formally as for traditional Tarski's semantics). While we are not giving a full inductive definition of the interpretation ϕ^S of a $\mathcal{L}$-formula ϕ in S, we recall that this notion can be written recursively on the complexity of ϕ itself, and we provide one of the inductive clauses as an instance, assuming that F is regular:

> if $\phi \equiv \exists b.\psi(a,b)$ where ψ is a formula in context $a : A, b : B$; if the objects A^S and B^S of $\mathcal{C}$ are the S-interpretations of the sorts A and B in $\mathcal{C}$; and if the subobject ψ^S of $A^S \times B^S$ in $\mathcal{C}$ is the S-interpretation of the formula ψ; then we define the S-interpretation ϕ^S of ϕ as the subobject $\exists_{\pi_A}\psi$ of A^S, where π_A the product projection $A \times B \to A$.

For a complete definition of the S-interpretation of ϕ we refer the reader to [27, 34]. Finally, we say that S **models** some $\mathcal{L}$-sequent $\phi \vdash \psi$ in some given context when it is the case that:

$$\phi^S \leq \psi^S$$

in the P-fibre of the S-interpretation of the given context. Therefore, the $\mathcal{L}$-structure S in P is a **model** of $\textsc{Th}$ if it is a model of every sequent that $\textsc{Th}$ proves.

According to this notion of semantics, it is the case that the P-models of $\textsc{Th}$, together with the model morphisms between them, are bijectively (equivalently) induced by the 1-cells $\mathcal{L}^{\textsc{Th}} \to P$ of Doc and the 2-cells between them, respectively (see Section 3.1 for the notions of 1-cell and 2-cell). Thus, the identity over $\mathcal{L}^{\textsc{Th}}$ constitutes the syntactic model of $\textsc{Th}$.

Conversely, if $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ is a doctrine satisfying those categorical properties that allow the interpretation in P of formulas of any F-language (e.g. if F is regular, then the fibres of P are required to be inf-semilattices and P has left-adjoints to pullbacks along product projections and diagonals satisfying Beck-Chevalley condition and Frobenius reciprocity), then an F-language $\mathcal{L}_P$ can be defined starting from P itself: the language $\mathcal{L}_P$ has a sort A for every object A of the base category $\mathcal{C}$, an n-ary function symbol $A_1, \ldots, A_n \xrightarrow{f} A$ for every morphism $A_1 \times \cdots \times A_n \xrightarrow{f} A$ of $\mathcal{C}$ and an n-relation symbol $R : A_1, \ldots, A_n$ for each element of $P(A_1 \times \cdots \times A_n)$, all of this for each finite list of objects $A_1, \ldots, A_n$ of $\mathcal{C}$ and every object A of $\mathcal{C}$. The terms and the formulas of $\mathcal{L}_P$ are the ones inductively generated as usual by applying to this signature the first-order symbols admitted by F. The language $\mathcal{L}_P$ is called the **internal language** of the doctrine P.

Let $\textsc{Th}_P$ be the theory whose sequents $\phi \vdash \psi$ in some context A are precisely those such that $\phi \leq \psi$ in $P(A)$. The doctrine $\mathcal{L}_P^{\textsc{Th}_P}$ is equivalent to P in Doc (see [34]) and the assignment $P \mapsto (\mathcal{L}_P, \textsc{Th}_P)$ extends to a pseudo 2-inverse to the 2-functor $(\mathcal{L}, \textsc{Th}) \mapsto \mathcal{L}^{\textsc{Th}}$, in such a way that Doc is equivalent to the 2-category of the theories in some language over F together with the models (of one of them into the other one) and the model morphisms (between them).

Whenever $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ is a doctrine, we know that $\mathcal{L}_P^{\mathrm{T{\scriptsize H}}_P}$ is equivalent to P in Doc (see [34]). Therefore, the doctrine P, together with the equivalence $\mathcal{L}_P^{\mathrm{T{\scriptsize H}}_P} \to P$, constitutes the syntactic model of its own theory $\mathrm{T{\scriptsize H}}_P$ in its own internal language $\mathcal{L}_P$. This fact means that, whenever ϕ and ψ are elements of $P(A)$, for some object A of $\mathcal{C}$, it is the case that $\phi \leq \psi$ precisely when $\phi \vdash \psi$ is a sequent of $\mathrm{T{\scriptsize H}}_P$ in context A. This is precisely why we can deduce properties of P through a purely syntactical procedure: every $\mathcal{L}_P$-sequent corresponds to a categorical statement or a condition involving P, and this is true precisely when that sequent belongs to $\mathrm{T{\scriptsize H}}_P$.

Taking advantage of this equivalence between satisfiability and provability for a doctrine P, from now on we write the assertions relative to P translated in its internal language $\mathcal{L}_P$ and, instead of deducing one of these assertions from other given ones by explicitly using the categorical properties of P, we usually employ the inference rules of the fragment F that these properties model, in order to achieve the same result. For example, if the fibres of P are inf-semilattices and P has left-adjoints to pullbacks along product projections and diagonals satisfying the Beck-Chevalley condition and Frobenius reciprocity, then F is regular, hence we are allowed to deduce assertions in $\mathcal{L}_P$ by means of the inference rules of many-sorted regular logic. We define the following notation for the internal language $\mathcal{L}_P$ of P. We write:

$$a_1 : A_1, \ldots, a_n : A_n \mid \phi(a_1, \ldots, a_n) \vdash \psi(a_1, \ldots, a_n)$$

to indicate the sequent in context associated to the inequality:

$$\phi \leq \psi$$

of the fibre $P(A_1 \times \cdots \times A_n)$. In particular, we write:

$$a : A \mid \exists b : B.\psi(a, b) \text{ and } a : A \mid \forall b : B.\psi(a, b)$$

to indicate the formulas in context associated to the predicates:

$$\exists_{\pi_A}\psi \text{ and } \forall_{\pi_A}\psi$$

in the fibre $P(A)$, where π_A is the projection $A \times B \to A$. In fact, as we said before regarding the notion of generalised interpretation of a formula, if a doctrine $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ is existential and $\alpha \in P(A \times B)$ is a formula in context:

$$a : A, b : B \mid \alpha(a, b)$$

then $\exists_{\pi_A}\alpha \in PA$ represents the formula $a : A \mid \exists b : B.\alpha(a, b)$. Analogously, if the doctrine P is universal, then $\forall_{\pi_A}\alpha \in PA$ represents the formula $a : A \mid \forall b : B.\alpha(a, b)$ in context A. The fact that this notion of interpretation is sound and complete is not surprising: this is how usual set-theoretic Tarski's semantics can be characterised in terms of categorical properties of the powerset functor $\mathcal{P}\colon \mathbf{Set}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$. This is what we mean when we say that generalised Tarski's semantics is formally identical to the classic set-theoretic one.

We write $a : A \mid \phi \dashv\vdash \psi$ to abbreviate $a : A \mid \phi \vdash \psi$ and $a : A \mid \psi \vdash \phi$. Moreover, when the type of a quantified variable is clear from the context, we will omit that type for the sake of readability. The substitution via given terms (i.e. reindexing and weakening) is modelled by pulling back along those given terms. The application of propositional connectives is interpreted by using the corresponding operations in the fibres of the given doctrine.

4 Towards Gödel doctrines

One of the fundamental notions of logic and proof theory is the notion of *quantifier-free formula*, and there are countless results built on the possibility of detecting quantifier-free formulae in the literature. For example, in the Dialectica interpretation, this notion is present at every stage, and we could say the entire translation depends on the fact that, syntactically, we can identify and distinguish formulae with no occurrences of quantifiers.

However, while from a syntactic perspective it is effortless and natural to speak of quantifier-free formulae, describing this notion algebraically is not so obvious. The main problem is that the property of being quantifier-free is totally syntactic, not involving any other entity different from the formula itself we are considering. It does not depend, for example, on the fact that we are working in classical, constructive, or intuitionistic logic. It only depends on how a formula is written in a given formal language.

Therefore, if we want to provide a complete categorical presentation of the Dialectica interpretation, able to represent all its logical details, we have to deal with the problem of representing quantifier-free formulae. We need to find a suitable *universal property* to represent predicates that are quantifier-free categorically. Notice that quantifier-free predicates may satisfy different properties depending on the logical system we are considering. Hence if we want to represent these predicates via universal properties, we have to relativise this notion to a given system, that is to the Dialectica interpretation in our case.

4.1 Existential and universal free predicates

We discuss how to identify those predicates of an existential doctrine:

$$P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$$

which are *free from left-adjoints* $\exists_\pi$, and then dualise this notion to define those predicates that are *free from right-adjoints* $\forall_\pi$. This idea was originally introduced by Trotta and Maietti [26] and, independently, by Frey in [11]. It was further developed and generalised to the fibrational setting in [38].

Definition 4.1. Let $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ be an existential doctrine and let A be an object of $\mathcal{C}$. A predicate α of the fibre $P(A)$ is said to be an **existential splitting** if it satisfies the following *weak* universal property: for every predicate

$\beta \in P(A \times B)$ such that $\alpha(a) \vdash \exists b : B.\beta(a,b)$ (i.e. $\alpha \leq \exists_{\pi_A}(\beta)$ in category-theoretic notation, where $A \times B \xrightarrow{\pi_A} A$ is a product projection of $\mathcal{C}$), there exists an arrow $A \xrightarrow{g} B$ such that:

$$\alpha(a) \vdash \beta(a, g(a)) \quad (\text{ i.e. } \alpha \leq P_{\langle 1_A, g\rangle}(\beta) \text{ in category-theoretic notation }).$$

Existential splittings stable under reindexing are called *existential-free predicates*. We introduce the following definition:

Definition 4.2. Let $P\colon \mathcal{C}^{\text{op}} \longrightarrow \mathbf{Pos}$ be an existential doctrine and let I be an object of $\mathcal{C}$. A predicate α of the fibre $P(I)$ is said to be **existential-free** if $P_f(\alpha)$ is an existential splitting for every morphism $A \xrightarrow{f} I$.

Employing the presentation of doctrines via internal language, we say that $i : I \mid \alpha(i)$ is *existential-free* if, whenever $a : A \mid \alpha(f(a)) \vdash \exists b : B.\beta(a,b)$ for some term $a : A \mid f(a) : I$, then there is a term $a : A \mid g(a) : B$ such that $a : A \mid \alpha(f(a)) \vdash \beta(a, g(a))$.

Observe that we always have that $a : A \mid \beta(a, g(a)) \vdash \exists b : B.\beta(a,b)$, in other words $P_{\langle 1_A, g\rangle}\beta \leq \exists_{\pi_A}\beta$. In fact, it is the case that $\beta \leq P_{\pi_A}\exists_{\pi_A}\beta$ (as this arrow of $P(A \times B)$ is nothing but the unit of the adjunction $\exists_{\pi_A} \dashv P_{\pi_A}$), hence a re-indexing by the term $\langle 1_A, g\rangle$ yields the desired inequality. Therefore, the property that we require for $i : I \mid \alpha(i)$ turns out to be the following: whenever there are proofs of $\exists b : B.\beta(a,b)$ from $\alpha(f(a))$, at least one of them factors through the canonical proof of $\exists b : B.\beta(a,b)$ from $\beta(a, g(a))$ for some term $a : A \mid g(a) : B$.

Requiring the stability under substitution as in Definition 4.2 is clear since, in logic, if a formula is existential-free, and we apply a substitution to this formula, then we obtain again an existential-free formula.

Definition 4.3. Let $P\colon \mathcal{C}^{\text{op}} \longrightarrow \mathbf{Pos}$ be an existential doctrine. Then we indicate by $P^{\exists\text{-free}}\colon \mathcal{C}^{\text{op}} \longrightarrow \mathbf{Pos}$ the subdoctrine of P whose predicates of the fibres $P^{\exists\text{-free}}(A)$ are existential-free predicates of $P(A)$.

Dualising the previous Definitions 4.1 and 4.2 we get the corresponding ones for the universal quantifier.

Definition 4.4. Let $P\colon \mathcal{C}^{\text{op}} \longrightarrow \mathbf{Pos}$ be a universal doctrine and let A be an object of $\mathcal{C}$. A predicate α of the fibre $P(A)$ is said to be a **universal splitting** if it satisfies the following weak universal property: for every predicate $\beta \in P(A \times B)$ such that $\forall b : B.\beta(a,b) \vdash \alpha(a)$, there exists an arrow $A \xrightarrow{g} B$ such that:

$$\beta(a, g(a)) \vdash \alpha(a).$$

Definition 4.5. Let $P\colon \mathcal{C}^{\text{op}} \longrightarrow \mathbf{Pos}$ be a universal doctrine and let I be an object of $\mathcal{C}$. A predicate α of the fibre $P(I)$ is said to be **universal-free** if $P_f(\alpha)$ is a universal splitting for every morphism $A \xrightarrow{f} I$.

Again, employing the presentation of a doctrine via its internal language, the property we require of the formula $i : I \mid \alpha(i)$, so that it is *universal-free*, is that whenever $a : A \mid \forall b : B.\beta(a,b) \vdash \alpha(f(a))$ for some term $a : A \mid f(a) : I$, then there is a term $a : A \mid g(a) : B$ such that $a : A \mid \beta(a, g(a)) \vdash \alpha(f(a))$.

Definition 4.6. Let $P \colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ be an existential doctrine. We say that P has **enough existential-free predicates** if, for every object I of $\mathcal{C}$ and every predicate $\alpha \in P(I)$, there exist an object A and an existential-free object β in $P(I \times A)$ such that $\alpha(i)$ is equiprovable with $\exists a : A.\beta(i,a)$ (i.e. $\alpha = \exists_{\pi_I}\beta$).

Analogously, we have the following definition for universal doctrines.

Definition 4.7. Let $P \colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ be a universal doctrine. We say that P has **enough universal-free predicates** if, for every object I of $\mathcal{C}$ and every predicate $\alpha \in PI$, there exist an object A and a universal-free object β in $P(I \times A)$ such that $\alpha(i) \dashv\vdash \forall a : A.\beta(i,a)$.

Definition 4.8. Let $P \colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ be a universal doctrine. Then we indicate by $P^{\forall\text{-free}} \colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ the subdoctrine of P whose predicates of the fibres $P^{\forall\text{-free}}(A)$ are universal-free predicate of $P(A)$.

4.2 Skolem and Gödel doctrines

Building over the notions corresponding to quantifier-free predicates in doctrines we introduced in the previous section, we now define two particular kinds of doctrines, called *Skolem doctrines* and *Gödel doctrines*.

The *Skolem doctrines* satisfy a version of the traditional principle of *Skolemisation*, namely:

$$\forall u \exists x \alpha(u,x) \to \exists f \forall u \alpha(u, fu).$$

The name *Gödel doctrine* is chosen because we will prove that these doctrines encapsulate the basic mathematical features of Gödel's Dialectica interpretation.

Definition 4.9. A doctrine $P \colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ is called a **Skolem doctrine** if:

(i) the category $\mathcal{C}$ is cartesian closed;

(ii) the doctrine P is existential and universal;

(iii) the doctrine P has enough existential-free predicates;

(iv) the existential-free objects of P are stable under universal quantification, i.e. if $\alpha \in P(A)$ is existential-free, then $\forall_\pi(\alpha)$ is existential-free for every projection π from A.

Remark 4.10. The last point (iv) of Definition 4.9 implies that, given a Skolem doctrine $P \colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$, the sub-doctrine $P^{\exists\text{-free}} \colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ of existential-free predicates of P as defined in 4.3 is a universal doctrine. From a purely logical perspective, requiring existential-free predicates to be stable under universal quantification is quite natural since this can be also read as *if* $\alpha(x)$ *is an existential-free formula, then* $\forall x.\alpha(x)$ *is again an existential-free formula.*

Proposition 4.11 (Skolemisation)**.** *Every Skolem doctrine* $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ *validates the Skolemisation principle:*

$$a : A \mid \forall b : B.\exists c : C.\alpha(a,b,c)$$

is equiprovable with:

$$a : A \mid \exists f : C^B.\forall b : B.\alpha(a,b,\mathrm{ev}(f,b))$$

where α *is any predicate in* $P(A \times B \times C)$*.*

Proof. Applying the properties of right and left adjoints, it is straightforward to check that $a : A \mid \exists f : C^B.\forall b : B.\alpha(a,b,\mathrm{ev}(f,b)) \vdash \forall b : B.\exists c : C.\alpha(a,b,c)$. Thus, we need to prove the converse.

Let us assume that $a : A \mid \gamma(a) \vdash \forall b.\exists c.\alpha(a,b,c)$ for some predicate $\gamma \in P(A)$. By point (iv) of Definition 4.9, we assume without loss of generality that $\gamma(a)$ is existential-free: otherwise, there is an existential-free predicate γ' covering $\gamma(a)$ and we get back to our hypothesis by using that P is existential.

Since P is universal, it is the case that $a : A, b : B \mid \gamma(a) \vdash \exists c.\alpha(a,b,c)$ and, being $\gamma(a)$ existential-free:

$$a : A, b : B \mid \gamma(a) \vdash \alpha(a,b,g(a,b))$$

for some term in context $a : A, b : B \mid g(a,b) : C$. Since $\mathcal{C}$ is cartesian closed, there is a context $f : C^B$ together with a term in context $f : C^B, b : B \mid \mathrm{ev}(f,b) : C$ such that there is a unique term in context $a : A \mid h(a) : C^B$ satisfying $a : A, b : B \mid g(a,b) = \mathrm{ev}(h(a),b) : C$. Hence:

$$a : A, b : B \mid \gamma(a) \vdash \alpha(a,b,\mathrm{ev}(h(a),b))$$

and P being universal, it is the case that:

$$a : A \mid \gamma(a) \vdash \forall b.\alpha(a,b,\mathrm{ev}(h(a),b)).$$

Finally, since:

$$a : A \mid \forall b.\alpha(a,b,\mathrm{ev}(h(a),b)) \vdash \exists f.\forall b.\alpha(a,b,\mathrm{ev}(f,b))$$

(this holds for any predicate $\delta(a,-)$ in place of the predicate $\forall b.\alpha(a,b,\mathrm{ev}(-,b))$) we conclude that:

$$a : A \mid \gamma(a) \vdash \exists f.\forall b.\alpha(a,b,\mathrm{ev}(f,b)).$$

We are done by taking $\forall b.\exists c.\alpha(a,b,c)$ as the predicate $\gamma(a)$. □

Example 4.12 (Hilbert epsilon calculus)**.** Let us consider the intuitionistic Hilbert's ϵ-τ calculus $\mathcal{H}$. In this framework we mean that $\mathcal{H}$ is a many-sorted intuitionistic first-order logic whose sorts satisfy simply typed lambda calculus, together with the following additional rules (observe that 1. and 2. would happen to be equivalent in classical first-order logic):

1. for every formula in context $a : A, b : B \mid \gamma(a, b)$ there is a choice of a term in context $a : A \mid \epsilon_\gamma(a) : B$ called ϵ-*operator*, such that:

$$a : A \mid \exists b.\gamma(a, b) \vdash \gamma(a, \epsilon_\gamma(a));$$

2. for every formula in context $a : A, b : B \mid \gamma(a, b)$ there is a choice of a term in context $a : A \mid \tau_\gamma(a) : B$ called τ-*operator*, such that:

$$a : A \mid \gamma(a, \tau_\gamma(a)) \vdash \forall b.\gamma(a, b).$$

For more details, we refer the reader to [2, 8]. Let us consider the syntactic doctrine $H : \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ associated to $\mathcal{H}$, that is:

- the category $\mathcal{C}$ is the one whose objects A are the $\mathcal{H}$-contexts and whose arrows are the $\mathcal{H}$-context morphisms i.e. the $\mathcal{H}$-substitutions;
- the poset $H(A)$ contains the well-formed $\mathcal{H}$-formulas $a : A \mid \phi(a)$. It is the case that $a : A \mid \phi(a)$ is smaller or equal than $a : A \mid \psi(a)$ precisely when the sequent:

  $$a : A \mid \phi(a) \vdash \psi(a)$$

  can be inferred in $\mathcal{H}$.

Then H is existential and universal, since it is a model of intuitionistic first-order logic. Moreover, $\mathcal{C}$ is cartesian closed, as the sorts of $\mathcal{H}$ model the simply typed lambda calculus.

Let $i : I \mid \alpha(i)$ be any predicate of H and let $A \xrightarrow{f} I$ be any substitution in $\mathcal{C}$. If $a : A, b : B \mid \beta(a, b)$ is a predicate such that:

$$\alpha(f(a)) \vdash \exists b : B.\beta(a, b)$$

then it is the case that $\exists b : B.\beta(a, b) \vdash \beta(a, \epsilon_\alpha(a))$, hence $\alpha(f(a)) \vdash \beta(a, \epsilon_\alpha(a))$. This proves that the predicate $i : I \mid \alpha(i)$ is existential-free, and $i : I \mid \alpha(i)$ was arbitrary. This implies:

- the existential-free predicates are stable under universal quantification;
- the doctrine H has enough existential-free predicates: if $a : A \mid \alpha(a)$ is any predicate, then its weakening $a : A, a' : A \mid \alpha(a)$ is existential-free and:

  $$a : A \mid \exists a'.\alpha(a) \dashv\vdash \alpha(a)$$

  where $\alpha(a)$ in the right-hand side is the result of substituting the term in context $a : A \mid \epsilon_\alpha(a) : A$ in the a'-component of the predicate $a : A, a' : A \mid \alpha(a)$;

Therefore H is a Skolem doctrine.

Now we introduce the main notion of this section:

Definition 4.13. A doctrine $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ is called a **Gödel doctrine** if:

(i) P is a Skolem doctrine;

(ii) the sub-doctrine $P^{\exists\text{-free}}\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ of the existential-free predicates of P has enough universal-free predicates.

As we prove in Theorem 4.17, the condition (ii) of Definition 4.13 is crucial to show that in a Gödel doctrine every formula admits a presentation of the 'quantifier' form ($\exists u\forall x.\alpha(u,x)$, where α is quantifier-free) used in the Dialectica translation.

Now we have all the tools needed to introduce the notion of *quantifier-free predicate* in the categorical setting of Gödel doctrines.

Definition 4.14. A predicate α of a fibre $P(A)$ of a Gödel doctrine P that is both an existential-free predicate of P and a universal-free predicate in the sub-doctrine $P^{\exists\text{-free}}$ of existential-free predicates of P is called a **quantifier-free predicate** of P. The sub-doctrine of quantifier-free predicates is denoted by $P^{\exists\forall\text{-free}}\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$.

Therefore, given a Gödel doctrine $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$, we have the following canonical inclusions of doctrines:

$$P^{\exists\forall\text{-free}} \xhookrightarrow{\iota_1} P^{\exists\text{-free}} \xhookrightarrow{\iota_2} P$$

where $P^{\exists\forall\text{-free}} \xhookrightarrow{\iota_1} P^{\exists\text{-free}}$ is a morphism of doctrines, while $P^{\exists\text{-free}} \xhookrightarrow{\iota_2} P$ is a morphism of universal doctrines.

Remark 4.15. Notice that a predicate of a Gödel doctrine that happens to be both existential-free and universal-free is in particular quantifier-free in the sense of Definition 4.14. However, a universal-free predicate of the sub-doctrine $P^{\exists\text{-free}}$ of a given Gödel doctrine P may not be a universal-free predicate in the whole doctrine P, because the universal property of being universal-free is relative only to the predicates of $P^{\exists\text{-free}}$. Therefore, the quantifier-free predicates of P as established in Definition 4.14 *are not the existential and universal free predicates* of P.

Example 4.16. The doctrine defined in Example 4.12 is a Gödel doctrine. In fact, by using the choice function τ analogously to how we used ϵ in Example 4.12, one can prove that every predicate of H is universal-free, which means that the sub-doctrine $H^{\exists\text{-free}}$ of the existential-free predicates of H (i.e. H itself) has enough universal-free predicates.

To simplify the notation and make clear the connection with the logical presentation of the Dialectica interpretation, for a given Gödel doctrine $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ we will use the notation α_D to indicate a predicate α of $P^{\exists\forall\text{-free}}$, i.e. a quantifier-free predicate.

Theorem 4.17 (Prenex normal form)**.** *Let $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ be a Gödel doctrine, and let α be a predicate of $P(I)$. Then there exists a quantifier-free predicate α_D of $P(I \times U \times X)$ such that:*

$$i : I \mid \alpha(i) \dashv\vdash \exists u : U.\forall x : X.\alpha_D(i, u, x).$$

Proof. By definition of Gödel doctrine, since the doctrine P has enough existential free objects, there exists an existential-free predicate $\beta \in P(I \times U)$ such that $i : I \mid \alpha(i)$ and $\exists u : U.\beta(i, u)$ are equiprovable. Then, since the subdoctrine of existential-free predicates has enough-universal free predicates, we can conclude that there exists a quantifier-free predicate α_D of $P(I \times U \times X)$ such that $i : I \mid \alpha(i)$ is equiprovable with $\exists u : U.\forall x : X.\alpha_D(i, u, x)$. □

The next result establishes the precise connection between Gödel doctrines and the Dialectica interpretation. Employing the properties of a Gödel doctrine, we can provide a complete categorical presentation of the chain of equivalences involved in the Dialectica interpretation of implicational formulae, described in Section 2.1. In particular, we show that the crucial steps where the principles (IP*) and (MP*) are applied are represented categorically via the notions of existential-free predicate and universal-free predicate.

Theorem 4.18 (Dialectica functionals categorically)**.** *Let $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ be a Gödel doctrine. Then for every $A_D \in P(I \times U \times X)$ and $B_D \in P(I \times V \times Y)$ quantifier-free predicates of P we have that:*

$$i : I \mid \exists u.\forall x.A_D(i, u, x) \vdash \exists v.\forall y.B_D(i, v, y)$$

if and only if there exist $I \times U \xrightarrow{f_0} V$ and $I \times U \times Y \xrightarrow{f_1} X$ such that:

$$i : I, u : U, y : Y \mid A(i, u, f_1(i, u, y)) \vdash B_D(i, f_0(i, u), y).$$

Proof. Let us consider two quantifier-free predicates $A_D \in P(I \times U \times X)$ and $B_D \in P(I \times V \times Y)$ of the Gödel doctrine P. The equivalence (where for the sake of readability we omit the types of quantified variables):

$$\begin{gathered} i : I \mid \exists u.\forall x.A_D(i, u, x) \vdash \exists v.\forall y.B_D(i, v, y) \iff \\ i : I, u : U \mid \forall x.A_D(i, u, x) \vdash \exists v.\forall y.B_D(i, v, y) \end{gathered}$$

follows from the definition of left adjoint functor. As the predicate $\forall x.A_D(i, u, x)$ is existential-free in P, it is the case that:

$$\begin{gathered} i : I, u : U \mid \forall x.A_D(i, u, x) \vdash \exists v.\forall y.B_D(i, v, y) \iff \\ i : I, u : U \mid \forall x.A_D(i, u, x) \vdash \forall y.B_D(i, f_0(i, u), y) \end{gathered}$$

for some arrow $I \times U \xrightarrow{f_0} V$. Therefore:

$$\begin{gathered} i : I, u : U \mid \forall x.A_D(i, u, x) \vdash \forall y.B_D(i, f_0(i, u), y) \iff \\ i : I, u : U, y : Y \mid \forall x.A_D(i, u, x) \vdash B_D(i, f_0(i, u), y) \end{gathered}$$

since the universal quantification is right adjoint to the weakening functor. Now we employ the fact that $B_D(i, f_0(u), y)$ is universal-free in the subdoctrine of existential-free predicates of P. Since $A_D(i, u, x)$ is a quantifier-free predicate of P, it is the case that $\forall x.A_D(i, u, x)$ is existential free: this follows from the fact that in every Gödel doctrine, existential-free predicates are stable under universal quantification (this is the last point of Definition 4.13). Therefore:

$$i : I, u : U, y : Y \mid \forall x.A_D(i, u, x) \vdash B_D(i, f_0(i, u), y) \iff$$
$$i : I, u : U, y : Y \mid A_D(i, u, f_1(i, u, y)) \vdash B_D(i, f_0(i, u), y)$$

for some arrow $I \times U \times Y \xrightarrow{f_1} X$ of $\mathcal{C}$. Combining the first and the last equivalences, we obtain that:

$$i : I \mid \exists u.\forall x.A_D(i, u, x) \vdash \exists v.\forall y.B_D(i, v, y) \iff \text{there exist } (f_0, f_1) \text{ such that}$$
$$i : I, u : U, y : Y \mid A_D(i, u, f_1(i, u, y)) \vdash B_D(i, f_0(i, u), y)$$

and we are done. □

Theorem 4.18 shows that the notion of Gödel doctrine encapsulates the basic mathematical feature of the Dialectica interpretation, namely its interpretation of implication, which corresponds to the existence of functionals of types $f_0 : I \times U \to V$ and $f_1 : I \times U \times Y \to X$ as described. One should think of this as saying that a proof of a formula of the form $\exists u.\forall x.A_D(i, u, x) \to \exists v.\forall y.B_D(i, v, y)$ is obtained by transforming it to:

$$\forall u.\exists v.\forall y.\exists x.(A_D(i, u, x) \to B_D(i, v, y))$$

by means of the Principle of Independence of Premises (IP) and Markov Principle (MP), and then Skolemising twice.

Therefore, combining the results in 4.11, 4.17 and 4.18 we obtain that the notion of a Gödel doctrine really provides a categorical abstraction of the main concepts involved in the Dialectica translation. We discuss this in more detail in the next section.

5 A characterisation of Dialectica doctrines

The concept of Dialectica category, originally introduced by de Paiva [6], was generalised to the fibrational setting by Hofstra [16].

We briefly recall the notion of a Dialectica category $\mathfrak{Dial}(\mathcal{C})$ associated to a finitely complete category $\mathcal{C}$ (see [6] for further details):

- An **object** of $\mathfrak{Dial}(\mathcal{C})$ is a triple (U, X, α), where α is a subobject of $U \times X$ in $\mathcal{C}$. We think of such a triple as a formula $\exists u.\forall x.\alpha(u, x)$.
- An **arrow** from $\exists u.\forall x.\alpha(u, x)$ to $\exists v.\forall y.\beta(v, y)$, for two objects (U, X, α) and (V, Y, β) in $\mathfrak{Dial}(\mathcal{C})$ is a pair $(U \xrightarrow{f} V, U \times Y \xrightarrow{F} X)$ of arrows of $\mathcal{C}$, i.e. a pair:
$$(u : U \mid f(u) : V, \quad u : U, y : Y \mid F(u, y) : X)$$

of terms in context (as usual, we are thinking of $\mathcal{C}$ as the category of contexts associated to some type theory), satisfying the condition:

$$\alpha(u, F(u, y)) \leq \beta(f(u), y)$$

between the reindexed subobjects, where the squares:

$$\begin{array}{ccc} \alpha(u, F(u,y)) & \longrightarrow & \alpha \\ \downarrow & & \downarrow \\ U \times Y & \xrightarrow[\langle \pi_U, F \rangle]{} & U \times X \end{array} \qquad \begin{array}{ccc} \beta(f(u), y) & \longrightarrow & \beta \\ \downarrow & & \downarrow \\ U \times Y & \xrightarrow[f \times 1_Y]{} & V \times Y \end{array}$$

are pullbacks.

A finitely complete category $\mathcal{C}$ admits a natural structure of *subobject doctrine* $\mathsf{Sub}\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$, if we look at $\mathcal{C}$ itself as a category of contexts associated to some type theory and at the subobjects α of a given object X of $\mathcal{C}$ as the predicates $\alpha(x)$ in context $x : X$.

So we can see that the notions of object and arrow of $\mathfrak{Dial}(\mathcal{C})$ are motivated by Gödel's notion of Dialectica interpretation (see Section 2.1), and in particular by its action on formulas in the language of arithmetic of the form $A \to B$.

5.1 Dialectica doctrines

The notion of Dialectica category was generalised to an arbitrary fibration by Hyland [17], Biering [3] and Hofstra [16]. In this section we deal with the proof-irrelevant version of the fibrational Dialectica construction associating a doctrine $\mathfrak{Dial}(P)$ called a *dialectica doctrine* to a given doctrine P:
Doctrinal Dialectica construction. Let $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ be a doctrine whose base category $\mathcal{C}$ is cartesian closed. The **dialectica doctrine** $\mathfrak{Dial}(P)\colon \mathcal{C}^{\mathrm{op}} \longrightarrow$ **Pos** is defined as the functor sending:

- an object I into the poset $\mathfrak{Dial}(P)(I)$ defined as follows:
  - **objects** are quadruples (I, U, X, α) where I, U and X are objects of the base category $\mathcal{C}$ and $\alpha \in P(I \times U \times X)$;
  - **partial order:** we stipulate that $(I, U, X, \alpha) \leq (I, V, Y, \beta)$ if there exists a pair (f_0, f_1), where $I \times U \xrightarrow{f_0} V$ and $I \times U \times Y \xrightarrow{f_1} X$ are morphisms of $\mathcal{C}$ such that:

$$\alpha(i, u, f_1(i, u, y)) \leq \beta(i, f_0(i, u), y).$$

- an arrow $J \xrightarrow{g} I$ into the poset morphism $\mathfrak{Dial}(P)(I) \to \mathfrak{Dial}(P)(J)$ sending a predicate (I, U, X, α) to the predicate:

$$(J, U, X, \alpha(g(j), u, x)).$$

Remark 5.1. Let $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ be a doctrine and let I be any object of $\mathcal{C}$. Then the poset $\mathfrak{Dial}(P)(I)$ is isomorphic to the poset reflection of the Dialectica category associated to some category.

5.2 Dialectica doctrines via quantifier completions

Our aim is to connect the notion of doctrinal Dialectica construction to the one of a Gödel doctrine and show that, under certain hypotheses, these notions are equivalent. In order to show this, we ask ourselves when is it the case that a doctrine is an instance of a Dialectica construction and, in this case, which doctrine do we need to complete in order to go back to the doctrine we started from.

The main background result we need is the following statement. (Here $Q^{\forall}$ and $Q^{\exists}$ denote the universal and the existential completions of any doctrine Q. We are going to recap these notions later.)

Theorem 5.2 (Hofstra [16]). *If $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ is a doctrine, then there is an isomorphism:*

$$\mathfrak{Dial}(P) \cong (P^{\forall})^{\exists}$$

which is natural in P.

We briefly recall the notion of existential completion of a doctrine, see [36] for more details:

Existential completion. Let $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ be a doctrine. The **existential completion** $P^{\exists}\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ of P is the doctrine such that, for every object A of $\mathcal{C}$, the poset $P^{\exists}(A)$ is defined as follows:

- **objects:** triples (A, B, α), where A and B are objects of $\mathcal{C}$ and α is a predicate in $P(A \times B)$.

- **order:** $(A, B, \alpha) \le (A, C, \beta)$ if there exists an arrow $A \times B \xrightarrow{f} C$ of $\mathcal{C}$ such that:

  $$\alpha(a, b) \vdash \beta(a, f(a, b)) \qquad (\text{ i.e. } \alpha \le P_{\langle \pi_A, f\rangle}(\beta) \text{ })$$

  in $P(A \times B)$ (here $A \times B \xrightarrow{\pi_A} A$ is the projection on A).

Whenever f is an arrow $A \to C$ of $\mathcal{C}$, the functor $P^{\exists}(C) \xrightarrow{P^{\exists}_f} P^{\exists}(A)$ sends an object (C, D, γ) of $P^{\exists}(C)$ to the object:

$$(A, D, \gamma(f(a), d)) \qquad (\text{ i.e. } (A, D, P_{\langle f\pi_A, \pi_D\rangle}(\gamma)) \text{ })$$

of $P^{\exists}(A)$ (here π_A, π_D are the projections from $A \times D$).

We think of a triple (A, B, α) in $P^{\exists}(A)$ as the predicate $(\exists b : B)\alpha(a, b)$. This construction provides a free completion, i.e. it extends to a 2-functor $(-)^{\exists} : \mathsf{Doc} \to \mathsf{ExD}$ which is left adjoint to the corresponding forgetful functor $\mathsf{ExD} \to \mathsf{Doc}$ (which sends an existential doctrine P to the doctrine P itself). We recall that the associated monad happens to be lax-idempotent. Analogously, let us recall the notion of universal completion of a doctrine.

Universal completion. Let $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ be a doctrine. The **universal completion** $P^{\forall}\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ of P is the doctrine such that, for every object A of $\mathcal{C}$, the poset $P^{\forall}(A)$ is defined as follows:

- **objects:** triples (A, B, α), where A and B are objects of $\mathcal{C}$ and α is a predicate in $P(A \times B)$.
- **order:** $(A, B, \alpha) \leq (A, C, \beta)$ if there exists an arrow $A \times C \xrightarrow{g} B$ of $\mathcal{C}$ such that:
$$\alpha(a, g(a, c)) \vdash \beta(a, c)$$
in $P(A \times C)$.

Again, if f is an arrow $A \to C$ of $\mathcal{C}$, the functor $P^{\forall}(C) \xrightarrow{P^{\forall}_f} P^{\forall}(A)$ sends an object (C, D, γ) of $P^{\forall}(C)$ to the object $(A, D, \gamma(f(a), d))$ of $P^{\forall}(A)$.

We think of a triple (A, B, α) in $P^{\forall}(A)$ as the predicate $(\forall b : B)\alpha(a, b)$. As for the notion of existential completion, this construction provides a free completion, i.e. it extends to a 2-functor which is right adjoint to the obvious forgetful functor inducing a colax-idempotent monad. The universal and the existential completions of a given doctrine P are related by the following natural isomorphism:

$$P^{\forall} \cong (-)^{\mathrm{op}}((-)^{\mathrm{op}}P)^{\exists} \tag{6}$$

where $(-)^{\mathrm{op}}$ is the functor $\mathbf{Pos} \to \mathbf{Pos}$ which inverts the order of any poset (see [37]).

We recall that the existential and universal completions are really well-behaved, as the doctrines obtained by applying (either of) these completions to a given doctrine P can be internally characterised without explicitly referring to P.

Proposition 5.3 (Intrinsic $\exists, \forall$-completions)**.** *Let $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ be a doctrine. Assume that P is existential. Then P is an existential completion of some other doctrine P' precisely when it has* enough existential-free predicates*, i.e. when, for every predicate $a : A \mid \alpha(a)$ of P, there is an existential-free predicate:*

$$a : A, b : B \mid \beta(a, b)$$

of P such that $\alpha(a) \cong (\exists b : B)\beta(a, b)$ in $P(A)$. In this case, such a doctrine P' is the full sub-doctrine $P^{\exists\text{-free}}$ of P whose predicates are the existential-free predicates of P.

Analogously, if P is a universal doctrine, then P is a universal completion of some doctrine P' precisely when it has enough universal-free predicates*, i.e. when, for every predicate $a : A \mid \alpha(a)$ of P, there is a universal-free predicate:*

$$a : A, b : B \mid \beta(a, b)$$

of P such that $\alpha(a) \cong (\forall b : B)\beta(a,b)$ in $P(A)$. In this case, such a doctrine P' is the full sub-doctrine $P^{\forall\text{-free}}$ of P whose predicates are the universal-free predicates of P.

By means of the Proposition 5.3 above, the following result follows. This provides a characterisation of the free-algebras of the monad $\mathfrak{Dial}(-)$.

Theorem 5.4 (Intrinsic Dialectica doctrines)**.** *Let us assume that the category $\mathcal{C}$ is cartesian closed. Then the doctrines $P \colon \mathcal{C}^{\mathrm{op}} \longrightarrow$ **Pos** that are dialectica completions of some doctrine P', i.e. $P \cong \mathfrak{Dial}(P')$, are precisely those that are Gödel doctrines. Moreover, in this case, such a doctrine P' is the full sub-doctrine $P^{\exists\forall\text{-free}}$ of the quantifier-free predicates of P.*

Theorem 5.4 establishes a way of re-defining the notion of doctrinal dialectica construction by only referring to the internal properties of a given doctrine P, i.e. without the need for mentioning the existence of a doctrine that P is the dialectica completion of. As the instances of the dialectica completion are precisely the Gödel doctrines, the properties defining the latter notion fully characterise the notion of dialectica doctrine itself. This is our second main theorem, following the result stated in Theorem 4.18 it shows that Gödel doctrines are a sensible doctrine adaptation of Dialectica categories.

Example 5.5. Let us consider the syntactic doctrine H of Hilbert's ϵ-τ calculus as in Example 4.12. Since every predicate of H is quantifier-free, by Theorem 5.4 it is the case that H is the dialectica completion of itself. In fact, one can prove that a Gödel doctrine is isomorphic to its dialectica completion precisely when it is a model of Hilbert's ϵ-τ calculus (see [26] for more details regarding the existential completion).

We end the current section with the following:

Remark 5.6. The existential completion of a universal doctrine whose base is cartesian closed happens to be universal as well (see [16] for more details). Therefore, by Theorem 5.4, it is the case that Dialectica doctrines already happen to be both existential and universal.

Complementing this description, one can also look at the Dialectica completion as a procedure to add existential and universal quantifications to the predicative part of a type theory containing at least the simply typed lambda calculus. Trotta and Spadetto [37] analyse which logical structure we might assume to be already present in the predicative part of our type theory, is preserved – or at least maintained – by this procedure.

6 Gödel doctrines and their Dialectica principles

A Gödel doctrine is essentially a structure where the main features underlying the Dialectica translation hold in a generalised form. One of these concepts is e.g. the existence of a witness function and a counterexample one, for every given implication $i : I \mid \exists u.\forall x.A_D(u,x,i) \vdash \exists v.\forall y.B_D(v,y,i)$. Intuitively, one can look at the notion of *existential-free* predicate as a reformulation of the *principle of independence of premises*, as well as the universal property characterising a *universal-free* predicate reformulates the content of *Markov principle*. In fact, in the proof of Theorem 4.18 existential and universal free predicates correspond to (IP) and (MP*) in the Dialectica interpretation of implicational formulae.

In this section, we are going to make this intuition precise, by formally connecting the notion of Gödel first-order hyperdoctrine to the principles (IP) and (MP*).

6.1 Gödel hyperdoctrines

We show in which sense the principles (IP*) and (MP*) are satisfied in a Gödel hyperdoctrine, using their rule versions. First, we have to equip Gödel doctrines with the appropriate Heyting structure in the fibres in order to be able to formally express these principles. Therefore, we consider Gödel hyperdoctrines.

Definition 6.1. A hyperdoctrine $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$ is said to be a **Gödel hyperdoctrine** when P is a Gödel doctrine.

Notice that from a logical perspective, one might want the quantifier-free predicates to be closed with respect to all the propositional connectives, since this is what happens in logic. However, we can demand less. So we start requiring only the Heyting structure on the fibres and study the logical principles in this setting.

Theorem 6.2. *Every Gödel first-order hyperdoctrine* $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$ *satisfies the* **Rule of Independence of Premise***, i.e. whenever* $\beta \in P(A \times B)$ *and* $\alpha \in P(A)$ *is an existential-free predicate, it is the case that:*

$$a : A \mid \top \vdash \alpha(a) \to \exists b.\beta(a,b) \text{ implies that } a : A \mid \top \vdash \exists b.(\alpha(a) \to \beta(a,b)).$$

Proof. Let us assume that $a : A \mid \top \vdash \alpha(a) \to \exists b.\beta(a,b)$. Then it is the case that $a : A \mid \alpha(a) \vdash \exists b.\beta(a,b)$. Since $\alpha(a)$ is free from the existential quantifier, it is the case that there is a term in context $a : A \mid t(a) : B$ such that:

$$a : A \mid \top \vdash \alpha(a) \to \beta(a,t(a)).$$

Therefore, since:

$$a : A \mid \alpha(a) \to \beta(a,t(a)) \vdash \exists b.(\alpha(a) \to \beta(a,b))$$

(as this holds for any predicate $\gamma(a,-)$ in place of the predicate $\alpha_D(a) \to \beta(a,-)$) we conclude that:

$$a : A \mid \top \vdash \exists b.(\alpha(a) \to \beta(a,b)).$$

□

Similarly, we can prove the following result.

Theorem 6.3. *Every Gödel first-order hyperdoctrine* $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow$ **Hey** *satisfies the following* **Modified Markov Rule***, i.e. whenever* $\beta_D \in P(A)$ *is a quantifier-free predicate and* $\alpha \in P(A \times B)$ *is an existential-free predicate, it is the case that:*

$$a : A \mid \top \vdash (\forall b.\alpha(a,b)) \to \beta_D(a) \text{ implies that } a : A \mid \top \vdash \exists b.(\alpha(a,b) \to \beta_D(a)).$$

Proof. Let us assume that $a : A \mid \top \vdash (\forall b.\alpha(a,b)) \to \beta_D(a)$. Then it is the case that $a : A \mid (\forall b.\alpha(a,b)) \vdash \beta_D(a)$. Hence, since β_D is quantifier-free and α is existential-free, there exists a term in context $a : A \mid t(a) : B$ such that:

$$a : A \mid \top \vdash \alpha(a, t(a)) \to \beta_D(a)$$

therefore, since:

$$a : A \mid \alpha(a, t(a)) \to \beta(a) \vdash \exists b.(\alpha(a,b) \to \beta_D(a))$$

we can conclude that:

$$a : A \mid \top \vdash \exists b.(\alpha(a,b) \to \beta_D(a)).$$

□

To obtain (the generalized) Markov rule from Theorem 6.3 we have to require that the predicate associate with falsum $\bot$ to be quantifier-free.

Corollary 6.4. *Every Gödel first-order hyperdoctrine* $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow$ **Hey** *such that* $\bot$ *is a quantifier-free predicate satisfies (the generalized)* **Markov Rule***, i.e. for every quantifier-free predicate* $\alpha_D \in P(A \times B)$ *it is the case that:*

$$b : B \mid \top \vdash \neg\forall a.\alpha_D(a,b) \text{ implies that } b : B \mid \top \vdash \exists a.\neg\alpha_D(a,b).$$

Proof. This follows from Theorem 6.3 by replacing β_D with $\bot$, which is quantifier-free by hypothesis. □

In Theorems 6.3 and 6.2 we proved that the universal properties of existential and universal free predicates allow us to prove that a Gödel first-order hyperdoctrine satisfies the Modified Markov Rule and the Rule of Independence of Premise.

From a logical perspective, the intuition behind Theorem,6.2 is that the existential-free predicates of a Gödel first-order hyperdoctrine correspond to formulae satisfying (R-IP). Similarly, we have that the quantifier-free predicates of a Gödel doctrine are exactly those satisfying an (R-MMP) by Theorem 6.3. Notice also that applying the definitions of existential-free and universal-free predicates, we immediately obtain the following presentation of the *Rule of Choice*, see [23] (also called *explicit definability* in [30]) and the *Counterexample Property*, previously defined in [37].

Corollary 6.5. *Every Gödel first-order hyperdoctrine* $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$ *such that* $\bot$ *is a quantifier-free object satisfies the* **Counterexample Property***, that is, whenever:*

$$a : A \mid \forall b.\alpha(a,b) \vdash \bot$$

for some predicate $\alpha(a,b) \in P(A \times B)$*, then it is the case that:*

$$a : A \mid \alpha(a, g(a)) \vdash \bot$$

for some term in context $a : A \mid g(a) : B$*.*

Corollary 6.6. *Every Gödel first-order hyperdoctrine* $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$ *such that* $\top$ *is existential-free satisfies the* **Rule of Choice***, that is, whenever:*

$$a : A \mid \top \vdash \exists b.\alpha(a,b)$$

for some existential-free predicate $\alpha \in P(A \times B)$*, then it is the case that:*

$$a : A \mid \top \vdash \alpha(a, g(a))$$

for some term in context $a : A \mid g(a) : B$*.*

6.2 Principles as a strengthening of rules

We have just looked at which *rules* hold in Gödel first-order hyperdoctrines. In this subsection we analyse the respective logical *principles* in Gödel first-order hyperdoctrines. Thus, we look for the right hypotheses that allow us to produce models of the stronger formulation of the rules as principles. These hypotheses involve closure properties of the existential-free and quantifier-free predicates under some propositional connectives (conjunction, implication, falsehood). These requests appear natural if we compare our notions of being categorically existential-free and quantifier-free, to the syntactic notion of being free from a quantifier. Applying connectives to predicates free from a quantifier produces predicates that continue being free from that quantifier. We are going to see a concrete instance of these closure properties in Example 6.13.

The following theorem is the first of this series of results and involves the Independence of Premise:

Theorem 6.7 (Independence of Premise in Gödel hyperdoctrines - strong version)**.** *Every Gödel first-order hyperdoctrine* $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$ *whose existential-free predicates are closed with respect to finite conjunctions satisfies the* **Principle of Independence of Premise***, i.e. whenever* $\beta \in P(A \times B)$ *and* $\alpha \in P(A)$ *is an existential-free predicate, it is the case that:*

$$a : A \mid \top \vdash (\alpha(a) \to \exists b.\beta(a,b)) \to \exists b.(\alpha(a) \to \beta(a,b)).$$

Proof. First, since every Gödel doctrine has enough existential-free predicates, there exists an existential-free predicate $\gamma(a,c) \in P(A \times C)$ such that:

$$a : A \mid \exists c.\gamma(a,c) \dashv\vdash \alpha(a) \to \exists b.\beta(a,b).$$

In particular, we have that $a : A, c : C \mid \gamma(a,c) \vdash \alpha(a) \to \exists b.\beta(a,b)$. Then we have that:

$$a : A, c : C \mid \gamma(a,c) \wedge \alpha(a) \vdash \exists b.\beta(a,b)$$

and $\gamma(a,c) \wedge \alpha(a)$ is an existential-free predicate since both $\gamma(a,c)$ and $\alpha(a)$ are existential-free predicates and existential-free predicates are closed under finite conjunctions by hypothesis. Therefore, we can conclude that there exists a term $a : A, c : C \mid t(a,c) : B$ such that:

$$a : A, c : C \mid \gamma(a,c) \wedge \alpha(a) \vdash \beta(a, t(a,c)).$$

Hence:

$$a : A, c : C \mid \gamma(a,c) \vdash \alpha(a) \to \beta(a, t(a,c))$$

and since $\alpha(a) \to \beta(a, t(a,c)$ is exactly $(\alpha(a) \to \beta(a,b))[t(a,c)/b]$ and it always holds that:

$$a : A, c : C \mid (\alpha(a) \to \beta(a,b))[t(a,c)/b] \vdash \exists b.(\alpha(a) \to \beta(a,b))$$

we can conclude that:

$$a : A, c : C \mid \gamma(a,c) \vdash \exists b.(\alpha(a) \to \beta(a,b)).$$

Therefore we get that:

$$a : A \mid \exists c.\gamma(a,c) \vdash \exists b.(\alpha(a) \to \beta(a,b)).$$

and, since $a : A \mid \exists c.\gamma(a,c) \dashv\vdash \alpha(a) \to \exists b.\beta(a,b)$, it is the case that:

$$a : A \mid \top \vdash (\alpha(a) \to \exists b.\beta(a,b)) \to \exists b.(\alpha(a) \to \beta(a,b)).$$

□

As a corollary of the previous result, we obtain the following presentation of the principle (IP*) introduced in Section 2.2 in terms of Gödel first-order hyperdoctrines. We recall that (IP*) is precisely the form of the Principle of Independence of Premise we need in the Dialectica interpretation.

Corollary 6.8. *Every Gödel first-order hyperdoctrine* $P \colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$ *such that the existential-free predicates are closed with respect to finite conjunction satisfies (IP*), i.e. whenever* $\beta \in P(C \times B)$ *and* $\alpha_D \in P(A)$ *is a quantifier-free predicate, it is the case that:*

$$- \mid \top \vdash (\forall a.\alpha_D(a) \to \exists b.\forall c.\beta(c,b)) \to \exists b.(\forall a.\alpha_D(a) \to \forall c.\beta(c,b)).$$

Proof. It follows from Theorem 6.7 and from the fact that if α_D is quantifier-free then $\forall a.\alpha_D$ is existential-free. □

Similarly, we can prove the following result for Markov principle. In this case, we make the additional request that there is an arrow from the terminal object of the base category to any other object. This corresponds to requiring that every context of the underlying type theory has a substitution from the empty context, that is equivalent to the assumption that every basic sort has a given term (hence that every type in context has a term in context) i.e. is not initial.

Theorem 6.9 (Modified Markov in Gödel hyperdoctrines - strong version)**.** *Every Gödel first-order hyperdoctrine* $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$ *such that there exists an arrow* $1 \to C$ *for every object* C *of* $\mathcal{C}$*, and whose existential-free predicates are closed with respect to implication, satisfies the following* **Modified Markov Principle***, i.e. whenever* $\beta_D \in P(A)$ *is a quantifier-free predicate and* $\alpha \in P(A \times B)$ *is an existential-free predicate, it is the case that:*

$$a : A \mid \top \vdash (\forall b.\alpha(a,b) \to \beta_D(a)) \to \exists b.(\alpha(a,b) \to \beta_D(a)).$$

Proof. Since α is an existential-free predicate and β_D is quantifier-free it is the case that: $\forall b.\alpha(a,b) \to \beta_D(a)$ lives in $P^{\exists\text{-free}}(A)$. Thus, since $P^{\exists\text{-free}}$ has enough universal-free predicates as P is a Gödel doctrine, there exists a universal-free predicate of $P^{\exists\text{-free}}$, i.e. a quantifier-free predicate:

$$\sigma_D \in P^{\exists\text{-free}}(A \times C)$$

of P, such that:

$$a : A \mid \forall c.\sigma_D(a,c) \dashv\vdash \forall b.\alpha(a,b) \to \beta_D(a).$$

In particular $a : A \mid \forall c.\sigma_D(a,c) \wedge \forall b.\alpha(a,b) \vdash \beta_D(a)$ and hence:

$$a : A \mid \forall c.\forall b.(\sigma_D(a,c) \wedge \alpha(a,b)) \vdash \beta_D(a)$$

since by hypothesis all types are inhabited. Now, since β_D is quantifier-free, i.e. it is universal-free in $P^{\exists\text{-free}}$, there exist two terms $a : A \mid t(a) : B$ and $a : A \mid t'(a) : C$ such that:

$$a : A \mid \sigma_D(a, t'(a)) \wedge \alpha(a, t(a)) \vdash \beta_D(a).$$

Therefore it is the case that $a : A \mid \sigma_D(a, t'(a)) \vdash (\alpha(a,b) \to \beta_D(a))[t(a)/b]$. Now, since:

$$a : A \mid \forall c.\sigma_D(a,c) \vdash \sigma_D(a, t'(a))$$

always holds and since $a : A \mid (\alpha(a,b) \to \beta_D(a))[t(a)/b] \vdash \exists b.(\alpha(a,b) \to \beta_D(a))$, we can conclude that:

$$a : A \mid \top \vdash (\forall b.\alpha(a,b) \to \beta_D(a)) \to \exists b.(\alpha(a,b) \to \beta_D(a)).$$

□

To obtain the presentation of the generalized Markov Principle (MP*) used in the Dialectica interpretation as a corollary of Theorem 6.9, we simply have to require the bottom predicate $\bot$ of a Gödel first-order hyperdoctrine to be *quantifier-free.*

Corollary 6.10. *Every Gödel first-order hyperdoctrine* $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$ *such that there exists an arrow* $1 \to C$ *for every object* C *of* $\mathcal{C}$*, and whose existential-free predicates are closed with respect to implication, and such that* $\bot$ *is a quantifier-free predicate, satisfies (the generalized)* **Markov Principle***, i.e. for every quantifier-free predicate* $\alpha_D \in P(A \times B)$ *it is the case that:*

$$b : B \mid \top \vdash \neg\forall a.\alpha_D(a,b) \to \exists a.\neg\alpha_D(a,b).$$

Proof. It follows by Theorem 6.9 just by replacing β_D with $\bot$, that is quantifier-free by hypothesis. □

We have proved that under suitable hypotheses, a Gödel first-order hyperdoctrine satisfies (IP*), (MP*), (MMP), and the principle of Skolemisation.

Therefore, combining Theorem 6.7, Theorem 6.9, and Proposition 4.11, we can repeat the chain of equivalences we provided in Section 2.1, and obtain the following main result.

Theorem 6.11 (Strengthened Dialectica functionals)**.** *Let* $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$ *be a Gödel first-order hyperdoctrine such that:*

- *there exists an arrow* $1 \to C$ *for every object* C *of* $\mathcal{C}$*;*
- *existential-free predicates are closed with respect to implication and finite conjunction;*
- *falsehood* $\bot$ *is a quantifier-free predicate.*

Then for every ψ_D *in* $P(I \times U \times X)$ *and* ϕ_D *in* $P(I \times V \times Y)$ *quantifier-free predicates of* P *we have that the formula:*

$$i : I \mid \exists u.\forall x.\psi_D(i,u,x) \to \exists v.\forall y.\phi_D(i,v,y)$$

is provably equivalent to:

$$i : I \mid \exists f_0, f_1.\forall u, y.(\psi_D(i,u,f_1(i,u,y)) \to \phi_D(i,f_0(i,u),y)).$$

Theorem 6.11 fully represents a categorical version of the translation of the implication connective in the Dialectica interpretation. In particular, it shows that the equivalence $(\psi \to \phi)^D \leftrightarrow (\psi^D \to \phi^D)$ presented in Section 2.1 is perfectly modelled by a Gödel first-order hyperdoctrine satisfying the natural additional closure properties of Theorem 6.11.

Remark 6.12. Observe that Theorem 6.11 can be considered a stronger version of Theorem 4.18. Hence, once more, it converts the rule stated in the latter theorem into an actual principle.

In detail, by the thesis of Theorem 6.11, it is enough to observe that the first sequent of the statement of Theorem 4.18 is equivalent to the sequent:

$$i : I \mid \top \vdash \exists u.\forall x.A_D(i,u,x) \to \exists v.\forall y.B_D(i,v,y)$$

by the elimination and introduction rules for the implication, and that the second one is equivalent to the following:

$$i : I \mid \top \vdash \exists f_0, f_1.\forall u, y.(A_D(i,u,f_1(i,u,y)) \to B_D(i,f_0(i,u),y)).$$

For this second equivalence one applies the implicational elimination and introduction to convert the second sequent of 4.18 into:

$$i : I, u : U, y : Y \mid \top \vdash A_D(i,u,f_1(i,u,y)) \to B_D(i,f_0(i,u),y)$$

which is actually equivalent to $i : I \mid \top \vdash \exists f_0, f_1.\forall u, y.(A_D(i,u,f_1(i,u,y)) \to B_D(i,f_0(i,u),y))$, by Corollary 6.6 and being the formula:

$$i : I \mid \forall u, y.(A_D(i,u,f_1(i,u,y)) \to B_D(i,f_0(i,u),y))$$

existential-free.

Theorem 6.11 follows as a consequence of the fragment of first-order logic under which the internal language of a Gödel first-order hyperdoctrine is closed. Observe that this fragment contains at least intuitionistic first-order logic together with the Principle of Independence of Premise, the Modified Markov Principle, and the Principle of Skolemisation. These principles, together with the rules of intuitionistic first-order logic, are precisely what is needed to get the equivalence $(A \to B)^D \leftrightarrow (A^D \to B^D)$ in a Gödel first-order hyperdoctrine.

Clearly, any boolean doctrine satisfies these principles as well, as it models every inference rule of classic first-order logic. However, in general, they are not satisfied by a usual hyperdoctrine, because they are not necessarily true in intuitionistic first-order logic. It turns out that *the fragment of first-order logic modelled by a Gödel hyperdoctrine is between intuitionistic first-order logic and classical first-order logic*: it is powerful enough to guarantee the equivalences in Section 2.1 that justify the Dialectica interpretation of the implication.

We end the current section with the following:

Example 6.13. Let us consider the syntactic doctrine H of Hilbert's ϵ-τ calculus of Example 4.12. Recall from Example 4.16 that H is a Gödel doctrine, hence we have that H is a Gödel hyperdoctrine in the sense of Definition 6.1. Since H is a model of many-sorted intuitionistic first-order logic and since every predicate of H is both existential-free and universal-free (hence every element is quantifier-free), it is the case that all of the hypotheses of the results of this section are trivially satisfied by H. Hence H happens to be a model of all of the versions of the Rule/Principle of Independence of Premise and Markov's Rule/Principle presented. Clearly, in order to have (MP*) and (MMP) satisfied by H one needs to make the further assumption that the sorts of Hilbert's ϵ-τ calculus be inhabited.

Observe that, in general, if a hyperdoctrine P has Hilbert's ϵ-operators only (hence it is a hyperdoctrine possibly without τ-operators), that is enough to have the required closure property under implication and finite conjunction of the existential-free predicates. This is actually enough to deduce the soundness of (IP) (and (IP*)), as in the proofs of Theorem 6.7 and Corollary 6.8 we do not use that $P^{\exists\text{-free}}$ has enough universal free predicates (see Example 4.12). However, in order to also have (MMP) (hence (MP*) with the additional closure properties) to be satisfied, one actually needs $P^{\exists\text{-free}}$ to have enough universal free predicates. This is ensured e.g. if P has τ-operators as well. We refer to [8] for details and proofs of the validity of (MMP) and (MP*) in Hilbert's τ-calculus.

A non-syntactic example of hyperdoctrine equipped with ϵ-operators only is presented in Example 5.14 of [23] (observe that the domain category of this example does not contain initial objects). Also, see [31] for further instances of doctrines equipped with Hilbert's ϵ-operators.

We turn now to an application of Gödel doctrines.

7 Tripos-to-topos and Gödel doctrines

The tripos-to-topos construction was originally introduced in [33, 18] as a generalisation of the construction of the category of sheaves on a locale. Recently, this construction has been proven to be an instance of the exact completion of an elementary existential doctrine. We refer to [25, 23] for the details, but we briefly recall it.

Tripos-to-topos. Given a first-order hyperdoctrine $P\colon \mathcal{C}^{\text{op}} \longrightarrow \mathbf{Hey}$, the category T_P consists of:

- **objects:** are pairs (A, ρ) where $\rho \in P(A \times A)$ satisfies:
  - *symmetry:* $a_1, a_2 : A \mid \rho(a_1, a_2) \vdash \rho(a_2, a_1)$;
  - *transitivity:* $a_1, a_2, a_3 : A \mid \rho(a_1, a_2) \wedge \rho(a_2, a_3) \vdash \rho(a_1, a_3)$.
- **arrows** $(A, \rho) \xrightarrow{\phi} (B, \sigma)$: are objects $\phi \in P(A \times B)$ such that:
  1. $a : A, b : B \mid \phi(a, b) \wedge \rho(a, a) \vdash \sigma(b, b)$;
  2. $a_1, a_2 : A, b : B \mid \rho(a_1, a_2) \wedge \phi(a_1, b) \vdash \phi(a_2, b)$;
  3. $a : A, b_1, b_2 : B \mid \sigma(b_1, b_2) \wedge \phi(a, b_1) \vdash \phi(a, b_2)$;
  4. $a : A, b_1, b_2 : B \mid \phi(a, b_1) \wedge \phi(a, b_2) \vdash \sigma(b_1, b_2)$;
  5. $a : A \mid \rho(a, a) \vdash \exists b.\phi(a, b)$.

Then the following holds:

Theorem 7.1. *Let $P\colon \mathcal{C}^{\text{op}} \longrightarrow \mathbf{Hey}$ be a hyperdoctrine. Then T_P is an exact category.*

The construction of the category T_P can be presented in the more general context of elementary and existential doctrines, and it is also called the **exact completion** of the elementary existential doctrine P, since it lifts to an adjunction between the 2-category of exact categories and that of elementary and existential doctrine. We refer to [25, Cor. 3.4] for a complete description of the construction in the general case.

7.1 Tripos-to-topos and exact completions

We recall from [26] a useful characterisation of the tripos-to-topos construction of a first-order hyperdoctrine arising as an existential completion. Again, in the present work we present the results for hyperdoctrines, but the characterisation presented in [26] works for an arbitrary elementary and existential doctrine.

To properly present such a characterisation we first need to recall from [23, 25, 24] the construction of the category of *predicates* of a first-order hyperdoctrine. The construction of this category is related to the *comprehension and comprehensive diagonal completions.*

Definition 7.2. Given a first-order hyperdoctrine $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$ we define the **comprehension completion** $P_c\colon \mathcal{G}_P^{\mathrm{op}} \longrightarrow \mathbf{Hey}$ of P as follows:

- an object of $\mathcal{G}_P$ is a pair (A, α) where A is a set and $\alpha \in P(A)$;
- an arrow $(A, \alpha) \xrightarrow{f} (B, \beta)$ is an arrow $A \xrightarrow{f} B$ such that:

$$a : A \mid \alpha(a) \vdash \beta(f(a)).$$

The fibres $P_c(A, \alpha)$ are given by those predicates γ of $P(A)$ such that $a : A \mid \gamma(a) \vdash \alpha(a)$ (i.e. $\gamma \leq \alpha$). Moreover, the action of P_c on a morphism $f : (B, \beta) \to (A, \alpha)$ is defined as $P_c(f)(\gamma) = P_f(\gamma) \wedge \beta$ i.e. the predicate:

$$b : B \mid \gamma(f(b)) \wedge \beta(b)$$

where $\gamma \in P(A)$ is such that $\gamma \leq \alpha$.

Similarly, the construction which freely adds a comprehensive diagonal is provided by the *extensional reflection*. We denote $\delta_A := \exists_\Delta(\top_A)$. According to the internal language of a given doctrine P, the predicate $\delta_A \in P(A \times A)$ corresponds to the predicate:

$$a_1 : A, a_2 : A \mid a_1 = a_2.$$

Definition 7.3. Given an elementary doctrine $P\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathbf{Hey}$ we can define the **extensional reflection** $P_x\colon \mathcal{X}_P^{\mathrm{op}} \longrightarrow \mathbf{Pos}$ of P as follows: the base category $\mathcal{X}_P$ is the quotient category of $\mathcal{C}$ with respect to the equivalence relation where $f \sim g$ when:

$$\top \vdash f(a) = g(a) \quad \text{(i.e. } \top_A \leq P_{\langle f,g\rangle}(\delta_B) \text{ in category-theoretic notation)}$$

in context $a : A$, for two parallel arrows $f, g : A \to B$. The equivalence class of a morphism f of $\mathcal{C}$, i.e. an arrow of $\mathcal{X}_P$, is denoted by $[f]$.

Finally, we denoted by $\mathbf{Pred}(P)$ the **category of predicates** of a doctrine P, i.e. the category defined as:

$$\mathbf{Pred}(P) := \mathcal{X}_{P_c}$$

where P_c is the comprehension completion of P. Again, we refer to [23, 25, 24] for a complete description of these constructions. Now we have all the instruments to recall the characterisation of tripos-to-topos of existential completions from [26]. Such a characterisation essentially shows that every tripos-to-topos of an existential completion is an instance of $(-)_{\mathsf{ex/lex}}$ completion, namely the exact completion of a lex category in the sense of [5, 4].

Theorem 7.4. *Let $P\colon \mathcal{C}^{\text{op}} \longrightarrow \mathbf{Hey}$ be a first-order hyperdoctrine. Then we have the equivalence:*

$$\mathsf{T}_{P^{\exists}} \equiv \mathbf{Pred}(P)_{\mathsf{ex/lex}}$$

of exact categories.

7.2 Tripos-to-topos for Gödel first-order hyperdoctrines

We recall that a Gödel first-order hyperdoctrine is in particular the existential completion of its subdoctrine of existential-free predicates by Theorem 5.4 and Proposition 5.3. Therefore, we are able to apply Theorem 7.4 and can conclude the following characterisation of the tripos-to-topos construction of Gödel first-order hyperdoctrines:

Theorem 7.5. *Let $P\colon \mathcal{C}^{\text{op}} \longrightarrow \mathbf{Hey}$ be a Gödel first-order hyperdoctrine. Then the equivalence of exact categories:*

$$\mathsf{T}_{P} \equiv \mathbf{Pred}(P^{\exists\text{-}free})_{\mathsf{ex/lex}}$$

holds.

We recall from [33, 18] that when a doctrine is a tripos, then its tripos-to-topos construction is a topos. Hence, we have the following corollary for Gödel first-order hyperdoctrines:

Corollary 7.6. *Let $P\colon \mathcal{C}^{\text{op}} \longrightarrow \mathbf{Hey}$ be a Gödel first-order hyperdoctrine. If P is a tripos, then $\mathbf{Pred}(P^{\exists\text{-}free})_{\mathsf{ex/lex}}$ is a topos.*

Given this corollary, we might be tempted to call these toposes Dialectica toposes. These are however different from Biering's Dialectica toposes.

Remark 7.7. The notion of *Dialectica topos* introduced in [3] as the tripos-to-topos of a suitable tripos called *dialectica tripos*. In [15] Hofstra characterises triposes arising in terms of ordered PCAs equipped with a filter. This characterisation includes Effective Topos-like triposes, but also the triposes for relative, modified and extensional realisability and the dialectica tripos. Therefore, the dialectica tripos can been seen as a tripos arising from a suitable ordered PCA.

Triposes given by PCAs are known to be instances of a general completion that freely adds left adjoints along arbitrary maps. Hofstra was the first to observe this fact, see [15, 14], but later in [11] and [26] it was proved that the construction identified by Hofstra in [14] is a particular case of the full existential completion of a primary doctrine.

In this paper we have proposed a different approach to the definition of doctrines related to the Dialectica interpretation, focusing on the logical principles and rules we need to properly translate the implication connective as in the Dialectica. Therefore, we could say that our approach is more syntactic, and less related to realisability in general. The Dialectica tripos introduced in [3] is not a Gödel doctrine in general, since Gödel doctrines are given by an existential completion just along projections (and by a universal completion), while the Dialectica tripos is an instance of the full existential completion. Therefore, the Dialectica tripos satisfies different structural properties with respect to an arbitrary Gödel doctrine. For example, the Dialectica tripos has left adjoints along every arrow, satisfying Beck-Chevalley conditions, while in an arbitrary Gödel doctrine the Beck-Chevalley conditions are not satisfied along arbitrary maps. Employing the universal properties of the existential and universal completions, one can show that the Dialectica tripos just contains a Gödel doctrine, but it is not equivalent to such a doctrine.

We can also relate this work to Maietti's work on Joyal's arithmetic universes.

Remark 7.8. Observe that categories arising as tripos-to-topos results of Gödel first-order hyperdoctrines have the same abstract presentation as the *Joyal-arithmetic universes* introduced by Maietti in [22]. Recall also that a Joyal-arithmetic universe is defined as the exact completion $\mathbf{Pred}(\mathcal{S})_{\mathsf{ex/lex}}$ of the category of predicates of a Skolem theory $\mathcal{S}$ as defined in [22, Def. 2.2], namely a cartesian category with a parameterised natural numbers object where all the objects are finite products of the natural numbers object.

8 Conclusion

This article is the culmination of various intertwined investigations begun in [38] and [39]. Inspired by Hofstra [16] and Hyland [17], as well as by the work of Maietti and Trotta [26], itself inspired by Trotta [36], we embarked on the programme of expanding the characterisation of the categorical version of the Dialectica Interpretation, to complete Hofstra's work. Hofstra studied the principle of Skolemisation showing that the Dialectica can be seen as a double completion of a cartesian category, under simple coproducts and products. We complete the work of Hofstra, by showing that the two other non-intuitionistic principles in the Dialectica interpretation, i.e. IP and MP can also be seen as instances of categorical constructions. Then we make sure that all the logical principles involved in the interpretation are precisely represented in the categorical models obtained.

Since the work on the purely fibrational setting seemed too abstract and hard to grasp, especially for our target audience of logicians, we opted for descriptions at the level of hyperdoctrines in [39, 40]. These doctrines are the poset reflections of the fibrations used early on. This crystallised our understanding of the issue of quantifier-free formulae in the categorical setting, but also made clear the import of non-intuitionistic principles such as Independence of Premise and the Markov Principle. These principles had been discussed by logicians, but not in categorical terms, as far as we are aware. Our investigation is, so far, restricted to the environment of the Dialectica interpretation, but it has a wider reach, helping to complete the program of categorification of logic, originally suggested by Lawvere.

We hope to carry on exploring other issues of this investigation. We started connecting this work to the work on categorical realisability and computability, as described by Pitt's tripos theory and the tripos to topos construction [33] in the final section of this article. Much remains to be done. A different direction that we have not even started to explore is the extension of our work to generalised versions of the Dialectica interpretation, as already hinted in the text, to dependent type theory in the style of [29]. Finally, the work in the original Dialectica category model [7] has had several applications to computer science problems like concurrency theory, in the shape of Petri Nets [9] and others [41]. We plan to investigate if these and other applications can be improved by our doctrinal version of the models.

Acknowledgements

It should be clear to the readers of this paper that we owe a large debt of gratitude to Peter Hofstra, who left us far too early this year. Besides clearly building up on his paper [16], we also started discussing these ideas with him, when Davide Trotta presented the work at the Category Theory Seminar in Ottawa (online) in October 2021. For these reasons, we would like to dedicate this work to him.

We also would like to thank Maria Emilia Maietti, for ideas and suggestions on ongoing joint work. Lastly, we also thank the anonymous referees.

References

[1] J. Avigad and S. Feferman. Gödel's functional ("Dialectica") interpretation. In Samuel R. Buss, editor, *Handbook of Proof Theory*, volume 137 of *Studies in Logic and the Fundations of Mathematics*, pages 337–405. Elsevier, 1998.

[2] J. Bell. Hilbert's ϵ-operator in intuitionistic type theories. *Mathematical Logic Quarterly*, 39(1):323–337, 1993.

[3] B. Biering. *Dialectica interpretations: a categorical analysis.* PhD thesis, University of Copenhagen, 2008. https://cs.au.dk/ birke/phd-students/BieringB-thesis.pdf.

[4] A. Carboni and R. Celia Magno. The free exact category on a left exact one. *Journal of the Australian Mathematical Society*, 33(3):295–301, 1982.

[5] A. Carboni and E.V. Vitale. Regular and exact completions. *Journal of Pure and Applied Algebra*, 125(1):79–116, 1998.

[6] V. de Paiva. The Dialectica categories. *Categories in computer science and logic (Boulder, CO, 1987), Contemporary Mathematics, vol 92, American Mathematical Society*, 92:47–62, 1989.

[7] V. de Paiva. *The Dialectica categories.* PhD thesis, University of Cambridge, 1991. https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-213.pdf.

[8] D. Devidi. Intuitionistic ϵ- and τ-calculi. *Mathematical Logic Quarterly*, 41(4):523–546, 1995.

[9] E. di Lavore, W. Leal, and V. de Paiva. Dialectica Petri nets. arXiv:2105.12801, 2021.

[10] S. Feferman, J.W. Dawson Jr., S.C. Kleene, G.H. Moore, R.M. Solovay, and J. van Heijenoort (Eds.). *Kurt Gödel: Collected Works, Volume II, Publications 1938-1974.* Oxford University Press, Oxford, 1990.

[11] J. Frey. Categories of partial equivalence relations as localizations. *Journal of Pure and Applied Algebra*, page 107115, 2022.

[12] J.Y. Girard. Linear logic. *Theoretical Computer Science*, 50(1):1–101, 1987.

[13] K. Gödel. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica*, 12(3-4):280–287, 1958.

[14] P. Hofstra. *Completions in Realizability.* PhD thesis, Universiteit Utrecht, 2003.

[15] P. Hofstra. All realizability is relative. *Mathematical Proceedings of the Cambridge Philosophical Society*, 141(2):239–264, 2006.

[16] P. Hofstra. The Dialectica monad and its cousins. *Models, Logics, and Higher-Dimensional Categories: A Tribute to the Work of Mihály Makkai*, 53:107–139, 2011.

[17] J.M. Hyland. Proof theory in the abstract. *Annals of Pure and Applied Logic*, 114(1):43–78, 2002.

[18] J.M. Hyland, P.T. Johnstone, and A.M. Pitts. Tripos theory. *Mathematical Proceedings of the Cambridge Philosophical Society*, 88:205–232, 1980.

[19] F.W. Lawvere. Adjointness in foundations. *Dialectica*, 23:281–296, 1969.

[20] F.W. Lawvere. Diagonal arguments and cartesian closed categories. In *Category Theory, Homology Theory and their Applications*, volume 2, pages 134–145. Springer, 1969.

[21] F.W. Lawvere. Equality in hyperdoctrines and comprehension schema as an adjoint functor. In A. Heller, editor, *Proc. New York Symposium on Application of Categorical Algebra*, pages 1–14. American Mathematical Society, 1970.

[22] M.E. Maietti. Joyal's arithmetic universe as list-arithmetic pretopos. *Theory and Applications of Categories*, 24(3):39–83, 2010.

[23] M.E. Maietti, F. Pasquali, and G. Rosolini. Triposes, exact completions, and Hilbert's ϵ-operator. *Tbilisi Mathematical Journal*, 10(3):141–166, 2017.

[24] M.E. Maietti and G. Rosolini. Quotient completion for the foundation of constructive mathematics. *Logica Universalis*, 7(3):371–402, 2013.

[25] M.E. Maietti and G. Rosolini. Unifying exact completions. *Applied Categorical Structures*, 23:43–52, 2015.

[26] M.E. Maietti and D. Trotta. A characterization of generalized existential completions. *Annals of Pure and Applied Logic*, 174(4):103234, 2023.

[27] M. Makkai and G.E. Reyes. *First Order Categorical Logic*, volume 611 of *Lecture Notes in Mathematics*. Springer-Verlag, 1977.

[28] M. Manighetti. Computational interpretations of Markov's principle. arXiv:1611.03714, 2016.

[29] S.K. Moss and T. von Glehn. Dialectica models of type theory. In *33rd Annual ACM/IEEE Symposium on Logic in Computer Science*, page 739–748, New York, NY, USA, 2018. Association for Computing Machinery.

[30] T. Nemoto and M. Rathjen. The independence of premise rule in intuitionistic set theories. arXiv:1911.08027, 2019.

[31] F. Pasquali. A categorical interpretation of the intuitionistic, typed, first order logic with Hilbert's ε-terms. *Logica Universalis*, 10:407–418, 2016.

[32] A.M. Pitts. Categorical logic. In S. Abramsky, D.M. Gabbay, and T.S.E. Maibaum, editors, *Handbook of Logic in Computer Science*, volume 6, pages 39–129. Oxford University Press, 1995.

[33] A.M. Pitts. Tripos theory in retrospect. *Mathematical Structures in Computer Science*, 12(3):265–279, 2002.

[34] R.A.G. Seely. Hyperdoctrines, natural deduction and the beck condition. *Mathematical Logic Quarterly*, 29(10):505–542, 1983.

[35] A.S. Troelstra. *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, volume 344 of *Lecture Notes in Mathematics*. Springer Berlin, 1973.

[36] D. Trotta. The Existential Completion. *Theory and Applications of Categories*, 35(43):1576–1607, 2020.

[37] D. Trotta and M. Spadetto. Quantifier completions, choice principles and applications. arXiv:2010.09111, 2020.

[38] D. Trotta, M. Spadetto, and V. de Paiva. The Gödel Fibration. In Filippo Bonchi and Simon J. Puglisi, editors, *46th International Symposium on Mathematical Foundations of Computer Science (MFCS 2021)*, volume 202 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 87:1–87:16, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[39] D. Trotta, M. Spadetto, and V. de Paiva. Dialectica logical principles. In S. Artemov and A. Nerode, editors, *Logical Foundations of Computer Science*, pages 346–363, Cham, 2022. Springer International Publishing.

[40] D. Trotta, M. Spadetto, and V. de Paiva. Dialectica logical principles, not only rules. *Manuscript (submitted)*, 2022.

[41] G. Winskel. Making concurrency functional. arXiv:2202.13910, 2022.