# Synthetic 1-Categories in Directed Type Theory

**Jacob Neumann**
University of Nottingham, UK

**Thorsten Altenkirch**
University of Nottingham, UK

## Abstract

The field of *directed type theory* seeks to design type theories capable of reasoning synthetically about (higher) categories, by generalizing the symmetric *identity* types of Martin-Löf Type Theory to asymmetric *hom-types*. We articulate the directed type theory of the *category model*, with appropriate modalities for keeping track of variances and a powerful directed-J rule capable of proving results about arbitrary terms of hom-types; we put this rule to use in making several constructions in synthetic 1-category theory. Because this theory is expressed entirely in terms of *generalized algebraic theories*, we know automatically that this directed type theory admits a syntax model.

**2012 ACM Subject Classification** Theory of computation → Type theory

**Keywords and phrases** Semantics, directed type theory, homotopy type theory, category theory, generalized algebraic theories

**Digital Object Identifier** 10.4230/LIPIcs.TYPES.2024.7

**Related Version** *Full Version*: `https://arxiv.org/abs/2410.19520` [26]

**Funding** *Jacob Neumann*: This publication is based upon work from COST Action EuroProofNet, supported by COST (European Cooperation in Science and Technology, www.cost.eu).

## 1 Introduction

A key motivation behind the emergent field of *homotopy type theory (HoTT)* [34] is the interpretation that *types are* $\infty$*-groupoids* [35]. Homotopy type theory can be understood as a *synthetic* theory of $\infty$-groupoids: all the higher structure is generated by the simple rules for manipulating identity types in Martin-Löf Type Theory [23, 24], permitting efficient reasoning with these complex structures.

Not long after homotopy type theory was established, the search for *directed homotopy type theory* – a synthetic theory of (higher) *categories* – began. In a directed type theory, the identity types of ordinary Martin-Löf Type Theory – which are provably symmetric in the theory, i.e. a witness $p\colon \mathsf{Id}(t,t')$ can be turned into $p^{-1}\colon \mathsf{Id}(t',t)$ – are replaced by asymmetric *hom-types*. However, building a type theory to effectively work with these hom-types is beset by difficulties, in particular the need to carefully track the *variances* of terms. A common feature of many approaches to directed type theory (e.g. [22, 29, 28]) is to track these variances by adopting some kind of modal typing discipline. However, no consensus ever emerged for exactly how to do this. The most developed branch of directed type theory (initiated by Riehl and Shulman [30]) avoids these issues by adopting a more indirect approach inspired by simplicial spaces. The resulting theory provides a synthetic language for $\infty$-categories, albeit at the cost of a more elaborate, multi-layered theory.

To synthetically study 0- and 1-categories, however, a more modest theory will suffice. This was true in the undirected/groupoidal case: the *setoid model* [15, 2] provides semantics for a synthetic theory of setoids (i.e. 0-groupoids) and the *groupoid model* [17] provides semantics for a synthetic theory of 1-groupoids. Both of these models are an instance of *categories with families (CwFs)* [12], which are the *generalized algebraic theory (GAT)* [9] articulation of the semantics of type theory. Working within the framework of GATs comes with numerous


30th International Conference on Types for Proofs and Programs (TYPES 2024).
Editors: Rasmus Ejlers Møgelberg and Benno van den Berg; Article No. 7; pp. 7:1–7:23
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

advantages: the universal algebraic features of GATs – such as homomorphisms, displayed models, products and coproducts of models, free and cofree models – are well-understood [20]. In particular, every GAT has an initial model, the *syntax model*, which can be constructed as a quotient inductive-inductive type [18]. Adopting CwFs as our notion of 'model of type theory' makes all this machinery available.

The purpose of the present work is, then, to propose a synthetic theory of 1-categories, whose semantics are given as a generalized algebraic theory, i.e. as a CwF with additional structure. Our intended model will be the *category model*, which is directly analogous to the groupoid and setoid models. We closely mirror the kind of metatheoretic arguments made by Hofmann and Streicher [17] and develop a directed analogue of their *universe extensionality*, an early statement of univalence. Like earlier directed type theories, our theory adopts a modal typing discipline to control variance (and prevent the theory from "collapsing" into undirected type theory); our hope is that a GAT-based study of this (relatively simple) modal type theory will help foster a connection between the study of multi-modal type theories [13] and *second-order* GATs [33, 32, 8, 19], though connecting the present work to either of these theories is left to future study.

## 1.1 Related Work

The present work draws some constructs from the theory of Licata and Harper [22], particularly their treatment of the *opposite category* construction as a modality on contexts and context extension, as well as their treatment of Π-types in the directed setting. Our theory is closest to that of North [28]; in particular, her semantics matches our 'category model' in the interpretation of the basic structural rules of type theory. However, we have different approaches to typing refl within our modal typing discipline: North restricts to groupoid *types* whereas we restrict to groupoid *contexts*. Recent extensions to North's theory [11] *do* include modalities on contexts, suggesting that we may be converging on a common theory.

Our theory, like North's, is "1-dimensional" in the sense of Licata-Harper in that we maintain *judgmental* equality as a symmetric notion, as opposed to "2-dimensional" theories [22, 29, 1] which introduce a theory of *directed reductions*. As with the identity types in Martin-Löf type theory, all our category-theoretic structure is emergent from the hom-type formation rule and all our reasoning about hom-types is done by the directed J-rule; this is in contrast to directed type theories (such as [22] and [27]) which must build in the synthetic category structure "by hand". As mentioned, we adopt a modal typing discipline for handling variances, unlike the theories of [30, 21, 37, 37, 14] and [36], which adopt approaches akin to simplicial and cubical type theories, respectively.

## 1.2 Contribution and Organization

We present a directed type theory satisfying the following constraints.

**(1)** It is presented as a generalized algebraic theory – our notion of "model" is a (GAT) extension of the GAT of *categories with families*.

**(2)** It is 1-dimensional in the sense of Licata-Harper: there are no 'directed reductions' introduced judgmentally.

**(3)** It is *deeply-polarized*: there is a modal typing discipline to keep track of variances, which operates not just on types but on contexts, substitutions, and context extension.

**(4)** The directed J-rule (directed path induction) permits reasoning about arbitrary terms of hom-types.

**(5)** Hom-types can be iterated, expressing synthetic higher categorical structure (though in the present work we only consider 1-category theoretic structure).

To our knowledge, there is no existing type theory satisfying all these criteria. The Licata-Harper theory satisfies (3), but their hom-types are not inductively structured (ruling out (4)), and cannot be iterated (ruling out (5)). North's theory satisfies (2), (4), and (5), but only has what we'll call *shallow polarity* – negation on types, but not contexts and substitutions – and consequently cannot have Π-types. The simplicial type theory of Riehl-Shulman does not satisfy (3) or (4) – they don't adopt a polarity calculus in the sense contemplated here, nor assert hom-types as primitive with judgmental rules – but do satisfy (2) and particularly (5) – it is in this sense that they provide a synthetic theory of $\infty$-categories. No directed type theory, to our knowledge, explicitly satisfies (1), though some of their model notions may be equivalent to an extension of the GAT of CwFs.

Starting with Section 2, we adopt a semantics-driven approach by investigating a particular model, the *category model* and abstracting its key features into a series of abstract notions of model (Section 3). These notions are all GATs (indeed, CwFs with additional structure), and therefore each give rise to a syntax model. Our main notion is that of a *Directed CwF (DCwF)*, a generalized algebraic theory of directed types with adequate polarity structure to properly track variances.

Achieving (4) while maintaining a modal typing discipline requires a novel approach. In the typing rules of existing directed type theories (including ours), the endpoint terms $t$ and $t'$ of a hom-type $\mathsf{Hom}(t, t')$ are assigned opposite variances: $t$ negative and $t'$ positive. However, this poses a difficulty for typing the identity morphism $\mathsf{refl}_t : \mathsf{Hom}(t, t)$ since $t$ must assume both variances. North [28] solves this by restricting $t$ to be a term of a *core type* (interpreted semantically by groupoids), but the consequent J-rule only operates on hom-terms with a core endpoint, not arbitrary ones. Our solution instead uses groupoid *contexts* rather than groupoid *types*.

In Section 4, we show that this is a viable framework for conducting synthetic category theory. In this section, we adopt an informal style reminiscent of [34], showing how this theory can be operated and how the groupoid context can be carefully maintained by a simple syntactic rule. We use our directed J-rule to give several basic constructions in synthetic (1-)category theory.

Finally, we consider the directed universe of sets in the category model, which serves as the category of sets. The existence of a directed universe allows us to make the metatheoretic argument that the syntax of DCwFs cannot prove the symmetry of hom-types (i.e. this is a genuinely *directed* type theory) or the uniqueness of homs (analogous to Hofmann and Streicher's proof that the groupoid model refutes the uniqueness of identity proofs). We conclude by sketching several possible routes for further study.

An expanded version of this work with more detailed calculations and further discussion is available on the arXiv, as [26]. A much more thorough development – including more careful development of the neutral-context method, Σ-types, and synthetic category theory in (1,1)-directed type theory – is given in the first author's forthcoming PhD thesis [25].

## 1.3 Metatheory and Notation

Throughout, we work in an informal type-theoretic metatheory, using pseudo-Agda notation to specify GATs, make category-theoretic constructions, and define terms in the syntax of Directed CwFs. We use the notations

$$(x\colon X) \to P(x) \qquad \text{and} \qquad (x\colon X) \times P(x)$$

for the dependent function and dependent sum types, respectively. When defining dependent functions, we'll enclose arguments in curly brackets to indicate that they're implicit. Any variables appearing free are also assumed to be implicitly universally quantified. We sometimes use underscores to indicate where the arguments to a function are written.

For equality, we'll use the Agda convention (which, note, is the opposite of the convention used in the HoTT Book [34]): use $=$ to mean *definitional* or *judgmental* equality in our metatheory, whereas $\equiv$ means *propositional* equality (though there's no reason they couldn't coincide, i.e. in an extensional metatheory). We tacitly make use of appropriate extensionality principles (particularly function extensionality) for both notions of equality, and the uniqueness of identity proofs for $\equiv$. We write $\mathsf{Prop}$ for the type of h-propositions in our metatheory, i.e. those $P$ such that $p \equiv p'$ for all $p, p' : P$.

We assume basic familiarity with category theory. The set of objects of a category $\Gamma$ is denoted $|\Gamma|$, the set of $\Gamma$-morphisms from $\gamma_0$ to $\gamma_1$ is denoted $\Gamma\,[\gamma_0, \gamma_1]$, and identities are written as $\mathsf{id}$. The *discrete* groupoid/category on a set $X$ is the category whose objects are elements of $X$ and whose morphisms from $x_0$ to $x_1$ are inhabitants of the identity type $x_0 \equiv x_1$. The *opposite category* construction is understood to be definitionally involutive, i.e. $|\Gamma^{\mathrm{op}}|$ is defined to be $|\Gamma|$ and $\Gamma^{\mathrm{op}}\,[\gamma_0, \gamma_1]$ is defined to be $\Gamma\,[\gamma_1, \gamma_0]$, and thus $(\Gamma^{\mathrm{op}})^{\mathrm{op}} = \Gamma$. We'll use the notation $\mathcal{C} \Rightarrow \mathcal{D}$ for the type of functors from $\mathcal{C}$ to $\mathcal{D}$.

## 2 The Category Interpretation of Type Theory

As mentioned, generalized algebraic theories (GATs) are a desirable formalism for expressing models of type theory, particularly when modelling numerous extensions to a 'basic' type theory. When a theory is given as a GAT, all operations and equations are made clear and explicit, making it easier to compare and contrast similar theories. The theory of **Categories with Families (CwFs)** (originally defined by Dybjer [12]) present the fundamental operations of type theory – contexts, variables, terms, types, and substitutions – encoded as a GAT; upon this basic framework, an endless variety of different type theories can be studied.

We assume basic familiarity with CwFs. The main components of a CwF are a category $\mathsf{Con}$ of *contexts*, whose morphisms are called *substitutions* (write $\mathsf{Sub}\ \Delta\ \Gamma$ for the set of substitutions from $\Delta$ to $\Gamma$); a presheaf $\mathsf{Ty}$ on $\mathsf{Con}$ (whose morphism part we denote $\_[\_]$)[1] and a dependent presheaf $\mathsf{Tm}$ over $\mathsf{Ty}$;[2] and a *context extension* operation $\triangleright$ guaranteeing that $\mathsf{Tm}$ is *locally representable* (in the sense of [8]). That is, there is an isomorphism (natural in $\Delta$) between the type of pairs $(\sigma, t)$ with $\sigma : \mathsf{Sub}\ \Delta\ \Gamma$ and $t : \mathsf{Tm}(\Delta, A[\sigma])$ and the type of substitutions from $\Delta$ to $\Gamma \triangleright A$. The left-to-right direction of this isomorphism is denoted $\langle \_, \_ \rangle$ and the opposite direction as $\mathsf{p} \circ \_, \mathsf{v}[\_]$, so

$$\tau \equiv \langle \mathsf{p} \circ \tau, \mathsf{v}[\tau] \rangle \qquad \text{and} \qquad \sigma \equiv \mathsf{p} \circ \langle \sigma, t \rangle \qquad \text{and} \qquad t \equiv \mathsf{v}[\langle \sigma, t \rangle]$$

for any $\sigma, t$ as above and $\tau : \mathsf{Sub}\ \Delta\ (\Gamma \triangleright A)$.

Two paradigm examples of CwFs are the *setoid model* of [15, 2] and the *groupoid model* of [17]. In the former, the contexts are setoids (i.e. sets equipped with equivalence relations), the types are families of setoids (functorially) indexed over their context setoid, and terms are given by the appropriate notion of *section* of their type (see the Coq formalization of [3] for a precise definition). The groupoid model is quite similar: contexts are groupoids, types are families of groupoids functorially indexed over their context groupoid, and terms are the appropriate notion of section. Indeed, we can view the groupoid model as generalizing the setoid model: a setoid can be viewed as a groupoid whose hom-sets are *subsingletons* (or *propositions*, in the terminology of [34]), sets with at most one element. In other words, the groupoid model is what results when the assumption of *proof-irrelevance* is dropped from the setoid model.

---

[1] I.e. if $\sigma : \mathsf{Sub}\ \Delta\ \Gamma$ and $A : \mathsf{Ty}\ \Gamma$, then $A[\sigma] : \mathsf{Ty}\ \Delta$.
[2] That is, a presheaf on the category of elements of $\mathsf{Ty}$. Its morphism part is also denoted $\_[\_]$, so if $t : \mathsf{Tm}(\Gamma, A)$, then $t[\sigma] : \mathsf{Tm}(\Delta, A[\sigma])$.

Both these models provide interpretations for numerous type formers, in particular the dependent types, identity types, and universes characteristic of Martin-Löf Type Theory [23, 24]. The difference in these models is reflected in the type theories they interpret: while both models permit arbitrary iteration of the identity type former (expressing identities between identities, and identities between identities between identities, and so on), these iterated identity types become trivial more quickly in the setoid model. More precisely, the setoid model validates the *uniqueness of identity proofs* principle, meaning that any two terms of an identity type, $p, q \colon \mathsf{Tm}(\Gamma, \mathsf{Id}(x, y))$ are themselves identical, $\mathsf{UIP}(p, q) \colon \mathsf{Tm}(\Gamma, \mathsf{Id}(p, q))$. The groupoid model famously violates this principle: in the type theory of the groupoid model, there are types (in particular, the universe of sets) which are not *h-sets*, i.e. they possess terms which are proved identical by multiple, *distinct* identity proofs.

This provides a roadmap for how we might develop a model of directed type theory. Since directed type theory can be described as "dependent type theory, but with *asymmetric* identity types", this leads us to suspect that models of directed type theory will result if we simply drop the assumption of symmetry from the setoid and groupoid models. A setoid without symmetry is a preorder, and a groupoid without symmetry is a category. A close inspection of the definition of the groupoid model reveals that nothing in its interpretation of *just the CwF structure* requires symmetry (i.e. that morphisms are invertible),[3] and thus we can define the preorder model of type theory and the **category model of type theory**. The category model is just a generalization of the groupoid model, obtained by dropping symmetry: contexts are categories (and substitutions are functors); types are families of categories and terms are sections, as written more precisely in Figure 1. We won't focus on the preorder model here, but leave it to future work to develop the directed analogue of setoid-model-specific considerations (e.g. [3]). Instead, we'll highlight those features of the category model which are relevant for modelling directed type theory, before abstracting those features into the notion of a *directed CwF* in the next section.

```
-- A : Ty Γ means A : Γ → Cat
record Ty (Γ : Con) : Set where
  field
    obj : |Γ| → Cat
    map : Γ [ γ₀, γ₁ ] → Cat [ obj γ₀, obj γ₁ ]
    fid : map (id_γ) ≡ id_obj(γ)
    fcomp : map (γ₁₂ ∘ γ₀₁) ≡ (map γ₁₂) ∘ (map γ₀₁)

record Tm (Γ : Con) (A : Ty Γ) : Set where
  field
    obj : (γ : |Γ|) → |A(γ)|
    map : (γ₀₁ : Γ [ γ₀, γ₁ ]) → (A γ₁) [A γ₀₁ (obj γ₀), obj(γ₁)]
    fid : map (id_γ) ≡ id_obj(γ)
    fcomp : map (γ₁₂ ∘ γ₀₁) ≡ (map γ₁₂) ∘ (A γ₁₂ (map γ₀₁))
```

**Figure 1** The family structure of the category model.

[3] Indeed, [18, Sect. 7] shows that any GAT gives rise to a CwF of algebras and displayed algebras. The groupoid model and setoid model are not literally instances of this construction, but can be viewed as the CwF of groupoids and displayed groupoids (resp. setoids and displayed setoids) with added fibrancy conditions attached (see [25, Chap. 2]). The same can be done with preorders and categories, yielding the preorder and category models.

While the basic CwF structure of the groupoid model doesn't require symmetry (i.e. that all morphisms are invertible), its interpretations of further type formers certainly do. After all, our hope is that by passing from the groupoid model to the category model, the symmetric identity types of the former will become asymmetric *hom-types* in the latter. Consider the semantics of the identity type former in the groupoid model. Here, and henceforth, we define a type (in this case $\mathsf{Id}(t, t')$) by giving its object- and morphism-parts, which are both written as just $\mathsf{Id}(t, t')$.

```
Id : Tm(Γ, A) → Tm(Γ, A) → Ty Γ      −− Taken from [17, Section 4.10]
  (Id(t,t')) γ = (A γ) [t γ, t' γ]      −− Discrete groupoid

  (Id(t,t')) (γ01 : Γ[ γ0 , γ1 ]) : (A γ0) [t γ0, t' γ0] → (A γ1) [t γ1, t' γ1]
  (Id(t,t')) γ01 x0 = (t' γ01) ∘ (A γ01 x0) ∘ (t γ01)^−1
```

Here, the fact that $A(\gamma_1)$ is a *groupoid* is used in an essential way (we must take the inverse of $t(\gamma_{01})$), and hence this definition doesn't work in the category model. But notice the following: the term $t$ is in the "negative" position (the domain) and the term $t'$ is in the "positive" position. Fittingly, we only use the *inverse* of $t(\gamma_{01})$ – never $t(\gamma_{01})$ itself – and only use $t'(\gamma_{01})$ but not its inverse. This observation will provide the key to adapting this definition for the category model.

What is needed is for $t$ to be a *contravariant* term of type $A$, while keeping $t'$ as *covariant*. We can treat this variance in the type theory of the category model, using a fundamental construct from category theory: *opposite categories*. A type $A\colon \mathsf{Ty}\ \Gamma$ in the category model consists of a family of categories $A(\gamma)$ for each object $\gamma\colon\ |\Gamma|$ and a functor $A(\gamma_{01})\colon A(\gamma_0) \Rightarrow A(\gamma_1)$ for each morphism $\gamma_{01}\colon \Gamma[\gamma_0, \gamma_1]$. Given such a family of categories $A$, we can form a new family $A^-$, where $A^-(\gamma)$ is defined as the opposite category of $A(\gamma)$. This extends to the morphism part as well, because any functor $f\colon C \Rightarrow D$ can be viewed as a functor on their opposites, $f\colon C^{\mathrm{op}} \Rightarrow D^{\mathrm{op}}$. Alternatively, we could view $A$ as a functor $\Gamma \Rightarrow \mathsf{Cat}$, and define $A^-$ to be the composition of $A$ with the endofunctor $(\_)^{\mathrm{op}}\colon \mathsf{Cat} \Rightarrow \mathsf{Cat}$. We can state generally that the category model validates the following rule:

$$\frac{A\colon \mathsf{Ty}\ \Gamma}{A^-\colon \mathsf{Ty}\ \Gamma.}$$

If $t\colon \mathsf{Tm}(\Gamma, A^-)$, this means that the object part of $t$ will still send objects $\gamma\colon\ |\Gamma|$ to objects of $A(\gamma)$, since $A(\gamma)$ and $A^-(\gamma)$ have the same objects. But observe the type of its morphism part:

$$t\colon (\gamma_{01}\colon \Gamma\ [\gamma_0, \gamma_1]) \to (A\ \gamma_1)\ [t\ \gamma_1, A\ \gamma_{01}\ (t\ \gamma_0)].$$

This is precisely what we need to give the definition of hom-types in the category model: see Figure 2. This definition is almost exactly the same as the semantics of $\mathsf{Id}$ in the groupoid model, but with $t$ changed to be a term of $A^-$, thus eliminating the need for the categories $A(\gamma_i)$ to be groupoids. Here's the hom-type formation, expressed as a rule:

$$\frac{t\colon \mathsf{Tm}(\Gamma, A^-) \quad t'\colon \mathsf{Tm}(\Gamma, A)}{\mathsf{Hom}(t, t')\colon \mathsf{Ty}\ \Gamma.}$$

The type annotation of $t$ as a "negative" term and the implicit annotation of $t'$ as "positive" serve as a kind of *modal typing discipline* for keeping track of the *variances* of terms.

For now, we just state the formation rule for hom-types; introducing and eliminating terms of hom-types will require more machinery. To see what kind of machinery, let's instead consider dependent function types. Like with the formation of hom-types, Π-types involve

```
Hom : Tm(Γ, A⁻) → Tm(Γ, A) → Ty Γ
  (Hom(t,t')) γ = (A γ) [t γ, t' γ] -- Discrete category

  (Hom(t,t')) (γ01 : Γ[ γ0 , γ1 ]) : (A γ0) [t γ0, t' γ0] → (A γ1) [t γ1, t' γ1]
  (Hom(t,t')) γ01 x0 = (t' γ01) ∘ (A γ01 x0) ∘ (t γ01)
```

**Figure 2** Semantics of the Hom-type former in the category model.

positive and negative "variance": a function is contravariant in its argument and covariant in its result. Therefore, as we might expect, the interpretation of Π-types in the groupoid model ([17, Section 4.6]) makes essential use of the invertibility of morphisms in a groupoid. Again, it only comes into play when defining the *morphism* part of the interpretation: the object part (reproduced in Figure 4) defines for each $\gamma\colon |\Gamma|$ an auxiliary type $B_\gamma$ in context $A(\gamma)$, and then specifies the category $\Pi(A,B)\ \gamma$ with terms $\theta\colon \mathsf{Tm}(A(\gamma), B_\gamma)$ as objects. This works fine in the category model. However, defining the morphism part of $\Pi(A,B)$ requires a kind of negative variance *deeper* than the shallow contravariance of $A^-$: in the type $\mathsf{Hom}(t,t')$ it was a *term* that occurred negatively ($t$), in the type $\Pi(A,B)$ it's a *type* ($A$) that occurs negatively.

To make sense of this, we must consider the *opposite category* operation, not just as acting on each $A(\gamma)$ in a family of categories over a context $\Gamma$, but as acting on the contexts themselves. In the category model, we have the following rules.

$$\frac{\Gamma\colon \mathsf{Con}}{\Gamma^-\colon \mathsf{Con}} \qquad \frac{\sigma\colon \mathsf{Sub}\ \Delta\ \Gamma}{\sigma^-\colon \mathsf{Sub}\ \Delta^-\ \Gamma^-}$$

That is, we can negate contexts and substitutions as well as types: $\Gamma^-$ is interpreted as $\Gamma^{\mathrm{op}}$, and this operation is (*co*variantly) lifted onto functors as before. Now consider the difference in the morphism parts of terms with these different kinds of variance.

```
-- t : Tm(Γ, A) where A : Ty Γ
t γ01 : (A γ1) [ A γ01 (t γ0), t γ1 ]
-- t : Tm(Γ, A⁻) where A : Ty Γ
t γ01 : (A γ1) [ t γ1, A γ01 (t γ0) ]
-- t : Tm(Γ⁻, A) where A : Ty Γ⁻
t γ01 : (A γ0) [ t γ0, A γ01 (t γ1) ]
```

This is why we referred to this as "shallow" and "deep" negation: the difference between the first two is that we've flipped around each $A(\gamma)$, whereas in the third term, the dependence of $A$ on $\Gamma$ has itself been flipped around ($A$ is now contravariant, so $A\ \gamma_{01}$ takes objects of $A(\gamma_1)$ to objects of $A(\gamma_0)$). It is this latter kind of contravariance that describes $A$'s position in $\Pi(A,B)$.

With this in mind, we might try to state the Π-formation rule as follows

$$\frac{A\colon \mathsf{Ty}(\Gamma^-) \quad B\colon \mathsf{Ty}(\Gamma^- \triangleright A)}{\Pi(A,B)\colon \mathsf{Ty}\ \Gamma.}$$

However, this doesn't get the variances quite right. The type $A$ depends negatively on $\Gamma$, as desired, but so does $B$: the morphism part of some $B\colon \mathsf{Ty}(\Gamma^- \triangleright A)$ has shape

$$B\colon (\gamma_{01}\colon \Gamma^-\ [\gamma_1,\gamma_0]) \to (x_0\colon (A\ \gamma_0)\ [A\ \gamma_{01}\ a_1, a_0]) \to B(\gamma_1,a_1) \Rightarrow B(\gamma_0,a_0).$$

This is too much contravariance. Semantically, to define the morphism part of $\Pi(A, B)$ – as we do precisely in [26] – we need $B$ to depend *covariantly* on $\Gamma$, but $A$ to depend contravariantly on $\Gamma$. Syntactically, this matches our understanding of $A$ being in "negative position" and $B$ "positive position" in the type $\Pi(A, B)$.

To achieve the desired arrangement of variances in the (modified) CwF we're building, we must introduce not one, but *two* notions of context extension. The first, or positive context extension, is the usual one: when $A\colon \mathsf{Ty}\ \Gamma$ then $\Gamma$ may be covariantly extended by $A$. As explicitly spelled out in Figure 3, the resulting context (which we'll henceforth denote $\Gamma \rhd^{+} A$, and use $\langle \_,_{+} \_\rangle$ for its associated pairing operation and $\mathsf{p}_{+}$ for the substitution from $\Gamma \rhd^{+} A$ to $\Gamma$), is interpreted in the category model by a Grothendieck construction for the category-valued covariant functor $A$. But if instead $A\colon \mathsf{Ty}\ \Gamma^{-}$, then, instead of forming the context $\Gamma^{-} \rhd^{+} A$ – which, as discussed above, has the wrong arrangement of variances – we could instead form $\Gamma \rhd^{-} A$, the negative context extension by $A$. A type $B$ in context $\Gamma \rhd^{-} A$ would have our desired variances for $\Pi$-types: $B$ depends covariantly on both $\Gamma$, even though $A$ depends contravariantly on $\Gamma$.

More concretely, the negative context extension operator satisfies the following "*modal* local representability", a version of the local representability condition required of the context extension operator of a CwF, which incorporates context negation, type negation, and negative context extension:

$$\mathsf{Sub}\ \Delta\ (\Gamma \rhd^{-} A) \cong (\sigma\colon \mathsf{Sub}\ \Delta\ \Gamma) \times (\mathsf{Tm}(\Delta^{-}, A[\sigma^{-}]^{-})) \tag{1}$$

for any $A\colon \mathsf{Ty}\ \Gamma^{-}$. We'll write $\langle \_,_{-} \_\rangle$ for the right-to-left direction of this isomorphism, and write $\mathsf{p}_{-,A}\colon \mathsf{Sub}\ (\Gamma \rhd^{-} A)\ \Gamma$ and $\mathsf{v}_{-,A}\colon \mathsf{Tm}((\Gamma \rhd^{-} A)^{-}, A[\mathsf{p}_{-,A}^{-}]^{-})$ for the data obtained from applying the left-to-right direction to the identity morphism on $\Gamma \rhd^{-} A$.

Notice that the $\sigma$ on the left is from $\Delta$ to $\Gamma$ (not $\Gamma^{-}$), and that a $\Gamma \rhd^{-} A$ morphism from $(\gamma_0, a_0)$ to $(\gamma_1, a_1)$ consists of a $\Gamma$-morphism from $\gamma_0$ to $\gamma_1$ (not $\gamma_1$ to $\gamma_0$). This is what we mean when we say types and terms in $\Gamma \rhd^{-} A$ depend positively on $\Gamma$, but $A$ depends negatively on $\Gamma$: a substitution into $\Gamma \rhd^{-} A$ must consist of a substitution into $\Gamma$ – not $\Gamma^{-}$ – paired with a context- *and* type-negated term. Though perhaps initially daunting, the number of negations on the right-hand side become much less of a problem later on, since for our practical purpose we'll restrict to the case where $\Gamma$ is a groupoid.

With this, we have everything needed to give the semantics of the $\Pi$-type former; this is partly done in Figure 4 – see [26, Appendix A] for full detail. The formation rule is

$$\frac{A\colon \mathsf{Ty}(\Gamma^{-}) \quad B\colon \mathsf{Ty}(\Gamma \rhd^{-} A)}{\Pi(A, B)\colon \mathsf{Ty}\ \Gamma} \tag{2}$$

As expected for $\Pi$-types, we have an isomorphism between $\mathsf{Tm}(\Gamma, \Pi(A, B))$ and $\mathsf{Tm}(\Gamma \rhd^{-} A, B)$, the application and lambda-abstraction rules. This is omitted here for reasons of space, but included in the appendix of [26], along with accompanying calculations.

Let's now return to the key feature of directed type theory, hom-types. Above, we gave just the formation rule for hom-types, but said nothing of how to introduce or eliminate terms of this type. Stating the introduction rule, the term $\mathsf{refl}$ inhabiting $\mathsf{Hom}(t, t)$ for each term $t$ proves rather subtle. The difficulty stems from the following *mixed-variance problem*: since our formation rule demands the domain term $t$ be of type $A^{-}$ and the codomain term $t'$ to be of type $A$, it's not immediately clear how to make $\mathsf{Hom}(t, t)$ well-formed. There is, in general, no way to coerce terms of type $A$ into terms of type $A^{-}$ or vice versa, and we have no rule permitting us to use a term in both variances.

```
_▷⁺_ : (Γ : Con) → Ty Γ → Con
  |Γ ▷⁺ A| = (γ : |Γ|) × |A γ|
  (Γ ▷⁺ A) [ (γ₀, a₀) , (γ₁, a₁) ] = (γ₀₁ : Γ [ γ₀, γ₁ ]) × (A γ₁) [ A γ₀₁ a₀, a₁ ]

_▷⁻_ : (Γ : Con) → Ty Γ⁻ → Con
  |Γ ▷⁻ A| = (γ : |Γ|) × |A γ|
  (Γ ▷⁻ A) [ (γ₀, a₀) , (γ₁, a₁) ] = (γ₀₁ : Γ [ γ₀, γ₁ ]) × (A γ₁) [ a₀ , A γ₀₁ a₁ ]
```

**Figure 3** Semantics of context extension in the category model.

```
Π : (A : Ty Γ⁻) → Ty(Γ ▷⁻ A) → Ty Γ
  |Π(A,B) γ| = Tm(A(γ), B_γ)
   where
    B_γ : Ty(A γ)
      B_γ a = B(γ,a)
      B_γ (x : (A γ)[ a , a' ]) = B(id_γ,x)
```

**Figure 4** Semantics of the Π-type former in the category model.

A solution which avoids this shortcoming is revealed by considering hom-types in the empty context. In the empty context, a type $A$ is the same thing[4] as a category, and a term of type $A$ is the same thing[5] as an object of type $A$. So then there's no difference between terms of type $A$ and terms of type $A^-$, since a category and its opposite have the same objects. Therefore, in the empty context, there is no mixed-variance problem, and we can state the introduction rule for $\mathsf{refl}_t$ simply by coercing $t$ to be positive and negative as needed.

This doesn't extend to arbitrary contexts: as we saw above, terms $t\colon \mathsf{Tm}(\Gamma, A^-)$ and $t'\colon \mathsf{Tm}(\Gamma, A)$ have different morphism parts. But here's the key observation: if $\Gamma$ is a *groupoid*, then we can still coerce between $A$ and $A^-$: given $t\colon \mathsf{Tm}(\Gamma, A^-)$, we can obtain $-t\colon \mathsf{Tm}(\Gamma, A)$, and vice-versa. The definition is given in Figure 5; there (and henceforth), we use $\Gamma\colon \mathsf{NeutCon}$ to indicate that $\Gamma$ is a groupoid, and therefore can invert $\Gamma$-morphisms as needed. So, rather than introduce a new type $A^0$ whose terms can be either positive or negative, we instead have identified those contexts – neutral contexts – where terms of the familiar types $A$ and $A^-$ can be inter-converted. Given this, we can introduce $\mathsf{refl}$:

$$\frac{\Gamma\colon \mathsf{NeutCon} \quad A\colon \mathsf{Ty}\,\Gamma \quad t\colon \mathsf{Tm}(\Gamma, A^-)}{\mathsf{refl}_t\colon \mathsf{Tm}(\Gamma, \mathsf{Hom}(t, -t)).}$$

We only need to assert $\mathsf{refl}_t$ for $t$ of type $A^-$, because the analogous rule for $t'$ of type $A$ can be derived: given $t'\colon \mathsf{Tm}(\Gamma, A)$, we observe that $t' = -(-t')$, so $\mathsf{refl}_{-t'}\colon \mathsf{Tm}(\Gamma, \mathsf{Hom}(-t', t'))$.

The solution to this problem proposed by North [28] is not groupoid *contexts*, rather to use *(core) groupoid types*. This solution consists of asserting a new type $A^0$ for each $A$ (interpreted in the category model as the (fiberwise) "core groupoid" of $A$), equipped with coercions $\mathsf{Tm}(\Gamma, A^0) \to \mathsf{Tm}(\Gamma, A)$ and $\mathsf{Tm}(\Gamma, A^0) \to \mathsf{Tm}(\Gamma, A^-)$. Then, for a term

[4] Silently coerce along $\mathbb{1} \Rightarrow \mathsf{Cat} \cong \mathsf{Cat}$.
[5] Silently coerce along $\mathbb{1} \Rightarrow A \cong |A|$ for any category $A$.

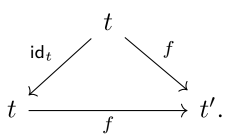
```
− : {Γ : NeutCon}{A : Ty Γ} → Tm(Γ,A) → Tm(Γ,A⁻)
  −t' γ = t' γ
  −t' γ01 = A γ01 (t'(γ01⁻¹))
− : {Γ : NeutCon}{A : Ty Γ} → Tm(Γ,A⁻) → Tm(Γ,A)
  −t γ = t γ
  −t γ01 = A γ01 (t(γ01⁻¹))
```

**Figure 5** Semantics of neutral-context coercion in the category model.

$t\colon \mathsf{Tm}(\Gamma, A^0)$, it makes sense to write $\mathsf{Hom}(t, t)$, as $t$ can be coerced to both the positive and negative modality, in order to fit the Hom formation rule. From there, a directed J-rule[6] can be stated for eliminating hom terms. The issue with this solution is that it forces homs to have core endpoints: the directed J-rule can *only* be used to prove claims about homs anchored at a term of type $A^0$ – the only terms $t$ for which $\mathsf{refl}_t$ can even be formed. *A priori*, it's not clear that proofs about arbitrary homs can be made with such a J-rule (recall that this was a key desideratum, number 4, for the present theory).

North's solution has the advantage of working in an arbitrary, non-groupoid context. But we find that the restriction to neutral contexts is needed anyway: as developed in [25, Chap. 2], the Π-types of the category model are generally unworkable outside of a groupoid context. Below, when we define the identity function and the composition of functions, notice that we make use of the groupoid-context facilities to do so. Moreover, characterizing the opposites and hom-types of Σ-types, defining the morphism part of functions as synthetic functors, and working with the universe of sets also need a groupoid context to work. So we might as well use this assumption to state refl too. We suspect that, if we work in a groupoid context but *also* have access to North's core types, then our rule of directed path induction (see below) becomes equivalent to North's, so the theories are fundamentally compatible.

Let's conclude this section by giving an eliminator for our hom-type, known as the *directed J-rule* or *directed path induction*. Following [17, Section 4.10], we study directed path induction in the empty context first, which can then be extended to an arbitrary neutral context. Given $A\colon \mathsf{Ty}\,\bullet$ and $t\colon \mathsf{Tm}(\bullet, A^-)$ and some $M\colon \mathsf{Ty}(\bullet \rhd^+ A \rhd^+ \mathsf{Hom}(t[\mathsf{p}_A], \mathsf{v}))$, our goal is to be able to prove $M[\mathsf{id},_+ t',_+ f]$ for arbitrary $t'$ and $f$, just by supplying a term $m$ of $M[-t, \mathsf{refl}_t]$. Translated into the category model semantics: $A$ is a category, $t$ and $t'$ are objects of $A$, $M$ is a functor from the coslice category $t/A$ into $\mathsf{Cat}$, $f$ is an $A$-morphism from $t$ to $t'$, and $m$ is an object of the category $M(t, \mathsf{id}_t)$. The key observation is that $f$ is then *also* a morphism in the coslice category from $(t, \mathsf{id})$ to $(t', f)$.




Therefore,

$$M(f)\colon M(t, \mathsf{id}_t) \Rightarrow M(t', f)$$

and so the object part of this functor turns objects of $M(t, \mathsf{id}_t)$ into objects of $M(t', f)$, that is, it turns terms $m\colon \mathsf{Tm}(\bullet, M[\mathsf{id},_+ -t,_+ \mathsf{refl}])$ into terms

$$(\mathsf{J}_{t,M}\ m)\ [\mathsf{id},_+ t',_+ f]\colon \mathsf{Tm}(\bullet, M[\mathsf{id},_+ t',_+ f]).$$

[6] Or, rather, two directed J-rules

And, since $M(\mathsf{id})$ is the identity functor, we have the $\beta$ law, saying that $\mathsf{J}_{t,M}\ m\,[\mathsf{id},_+ -t,_+ \mathsf{refl}_t] \equiv m$. The general law replaces $\bullet$ with an arbitrary neutral context:

$$\frac{\Gamma\colon \mathsf{NeutCon} \quad t\colon \mathsf{Tm}(\Gamma, A^-) \quad m\colon \mathsf{Tm}(\Gamma, M[\langle \mathsf{id},_+ -t,_+ \mathsf{refl}_t\rangle]) \qquad A\colon \mathsf{Ty}\,\Gamma \quad M\colon \mathsf{Ty}(\Gamma \triangleright^+ A \triangleright^+ \mathsf{Hom}(t[\mathsf{p}_A], \mathsf{v}))}{\mathsf{J}_{t,M}\ m : \mathsf{Tm}(\Gamma \triangleright^+ A \triangleright^+ \mathsf{Hom}(t[\mathsf{p}_A], \mathsf{v}), M)} \tag{3}$$

but the category model interpretation – see Figure 6 – essentially follows this same idea. If $M$ doesn't need to depend on the term of type $\mathsf{Hom}(t, \mathsf{v})$, then we can instead use the simpler rule

$$\frac{\Gamma\colon \mathsf{NeutCon} \quad A\colon \mathsf{Ty}\,\Gamma \quad t\colon \mathsf{Tm}(\Gamma, A^-) \quad M\colon \mathsf{Ty}(\Gamma \triangleright^+ A) \quad m\colon \mathsf{Tm}(\Gamma, M[\mathsf{id},_+ -t])}{\mathsf{J}_{t,M}\ m : \mathsf{Tm}(\Gamma \triangleright^+ A \triangleright^+ \mathsf{Hom}(t[\mathsf{p}_A], \mathsf{v}), M[\mathsf{p}_{\mathsf{Hom}(t[\mathsf{p}_A],\mathsf{v})}])} \tag{4}$$

In Section 4 we put this rule to use in synthetic category theory constructions and proofs.

```
J : (t : Tm(Γ, A⁻)) → (M : Ty (Γ ▷⁺ A ▷⁺ Hom(t[p_A],v)))
  → Tm(Γ, M[⟨id,₊−t,₊refl_t⟩]) → Tm(Γ ▷⁺ A ▷⁺ Hom(t[p_A],v), M)
  (J_{t,M} m) : (γ : |Γ|) → (a : |A γ|) → (x : (A γ) [t γ , a]) → | M(γ, a, x) |
  (J_{t,M} m) γ a x = M (id_γ, x, ρ) (m γ)  −− ρ : x ∘ A id_γ id_{tγ} ∘ t id_γ ≡ x

  (J_{t,M} m) : (γ_01 : Γ [γ_0,γ_1]) → (a_01 : Aγ_1 [A γ_01 a_0, a_1])
  → (φ_01 : a_01 ∘ (A γ_01 x_0) ∘ t(γ_01) ≡ x_1)
  → M(γ_1, a_1, x_1)[ M(γ_01,a_01,φ_01) ((J_{t,M} m) γ_0 a_0 x_0), ((J_{t,M} m) γ_1 a_1 x_1) ]
  (J_{t,M} m) γ_01 a_01 ρ_01 = M(id_{γ_1}, x_1, ρ_1) (m γ_01)−− ρ_1 : x_1 ∘ A id_{γ_1} id_{tγ_1} ∘ t id_{γ_1} ≡ x_1

Jβ : (J_{t,M}m) [⟨id,₊−t,₊refl_t⟩] ≡ m
```

**Figure 6** Semantics of directed path induction in the category model.

## 3 Directed Categories with Families

The aim of the present work is not just to establish the category model as a suitable interpretation of directed type theory, but to abstract the category model to a general, abstract notion of 'model' of directed type theory. Specifically, we wish to present this model notion as a generalized algebraic theory, that is, as a CwF with further structure. We do so in several stages, progressively capturing more of the structure described in the previous section. In addition to making the complex and multifaceted notion of 'directed CwF' more digestible, this approach will also give us several intermediate notions, each of which is worthy of further study in its own right. First, we encapsulate the 'negation' structure.

▶ **Definition 1** (Polarized CwF). *A **polarized category with families (PCwF)** consists of a CwF $\mathcal{C} = (\mathsf{Con}, \mathsf{Ty}, \mathsf{Tm}, \triangleright^+, \ldots)$ equipped with the following operations.*

- *An endofunctor $(\_)^- : \mathsf{Con} \to \mathsf{Con}$ such that $(\Gamma^-)^- \equiv \Gamma$ and $(\sigma^-)^- \equiv \sigma$ for all $\Gamma$ and $\sigma$*
- *A natural transformation $(\_)^- : \mathsf{Ty} \to \mathsf{Ty}$ such that $(A^-)^- \equiv A$ for all $A$.*

So a PCwF is just a CwF equipped with context-, substitution-, and type-negation involutions. The fact that the type-negation operation is a natural transformation just says that it is stable under substitution, i.e. $A[\sigma]^- \equiv A^-[\sigma]$. Now, notably absent from this definition is the negative context extension operation $\triangleright^-$; by this definition, a PCwF only has the positive one. This is because the negative operation is, in fact, definable: in the category model, the following equation holds for any $\Gamma$ and any $A\colon \mathsf{Ty}\ \Gamma$:

$$(\Gamma \triangleright^+ A)^- \equiv \Gamma^- \triangleright^- A^-. \tag{5}$$

Here we use the fact that $(\Gamma^-)^- \equiv \Gamma$, and hence $A\colon \mathsf{Ty}\ (\Gamma^-)^-$, making the right-hand side well-formed. Consequently, we can turn this equation around to *define* negative context extension: for $A\colon \mathsf{Ty}(\Gamma^-)$, let $\Gamma \triangleright^- A$ be $(\Gamma^- \triangleright^+ A^-)^-$. The isomorphism characterizing $\triangleright^-$ (Equation 1) can then be proved as a consequence of the one for $\triangleright^+$.

Also absent from Definition 1 is any mechanism connecting the context/substitution negation endofunctor to the type-negation operation. It's unclear if this ought to be rectified, or if their connection is just a peculiarity of the category model. Not every CwF fits the same mold of "contexts are structures, types are families of structures", so it's not possible to require in general that the type-negation operation is just post-composition with the context-negation functor. There *are* suitably abstract ways of connecting the two – for instance, we can note that the category model is *democratic* in the sense of [10, Defn. 3]: there is an isomorphism $K$ between contexts $\Gamma$ and closed types; this isomorphism is compatible with both negation operations, in that $K(\Gamma^-) = K(\Gamma)^-$. However, we don't need need such strong assumptions for the results of Section 4, so we omit them from the general definition of PCwFs.

Of course, the category model and the preorder model are both examples of PCwFs, where the negation is the 'opposite' construction. But so are the groupoid and setoid models. Indeed, the groupoid model is a *sub-PCwF* of the category model: a groupoid $\Gamma$ is a category, and so it makes perfect sense to take the opposite category of $\Gamma$, obtaining $\Gamma^-$, which is also a groupoid. What makes the groupoid model a peculiar instance of a PCwF is that $\Gamma \cong \Gamma^-$ for every $\Gamma$, and $A(\gamma) \cong A^-(\gamma)$ for every $A$. It is what we'll call a *symmetric PCwF*. The setoid model is also a symmetric PCwF, but strictly so: there, $\Gamma \equiv \Gamma^-$. The situation exemplified by the groupoid/category and setoid/preorder models – a symmetric sub-PCwF of another PCwF – is what we capture in our next notion.[7]

▶ **Definition 2** (Neutral-Polarized CwF). *A* ***sub-PCwF*** [8] *$\mathcal{D}$ of a PCwF $\mathcal{C}$ consists of predicates*[9] $\mathsf{D}_{\mathsf{Con}}\colon \mathsf{Con} \to \mathsf{Prop}$ *and* $\mathsf{D}_{\mathsf{Ty}}\colon \{\Gamma\colon \mathsf{Con}\} \to \mathsf{Ty}\ \Gamma \to \mathsf{Prop}$ *such that*

- $\mathsf{D}_{\mathsf{Con}}\ \bullet$*;*
- *if* $\mathsf{D}_{\mathsf{Con}}(\Gamma)$*, then* $\mathsf{D}_{\mathsf{Con}}(\Gamma^-)$*;*
- *if* $\mathsf{D}_{\mathsf{Ty}}(A)$*, then* $\mathsf{D}_{\mathsf{Ty}}(A^-)$*;*
- *if* $A\colon \mathsf{Ty}\ \Gamma$ *is such that* $\mathsf{D}_{\mathsf{Ty}}(A)$*, then* $\mathsf{D}_{\mathsf{Ty}}(A[\sigma])$ *for any* $\sigma\colon \mathsf{Sub}\ \Delta\ \Gamma$*; and*
- *if* $\mathsf{D}_{\mathsf{Con}}(\Gamma)$ *and* $\mathsf{D}_{\mathsf{Ty}}(A)$*, then* $\mathsf{D}_{\mathsf{Con}}(\Gamma \triangleright A)$*.*

---

[7] The definition given here is, admittedly, somewhat *ad hoc*. These are just the constructs we'll need in order to make the subsequent definitions and synthetic category theory development proceed. A more detailed explication of this definition is given in [25, Chap. 2].

[8] A more descriptive name might be "full and *locally full* sub-PCwF": it's a *full subcategory* $\mathsf{DCon} \hookrightarrow \mathsf{Con}$ in that $\mathsf{DCon}\ [\Delta, \Gamma] = \mathsf{Sub}\ \Delta\ \Gamma$, but also "locally full" in the sense that $\mathsf{Tm}_{\mathcal{D}}(\Gamma, A) = \mathsf{Tm}_{\mathcal{C}}(\Gamma, A)$ for $\Gamma\colon \mathsf{DCon}$ and $A\colon \mathsf{DTy}(\Gamma)$.

[9] The first part of this definition articulates a notion of "$\mathsf{Prop}$-valued logical predicate" $\mathsf{D} = (\mathsf{D}_{\mathsf{Con}}, \mathsf{D}_{\mathsf{Sub}}, \mathsf{D}_{\mathsf{Ty}}, \mathsf{D}_{\mathsf{Tm}}, \ldots)$ on the PCwF $\mathcal{C}$, except we don't mention the components $\mathsf{D}_{\mathsf{Sub}}$ and $\mathsf{D}_{\mathsf{Tm}}$ because fullness and local fullness means that $\mathsf{D}_{\mathsf{Sub}}$ and $\mathsf{D}_{\mathsf{Tm}}$ (respectively) are *always satisfied.*

*We indicate a sub-PCwF by* $\mathcal{D} = (\mathsf{DCon}, \mathsf{DTy})$ *to indicate that* $\mathsf{DCon}$ *is the subcategory of* $\mathsf{D}_{\mathsf{Con}}$*-contexts, and* $\mathsf{DTy}$ *is the subpresheaf of* $\mathsf{D}_{\mathsf{Ty}}$*-types.*

*A* ***neutral-polarized category with families (NPCwF)*** *consists of a PCwF* $\mathcal{C}$ *and a sub-PCwF* $\mathcal{N} = (\mathsf{NeutCon}, \mathsf{NeutTy})$ *such that*

- $\mathsf{NeutCon}$ *is symmetric: every* $\Gamma : \mathsf{NeutCon}$ *comes equipped with an* $\mathsf{e} : \mathsf{Sub}\,\Gamma^-\,\Gamma$ *such that* $\mathsf{e}^- : \mathsf{Sub}\,\Gamma\,\Gamma^-$ *is an inverse of* $\mathsf{e}$*; moreover, this isomorphism is natural in the sense that, for* $\Delta, \Gamma : \mathsf{NeutCon}$ *and* $\sigma : \mathsf{Sub}\,\Delta\,\Gamma$*, we have* $\sigma \equiv \mathsf{e}_\Gamma \circ \sigma^- \circ \mathsf{e}_\Delta^-$*;*
- *if* $\Gamma$ *is neutral and* $A : \mathsf{Ty}\,\Gamma$*, then there is a coercion operation* $- : \mathsf{Tm}(\Gamma, A^-) \to \mathsf{Tm}(\Gamma, A)$ *such that* $-(-t) \equiv t$ *for all* $t$ *and such that* $(-t)[\sigma] \equiv -(t[\sigma])$ *for any* $\sigma : \mathsf{Sub}\,\Delta\,\Gamma$ *where* $\Delta$ *is neutral;*
- *for every* $\Gamma : \mathsf{NeutCon}$ *and* $A : \mathsf{Ty}\,\Gamma$*,there is an isomorphism*

$$\mathsf{ee} : \Gamma \triangleright^- A[\mathsf{e}] \cong \Gamma \triangleright^+ A$$

  *such that*
  - $\mathsf{p}_{+,A} \circ \mathsf{ee} \equiv \mathsf{p}_{-,A[\mathsf{e}]}$
  - *for every* $\Delta, \Gamma : \mathsf{NeutCon}$*,* $\sigma : \mathsf{Sub}\,\Delta\,\Gamma$*,* $A : \mathsf{Ty}\,\Gamma$ *and every* $s : \mathsf{Tm}(\Delta, A[\sigma])$*,*

$$\mathsf{ee} \circ \langle \sigma,_+ s \rangle \equiv \langle \sigma,_- (-s)[\mathsf{e}] \rangle \,. \tag{6}$$

To summarize: an NPCwF is a PCwF with a (locally) full sub-PCwF of “neutral” contexts and types, which come equipped with machinery for overcoming the polarity calculus (coercing between the context and its opposite, and between a type and its opposite). Our goal is to talk about the category model: we want to take the category theoretic statement “Grpd is a full subcategory of Cat” and extend it to a statement about the semantics of polarized type theory: “the groupoid model is a (full and locally full) symmetric sub-PCwF of the category model”. This is what we do in Definition 2: an “NPCwF” is PCwF equipped with an appropriate symmetric sub-PCwF.

Let us also note that a common feature in directed type theories (e.g. [29, 28]) is to include *core types*, i.e. an operation of the form $(\_)^0 : \mathsf{Ty}\,\Gamma \to \mathsf{NeutTy}\,\Gamma$. In the category model, this is interpreted as applying the *core groupoid* construction to each category $A(\gamma)$, producing a family of groupoids indexed over $\Gamma$. We might as well have a deep version too, and consider the core groupoid construction as operating on contexts $(\_)^0 : \mathsf{Con} \to \mathsf{NeutCon}$ too. So the operation on *contexts* is the coreflector, the right adjoint, of the inclusion $\mathsf{NeutCon} \hookrightarrow \mathsf{Con}$; the operation on *types* is a “local coreflector”, a dependent right adjoint [7] to $\mathsf{NeutTy} \hookrightarrow \mathsf{Ty}$. What this means in the category model (equipped with core types) is that $\mathsf{Tm}(\Gamma, A) \cong \mathsf{Tm}(\Gamma, A^0)$ for every $\Gamma : \mathsf{NeutCon}$ and $A : \mathsf{Ty}\,\Gamma$. So this is another coercion, between terms of $A^0$ and $A$, in addition to the NPCwF coercion operations $\mathsf{Tm}(\Gamma, A^-) \cong \mathsf{Tm}(\Gamma, A)$. This is why the above-mentioned concern about North’s theory – that the directed J-rule is restricted to only core terms – is resolved in a neutral context, because in neutral contexts we can coerce freely between $A$, $A^-$, and $A^0$. We’ll only need *neutral* contexts and types for the present work, not *core* contexts and types, but a thorough study of the latter is certainly needed in order to join the present theory to either North or to multi-modal type theory[13].

Recall that CwFs are not a single notion of model for a single type theory, but rather that CwFs encode the basic structural operations of type theory, upon which innumerable different type theories can be specified by defining the desired term- and type-formers. We have arrived at the same point in our development of a semantics for directed type theory: the notion of NPCwF consists solely of structural components, but nothing that actually allows for the construction of interesting types and terms. So let’s rectify this by giving the directed analogue of the standard core of undirected type theory: identity types, dependent function types, and universes. We start with the directed analogue of identity types, hom-types.

▶ **Definition 3** (Directed CwF). *A **directed CwF (DCwF)** is a NPCwF equipped with the following structure:*

- *a type former*

$$\mathsf{Hom}\colon \{\Gamma\colon \mathsf{Con}\}\{A\colon \mathsf{Ty}\,\Gamma\} \to \mathsf{Tm}(\Gamma, A^-) \to \mathsf{Tm}(\Gamma, A) \to \mathsf{Ty}\,\Gamma$$

  *which is stable under substitution:*

$$\mathsf{Hom}(t, t')[\sigma] \equiv \mathsf{Hom}(t[\sigma], t'[\sigma]);$$

- *in any* $\Gamma\colon \mathsf{NeutCon}$*, a term* $\mathsf{refl}_t\colon \mathsf{Hom}(t, -t)$ *for each term* $t\colon \mathsf{Tm}(\Gamma, A^-)$*, also stable under substitution by* $\sigma\colon \mathsf{Sub}\ \Delta\ \Gamma$ *for* $\Delta\colon \mathsf{NeutCon}$*; and*
- *a term former* $\mathsf{J}$ *as given in Equation 3, also appropriately stable under substitution.*

*For any* $\Gamma\colon \mathsf{Con}$ *and* $A\colon \mathsf{NeutTy}\,\Gamma$*, write* $\mathsf{Id}(t, t')$ *for* $\mathsf{Hom}(t, t')$*.*[10]

The naming of Hom versus Id is suggestive: the types in a DCwF are supposed to function like synthetic categories (with Hom encoding their morphisms), and the neutral types are synthetic groupoids, whose homs are symmetric like an identity type. This point is best illustrated by the following claim.

▶ **Proposition 4.** *Every DCwF has an operation*

$$\begin{aligned}\mathsf{symm}\colon &\{\Gamma\colon \mathsf{NeutCon}\}\{A\colon \mathsf{NeutTy}\}\{t\colon \mathsf{Tm}(\Gamma, A^-)\}\{t'\colon \mathsf{Tm}(\Gamma, A)\}\\ &\to \mathsf{Tm}(\Gamma, \mathsf{Id}(t, t')) \to \mathsf{Tm}(\Gamma, \mathsf{Id}(-t', -t))\end{aligned}$$

**Proof.** By the following construction in the DCwF syntax:

```
symm : {Γ : NeutCon}{A : NeutTy Γ}{t : Tm(Γ, A⁻)}{t' : Tm(Γ, A)}
   → Tm(Γ, Id(t, t')) → Tm(Γ, Id(−t', −t))
symm f = (J_t,S refl_t)[ id ,+ t',+ f ]  where
      S : Ty (Γ ▷+ A)
      S = Id(−v , −t)
```

◀

This proof relies on the neutrality of $A$ in a very subtle, but critical way: in the definition of the type family $\mathsf{S}$, the variable term $\mathsf{v}\colon \mathsf{Tm}(\Gamma \rhd^+ A, A[\mathsf{p}_A])$ is negated, so that it is of type $A[\mathsf{p}_A]^-$ and therefore able to stand as the first argument to Id. But this is only possible if $\Gamma \rhd^+ A\colon \mathsf{NeutCon}$, because term-negation is only defined in neutral contexts. This reasoning will prove important for the style of reasoning we employ in Section 4, so we isolate it as a principle.

▶ **Principle** (Var Neg). For $\Gamma\colon \mathsf{NeutCon}$, the variable term $\mathsf{v}\colon \mathsf{Tm}(\Gamma \rhd^+ A, A[\mathsf{p}_A])$ can only be negated (i.e. forming $-\mathsf{v}$) if $A\colon \mathsf{NeutTy}\,\Gamma$.

In Section 5, we'll argue that there's *no* way to construct this symmetry term (for arbitrary DCwFs[11]) if $A$ is not assumed to be neutral.

Before proceeding, it's worth explaining what is "the DCwF syntax" mentioned in the proof above. This is where it becomes relevant that DCwFs are presented as generalized algebraic theories: as mentioned in the introduction, [18] proves that any GAT has an

---

[10] In general, we'll usually want to require that, if $A$ is neutral, then so are its identity types. However, in the present theory, we'll require that *all* hom-types are neutral (see below).

[11] with some nontrivial amount of structure.

initial syntax model. Therefore, any construction done in the syntax model (such as the construction of $\mathsf{symm}$ above) can be interpreted into *any* DCwF. This is why a syntactic construction was adequate to prove a claim about *all* DCwFs in the foregoing proof. In the next section, our proofs will all be syntactic, and thereby apply to arbitrary DCwFs.

Let us make an important observation about the syntax of DCwFs. An important criterion for our theory is that hom-types can be *iterated*, that is, our syntax allows for the formation of homs between homs, and homs between homs between homs, and so on. The iteration of identity types is, after all, how homotopy type theory is able to serve as a synthetic language for higher groupoids; and since hom types are iterable in the DCwF syntax, it is a synthetic language for higher categories. However, a given model may be *truncated*, in that the higher structure may become trivial after a certain point. This is the case with the groupoid model: while its types do not all obey the *uniqueness of identity proofs (UIP)* principle and are therefore not mere *h-sets*, they do obey "UIP, one level up": in the groupoid model, identity proofs of identity proofs *are* unique.

The same happens in the category model: in general, there may be terms $f$ of type $\mathsf{Hom}(t, t')^-$ and $g$ of type $\mathsf{Hom}(t, t')$ but no term of type $\mathsf{Hom}(f, g)$ – the same way the groupoid model doesn't validate the uniqueness of *identities*, the category model doesn't validate the uniqueness of *homs*. But it does trivialize "one level up": the type $\mathsf{Hom}(f, g)$ may be either inhabited or uninhabited, sure, but $\mathsf{Hom}(f, g)$ is an h-prop, a subsingleton type; it has at most one element. This is to say that the category model doesn't support synthetic *higher* categories: the synthetic category structure of $\mathsf{Hom}(t, t')$ is a synthetic preorder. But there's another sense in which the category model structure trivializes "one level up": all the hom-types are interpreted as discrete categories, which are necessarily *groupoids*. So $\mathsf{Hom}(t, t')$ is not just a synthetic preorder, it's actually a synthetic *setoid*. This is appropriate for doing synthetic 1-category theory: it makes sense that the hom-types are trivial *as categories*: to do 1-category theoretic arguments, we want our hom-sets to be hom-*sets*, that is, types whose only synthetic category-theoretic structure is propositional, symmetric identity types.

We encapsulate DCwFs like this into a definition for further study.

▶ **Definition 5.** *A* $(1, 1)$*-**truncated DCwF** is a DCwF such that*

- $\mathsf{Hom}(t, t')$ *is a neutral type for any terms* $t, t'$*; and*
- *UIP holds for identities of hom-terms:*

$$\mathsf{UIP}^1 \colon (\alpha \colon \mathsf{Tm}(\Gamma, \mathsf{Id}(p, q)^-)) \to (\beta \colon \mathsf{Tm}(\Gamma, \mathsf{Id}(p, q))) \to \mathsf{Tm}(\Gamma, \mathsf{Id}(\alpha, \beta))$$

The numbering follows the well-known indexing of $(n, m)$-categories (see e.g. [6, Defn. 8]) to refer to $\infty$-categories where all parallel $k$-morphisms are equal when $k > n$ and all $k$-morphisms are invertible for $k > m$. We could define $(n, m)$-truncated DCwFs for arbitrary $n$ and $m$ (for instance, the preorder model would be $(0, 1)$-truncated, the groupoid model $(1, 0)$-truncated, etc.), but that would take us too far afield. For the present work, we will work with $(1, 1)$-truncated DCwFs, and develop the theory of synthetic 1-category theory (i.e. synthetic $(1, 1)$-category theory) in that language. The practical consequence of working in the syntax of $(1, 1)$-truncated DCwFs is that we only have one "layer" of homs, and the type $\mathsf{Hom}(t, t')$ itself is neutral, i.e. its homs are symmetric identity types.

To conclude this section, we return to a key construct from the previous section, namely Π-types, and define what they look like in NPCwFs. This is just the appropriately-polarized analogue of [16, Defn. 3.15], and is approximately the same rule for Π-types in [22].

▶ **Definition 6.** *A PCwF supports* ***polarized*** $\Pi$***-types*** *if it comes equipped with a type former*

$$\Pi\colon (A\colon \mathsf{Ty}\,\Gamma^-) \to \mathsf{Ty}(\Gamma \rhd^- A) \to \mathsf{Ty}\,\Gamma$$

*which is stable under substitution, along with a natural isomorphism*

$$\mathit{lam}\colon \mathsf{Tm}(\Gamma \rhd^- A, B) \cong \mathsf{Tm}(\Gamma, \Pi(A, B))\colon \mathit{app}.$$

The $\beta$-law is that $\mathsf{app} \circ \mathsf{lam} \equiv \mathsf{id}$ and the $\eta$-law the other way around.

Now, the operation of $\Pi$-types provide an independent reason to work in neutral contexts: in a non-neutral context, these polarized $\Pi$-types are rather difficult to work with. For instance, consider the task of writing the identity function on some type $A\colon \mathsf{Ty}\,\Gamma$. The type $A \to A$ is not even well-formed, as the domain type of a function must be a type in context $\Gamma^-$ but the codomain in context $\Gamma$. If $\Gamma$ is neutral, we can fix this using the $\mathsf{e}$ isomorphism, taking $A[\mathsf{e}] \to A$ to be the type of endofunctions on $A$. Likewise, it's unclear how to *write* the identity function without using the tools of neutral contexts, as the variable term $\mathsf{v}$ is an element of $\mathsf{Tm}(\Gamma \rhd^+ A, A[\mathsf{p}])$, but $\mathsf{lam}$ needs an element of $\mathsf{Tm}(\Gamma \rhd^- A[\mathsf{e}], A[\mathsf{e}][\mathsf{p}_-])$. Here, the $\mathsf{ee}$ isomorphism solves the problem:

$$\mathsf{lam}(\mathsf{v}[\mathsf{ee}])\colon \mathsf{Tm}(\Gamma, A[\mathsf{e}] \to A).$$

Likewise for application: if we define the application operator by $F \mathbin{\$} t = (\mathsf{app}\ F)[\mathsf{id},_-\, t]$, it will have type

$$\_\$\_\colon \mathsf{Tm}(\Gamma, A[e] \to B) \to \mathsf{Tm}(\Gamma^-, A[e]^-) \to \mathsf{Tm}(\Gamma, B).$$

This is a bit unfortunate: a function and its argument need to come from different contexts! Using that $\Gamma\colon \mathsf{NeutCon}$, however, then we can freely substitute terms between $\Gamma$ and $\Gamma^-$ using $\mathsf{e}$, and coerce terms between $A$ and $A^-$ using the term-negation operator, alleviating this difficulty and allowing us to, for instance, construct the composition of functions.

```
_∘_ : {Γ : NeutCon}{A B C : Ty Γ} →
  Tm(Γ, B[e] → C) → Tm(Γ, A[e] → B) → Tm(Γ, A[e] → C)
G ∘ F = lam( (app G)[ ee⁻¹ ][ p₋,A ,+ app F ] )
```

There are several other ways which neutral contexts prove necessary – see [25, Chap. 2] for a more thorough exploration. We view the neutral-context method as essential to making the type theory of the category model viable: they provide not just an alternative solution to the issue of $\mathsf{refl}$'s divariance, but, as we can see here with $\Pi$-types, solve numerous problems caused by the polarity calculus in arbitrary contexts being too strict. Working in a neutral context in the syntax of (1,1)-directed type theory with $\Pi$-types, we're able to actually develop synthetic category theory; we turn our attention to that task now.

## 4 Synthetic Category Theory

In this section, we work in an arbitrary $(1,1)$-truncated DCwF with polarized $\Pi$-types by only working in the syntax. In the main body of the text, we'll adopt an informal type theoretic style (inspired by [34]). We assume that we're working in some neutral context $\Gamma$, though we don't explicitly reference $\Gamma$. We'll write $t\colon A$ to indicate $t\colon \mathsf{Tm}(\Gamma, A)$. In what follows, we'll use the letters $p, q, r, s, t, u, v, w, f, g$ to name terms (of various types) in $\Gamma$, whereas the letters $x, y, z$ will be the names of *variables* obtained by extending $\Gamma$. We'll have to be careful to abide by the variable negation rule:

▶ **Principle** (Var Neg)**.** An expression $e$ can only be negated if all the variables occurring in it are of neutral types.

We suppress the distinction between $\Gamma$ and $\Gamma^-$, since we can substitute back and forth with $\mathsf{e}$ behind the scenes, as needed. Negative context extension will just behave like positive extension by a negative type: recall that $\mathsf{v}_{-,A} \colon \mathsf{Tm}((\Gamma \triangleright^- A)^-, A[\mathsf{p}^-_{-,A}]^-)$ i.e.

$$\mathsf{v}_{-,A} \colon \mathsf{Tm}(\Gamma^- \triangleright^+ A^-, A[\mathsf{p}_{A^-}]^-)$$

so, if we're suppressing the distinction between $\Gamma$ and $\Gamma^-$, then this is just a variable $x$ of $A^-$. Accordingly, we'll apply and form functions like this:

$$\frac{f \colon \prod_{(x \colon A^-)} B(x) \quad t \colon A^-}{f(t) \colon B(t)} \qquad \frac{x \colon A^- \vdash e \colon B(x)}{(\lambda(x \colon A^-) \to e) \colon \prod_{x \colon A^-} B(x).}$$

Finally, here's our principle of directed path induction:

▶ **Principle** (Directed Path Induction)**.** For every $t \colon A^-$, if $M(x, y)$ is a type family depending on $x \colon A$ and $y \colon \mathsf{Hom}(t, x)$, then, for each $m \colon M(-t, \mathsf{refl}_t)$, we get an $\mathsf{ind}_M(m, x, y) \colon M(x, y)$ for all $x, y$.

So, for instance, the construction of symmetry above (the proof of Proposition 4) would be expressed informally as follows: given a neutral type $A$ and a term $t \colon A^-$, define a type family over $x \colon A, y \colon \mathsf{Id}(t, x)$ by $S(x, y) = \mathsf{Id}(-x, -t)$. We have not violated (Var Neg) because $x \colon A$ and $A$ is neutral. We have a term of type $S(-t, \mathsf{refl}_t)$, i.e. $\mathsf{Id}(t, -t)$, namely $\mathsf{refl}_t$. So therefore we get $S(x, y)$ for arbitrary $x, y$. If we have a particular $t' \colon A$ and $p \colon \mathsf{Id}(t, t')$, we can put $\mathsf{symm}\ p = \mathsf{ind}_S(\mathsf{refl}_t, t', p)$. Again, we emphasize that it is (Var Neg) which prevents this argument from working for non-neutral types, as desired. Below, we are more casual with our application of directed path induction (e.g. not defining the type family explicitly) in cases where (Var Neg) is not a concern.

With that, we can proceed to the informal constructions. Along the way, the explicit constructions in the DCwF syntax are carried out in the accompanying figures. For the full details, see the calculations in [26].

## 4.1 Composition of Homs (Figure 7)

As mentioned, a type $A$ in directed type theory is supposed to be a *synthetic category*. The terms $t : A$ represent objects, and the terms $p \colon \mathsf{Hom}(t, t')$ represent morphisms. For this to truly be category theory, however, we must be able to compose morphisms. We'll write composition in diagrammatic order: given $t, u \colon A^-$ and $v' \colon A$, we should be able to compose $p \colon \mathsf{Hom}(t, -u)$ with $q \colon \mathsf{Hom}(u, v')$ to get $p \cdot q \colon \mathsf{Hom}(t, v')$. We do this by directed path induction on $q$, by putting $p \cdot \mathsf{refl}_u = p$.

The $\mathsf{refl}$ terms serve as the identity morphisms of the category: by the above, we know that $p \cdot \mathsf{refl}_u \equiv p$, and thus $\mathsf{refl}_p \colon \mathsf{Id}(p \cdot \mathsf{refl}_u, p)$. As for the other unit law, we must again use directed path induction: since $\mathsf{refl}_u \cdot \mathsf{refl}_u \equiv \mathsf{refl}_u$, we have that $\mathsf{refl}_{\mathsf{refl}_u} \colon \mathsf{Id}(\mathsf{refl}_u \cdot \mathsf{refl}_u, \mathsf{refl}_u)$, and, by induction we get for each $q \colon \mathsf{Hom}(u, v')$ a term

$$\mathsf{r{-}unit}\ q = \mathsf{ind}(\mathsf{refl}_{\mathsf{refl}_u}, v, q) \colon \mathsf{Id}(\mathsf{refl}_u \cdot q, q).$$

Finally, we get that the composition operation is associative. Given $t, u, v \colon A^-$ and $w' \colon A$ as well as $p \colon \mathsf{Hom}(t, -u)$, $q \colon \mathsf{Hom}(u, -v)$, and $r \colon \mathsf{Hom}(v, w')$, we construct

$$\mathsf{assoc}\ p\ q\ r \colon \mathsf{Id}(p \cdot (q \cdot r), (p \cdot q) \cdot r)$$

by directed path induction on $r$. If $r = \mathsf{refl}_v$, then $q \cdot r \equiv q$ and $(p \cdot q) \cdot r \equiv p \cdot q$. Thus, we have $\mathsf{refl}_{p \cdot q} \colon \mathsf{Id}(p \cdot (q \cdot \mathsf{refl}_v), (p \cdot q) \cdot \mathsf{refl}_v)$, and then the induction carries through, and we get $\mathsf{assoc}\ p\ q\ r$ as desired.

```
C : {t : Tm(Γ, A⁻)} → Ty (Γ ▷⁺ A ▷⁺ Hom(t'[p_A], v))
C = Hom(t[p_A],v_A)[ p ]

_·_ : {t t' : Tm(Γ, A⁻)}{t'' : Tm(Γ, A)} → Tm(Γ, Hom(t,−t')) → Tm(Γ, Hom(t',t''))
   → Tm(Γ, Hom(t,t''))
f · g = (J_t',C f) [ id,+ t'' ,+ g ]
```

**Figure 7** Composition of Homs.

## 4.2 Synthetic Functors (Figure 8)

If types $A, B$ are synthetic categories, it should come as no surprise that terms $F \colon A \to B$ are synthetic *functors*. The object part is given by the usual function application, but the variances are somewhat mixed: if $t \colon A^-$, then we can say $f(t) \colon B$. However, we can still apply $f$ to a term $t' \colon A$, we just have to put a minus on $t'$, i.e. $f(-t')$.

Unlike usual ("analytic") category theory, we don't have to explicitly define the morphism part of a functor; any term of type $A \to B$ we can write down will come with a morphism part for free. To obtain this morphism part, again we use directed path induction: given an $F \colon A \to B$ and some $t \colon A^-$, we can define a $B$-morphism

$$\mathsf{map}\ F\ f \colon \mathsf{Hom}(-F(t), F(-t'))$$

for every $t' \colon A$ and $f \colon \mathsf{Hom}(t, t')$ by defining $\mathsf{map}\ F\ \mathsf{refl}_t$ to be $\mathsf{refl}_{-F(t)}$. By definition $(J\beta)$, this operation preserves identities (sending $\mathsf{refl}$ to $\mathsf{refl}$), and respects composition: if we have $t, u \colon A^-$ and $f \colon \mathsf{Hom}(t, -u)$, then, since $\mathsf{map}\ F\ \mathsf{refl}_u \equiv \mathsf{refl}_{-F(u)}$ and $f \cdot \mathsf{refl}_u \equiv f$ and $(\mathsf{map}\ F\ f) \cdot \mathsf{refl}_{-F(u)} \equiv \mathsf{map}\ F\ f$, we have

$$\mathsf{refl}_{(\mathsf{map}\ F\ f)} \colon \mathsf{Id}(\mathsf{map}\ F\ (f \cdot \mathsf{refl}_u), (\mathsf{map}\ F\ f) \cdot (\mathsf{map}\ F\ \mathsf{refl}_t)).$$

By induction, we get an identity between $\mathsf{map}\ F\ (f \cdot g)$ and $(\mathsf{map}\ F\ f) \cdot (\mathsf{map}\ F\ g)$ for arbitrary $g$.

```
_$_ : (Tm(Γ, A[e] → B)) → Tm(Γ,A⁻) → Tm(Γ,B)
F $ t = (app F)[ id ,− t[e] ]

map : {Γ : NeutCon}{A}{B}(f : Tm(Γ, A[e] → B)){t : Tm(Γ,A⁻)}{t' : Tm(Γ,A)}
   → (Tm(Γ, Hom(t,t'))) → Tm(Γ, Hom( −(F $ t) , F $ (−t')))
map F f = (J_t,MAP refl_−(F $ t))[id ,+ t' ,+ f ]     where
      MAP : Ty (Γ ▷⁺ A)
      MAP = Hom( (−(F $ t))[p_A], (app F)[ ee⁻¹ ] ) −− See Equation 6
```

**Figure 8** Morphism part of Functors.

## 5 Further observations about the Category Model

As the previous section showed, the syntax of 1-truncated Directed CwFs provides a nice setting for some very basic constructions in synthetic category theory. However, further expansion of the DCwF syntax is needed to be able to capture the full range of constructions in category theory. In this section, we'll observe some constructions that can be made (and some equivalences that hold) in the category model, which require further study to be internalized into the DCwF syntax.

Probably the most significant omission from the synthetic category theory of the previous section is natural transformations. As we discuss in [26, Sect. 5], there is some work to be done to appropriately accommodate natural transformations in our framework. But for the present work, we'll focus on another important feature: *universes*. The category model comes equipped with several type universes, most significantly the universe of sets. More precisely, we can regard the category $\mathsf{Set}$ as a type in each context $\Gamma$ of the category model, interpreted by the category of sets and functions. The operation $\mathsf{El}\colon \mathsf{Tm}(\Gamma, \mathsf{Set}) \to \mathsf{NeutTy}\,\Gamma$ takes a set $X$ and views it as a discrete category. We can then define

$$\mathsf{Hom{-}to{-}func}\colon \mathsf{Tm}(\Gamma, \mathsf{Hom}(-X, Y)) \to \mathsf{Tm}(\bullet, \mathsf{El}(X)[\mathsf{e}] \to \mathsf{El}(Y))$$

by directed path induction: $\mathsf{Hom{-}to{-}func}\ \mathsf{refl}_{-X}$ is the identity function, $\mathsf{lam}\ (\mathsf{v}[\mathsf{ee}])$ of type $\mathsf{Tm}(\Gamma, \mathsf{El}(X)[\mathsf{e}] \to \mathsf{El}(X))$. We can use this to state the following principle.

▶ **Principle** (Directed Univalence)**.** $\mathsf{Hom{-}to{-}func}$ is a bijection.

Spelling out the category model semantics, we see that every function is sent to itself. Sufficiently internalized, this principle of Directed Univalence serves as the directed analogue of Hofmann and Streicher's *universe extensionality* axiom [17, Section 5.4]. Further work is required to better develop the theory of isomorphisms in the synthetic category theory, and to compare this principle of directed univalence to existing ones (e.g. [21, 14]).

Let us conclude by observing that the existence of a universe allows for metatheoretic reasoning as well, specifically negative proofs about what *cannot* be done in the syntax. We can view the directed universe $\mathsf{Set}$ as a source of nontrivial directedness: if we affirm Directed Univalence, then $\mathsf{Set}$ cannot possibly be a neutral type. We show that, in DCwFs equipped with a directed univalent universe $\mathsf{Set}$, hom-types must be asymmetric in general. That is, we cannot construct a term $\mathsf{symm}$ like in Proposition 4 for non-neutral types in the syntax DCwF+$\mathsf{Set}$ (the initial model of the GAT of DCwFs with a universe $\mathsf{Set}$). We do so the same way Hofmann and Streicher [17] proved that ordinary Martin-Löf Type Theory couldn't prove the Uniqueness of Identity Proofs: by countermodel. Hofmann and Streicher's countermodel was the groupoid model, and, of course, ours is the category model.

▶ **Proposition 7.** *There cannot be an operation*

$$\mathsf{symm}'\colon \{\Gamma\}\{A\}\{t\colon \mathsf{Tm}(\Gamma, A^-)\}\{t'\colon \mathsf{Tm}(\Gamma, A)\} \to \mathsf{Tm}(\Gamma, \mathsf{Hom}(t, t')) \to \mathsf{Tm}(\Gamma, \mathsf{Hom}(-t', -t))$$

*definable in the syntax of DCwFs+*$\mathsf{Set}$*.*

**Proof.** If the syntax model of DCwF+$\mathsf{Set}$ had such an operation $\mathsf{symm}'$, then, by initiality, so too would every DCwF with $\mathsf{Set}$, in particular the category model. But then for any $X\colon \mathsf{Tm}(\bullet, \mathsf{Set}^-)$ and $Y\colon \mathsf{Tm}(\bullet, \mathsf{Set})$ and $f\colon \mathsf{Tm}(\bullet, \mathsf{Hom}(X, Y))$, we would obtain $\mathsf{symm}'\ f\colon \mathsf{Tm}(\bullet, \mathsf{Hom}(Y, X))$. But this is absurd, because the function $?\colon \emptyset \to \mathbb{1}$ is a term of type $\mathsf{El}(\emptyset) \to \mathsf{El}(\mathbb{1})$ in the category model, and, by Directed Univalence, corresponds to a term of type $\mathsf{Hom}(\emptyset, \mathbb{1})$, but there cannot be any terms of $\mathsf{Hom}(\mathbb{1}, \emptyset)$, because the set of terms of this type is in bijection with the set of functors $\mathsf{El}(\mathbb{1})$ to $\mathsf{El}(\emptyset)$, of which there are none. ◀

Basically the same argument will show the *uniqueness of homs* principle – that for any hom terms $p\colon \mathsf{Tm}(\Gamma, \mathsf{Hom}(t,t')^-)$ and $q\colon \mathsf{Tm}(\Gamma, \mathsf{Hom}(t,t'))$, there is a witness of $\mathsf{Id}(p,q)$ – is violated in the category model (a counterexample being Set-homs from the two-element set to itself), and therefore not provable in the syntax of DCwFs+Set. Indeed, in the same way that Hofmann and Streicher remark that UIP is contradicted by their universe extensionality, our Directed Univalence is incompatible with the uniqueness of homs. So, in sum, we can conclude that the difference between $(1,1)$-truncated DCwFs, $(1,0)$-truncated DCwFs, and $(0,1)$-truncated CwFs is reflected internally in the syntax.

## 6 Conclusion and Future Work

We have laid the foundation for a generalized algebraic theory of directed types, and began to conduct synthetic category theory in that setting. Our semantics-forward approach was to study the category model first, and extract its key features into a series of abstract definitions – the GATs of polarized CwFs, neutral-polar CwFs, directed CwFs, $(1,1)$-truncated directed CwFs, and directed CwFs with features like polarized dependent types and a directed univalent set universe. Working within the directed type theory of these models, we found that it was possible to work informally with the powerful directed path induction principle to make basic constructions in category theory, with our careful discipline about variable negation preventing the directed type theory from collapsing into undirected type theory.

The reviewers have claimed that our system isn't closed under substitution: the negation operation is only applicable in a neutral context, but there may be a substitution from a context that is not neutral and then reducing the substitution would result in a non-welltyped term. However, since we formulated our theory as a GAT, substitution is an explicit operation and hence this isn't an issue for our presentation. Indeed the naturality equation $(-t)[\sigma] = -(t[\sigma])$ only applies if $\sigma$ is a substitution between neutral contexts. This does entail for an implementation that applying substitutions to negated terms has to be delayed, i.e. we need to deal with explicit closures of the form $(-t)[\sigma]$, if the domain of $\sigma$ is not a neutral context.

Much remains to be done.[12] The category theory of Section 4 serves as a proof-of-concept, but needs to be fleshed out into a full theory. As mentioned, work is needed to articulate natural transformations in the theory; our current investigations concern possible generalization of Π-types to di-variant *end types* which address some of the above-mentioned variance issues with natural transformations. For reasons of space, we omitted dependent sum types from the theory. But with them added, much of basic category theory should be expressible in this language, such as isomorphisms, (co)slice categories, (co)limits, exponentials, and adjunctions.[13] Better development of the category of sets should put representability and some properties of presheaf categories into reach, though internal statement and proof of the Yoneda Lemma will likely rely on the resolution of the above-mentioned dilemma regarding natural transformations. We also leave it to future work to study whether this theory can capture higher category theory by weakening or dropping the assumption of (1,1)-truncation, and, if so, how it compares to existing synthetic higher category frameworks, such as [30].

[12] Some details which are omitted here are included in the extended arxiv version of this paper, [26]. A broader and more thorough treatment of this material is given in the first author's forthcoming PhD thesis [25]. There, some of the issues listed here are addressed.

[13] See [25, Chap. 4].

There are further avenues for developing the type theory of DCwFs. Two important metatheoretic results about the syntax of DCwFs currently being pursued are *canonicity* and *normalization* (modulo the concern above about closure under substitution). Moreover, we would like to verify the correctness of these results by formalizing them in a computer proof assistant. A further goal would be to implement the syntax of DCwFs as a computer proof language itself, hopefully with syntax nearly as convenient as the constructions of Section 4, and formalize larger swaths of category theory in it.

A further motivation for the present work's focus on generalized algebraic theories is the possibility of expressing this style of directed types theory in a *second-order* generalized algebraic theory, following [8, 33, 32, 4, 19]. Since our notions of PCwFs and NPCwFs include explicit operations on contexts (as seen in the substructural character of the (Var Neg) rule), it's clear that either an extension to the SOGAT signature language of [19], and/or a partial internalization of the first-order theory into the second-order theory – à la [4] – will be necessary. Another task is to compare our $(\_)^-$ modality (and the $(\_)^0$ modality definable in the category model using core groupoids) to the modalities studied in [13], particularly the kinds employed in directed type theory [37]. We indicated above that the core modality ought to be a dependent right adjoint in the appropriate sense, but work remains to spell out this explicitly. Another possible avenue: this work is conducted with a possible future connection to *higher observational type theory* [31, 5, 4] in mind; to begin to prepare the way for such a project, at the very least further study of the observational equivalences of this theory (e.g. a characterization of the $\mathsf{Hom}$-types of $\Pi$-types) is needed.

Finally, the present framework provides a setting for studying *directed higher-inductive types* – inductively-defined types with both term constructors and *hom constructors*. Some simple examples would be the directed interval (as is studied in [36, 30]) and a directed analogue of the *circle* type [34, Section 6.2]. These examples are modelled by the category model, and can therefore be soundly added to the present theory. Higher examples (such as directed versions of higher tori and spheres) would require more careful metatheoretic work to justify, but could perhaps lead to a number of interesting considerations.

## References


1 Benedikt Ahrens, Paige Randall North, and Niels van der Weide. Bicategorical type theory: semantics and syntax. *Mathematical Structures in Computer Science*, 33(10), 2023. `doi:10.1017/s0960129523000312`.

2 Thorsten Altenkirch. Extensional equality in intensional type theory. In *Proceedings. 14th Symposium on Logic in Computer Science (Cat. No. PR00158)*, pages 412–420. IEEE, 1999. `doi:10.1109/LICS.1999.782636`.

3 Thorsten Altenkirch, Simon Boulier, Ambrus Kaposi, Christian Sattler, and Filippo Sestini. Constructing a universe for the setoid model. In *FoSSaCS*, pages 1–21, 2021. `doi:10.1007/978-3-030-71995-1_1`.

4 Thorsten Altenkirch, Yorgo Chamoun, Ambrus Kaposi, and Michael Shulman. Internal parametricity, without an interval. *Proceedings of the ACM on Programming Languages*, 8(POPL):2340–2369, 2024. `doi:10.1145/3632920`.

5 Thorsten Altenkirch, Ambrus Kaposi, and Michael Shulman. Towards higher observational type theory. In *28th International Conference on Types for Proofs and Programs (TYPES 2022). University of Nantes*, 2022.

6 John C. Baez and Michael Shulman. Lectures on n-categories and cohomology, 2007. `arXiv:math/0608420`.

7 Lars Birkedal, Ranald Clouston, Bassel Mannaa, Rasmus Ejlers Møgelberg, Andrew M Pitts, and Bas Spitters. Modal dependent type theory and dependent right adjoints. *Mathematical Structures in Computer Science*, 30(2):118–138, 2020. doi:10.1017/S0960129519000197.

8 Rafaël Bocquet. External univalence for second-order generalized algebraic theories. *CoRR*, abs/2211.07487, 2022. doi:10.48550/arXiv.2211.07487.

9 John Cartmell. Generalised algebraic theories and contextual categories. *Annals of pure and applied logic*, 32:209–243, 1986. doi:10.1016/0168-0072(86)90053-9.

10 Simon Castellan, Pierre Clairambault, and Peter Dybjer. Categories with families: Unityped, simply typed, and dependently typed. *CoRR*, abs/1904.00827:135–180, 2019. doi:10.48550/arXiv.1904.00827.

11 Fernando Chu, Éléonore Mangel, and Paige Randall North. A directed homotopy type theory for 1-categories. In *30th International Conference on Types for Proofs and Programs (TYPES 2024). IT University of Copenhagen*, 2024.

12 Peter Dybjer. Internal type theory. In *International Workshop on Types for Proofs and Programs*, pages 120–134. Springer, 1995. doi:10.1007/3-540-61780-9_66.

13 Daniel Gratzer, G. A. Kavvos, Andreas Nuyts, and Lars Birkedal. Multimodal dependent type theory. In Holger Hermanns, Lijun Zhang, Naoki Kobayashi, and Dale Miller, editors, *LICS '20: 35th Annual ACM/IEEE Symposium on Logic in Computer Science, Saarbrücken, Germany, July 8-11, 2020*, pages 492–506. ACM, 2020. doi:10.1145/3373718.3394736.

14 Daniel Gratzer, Jonathan Weinberger, and Ulrik Buchholtz. The yoneda embedding in simplicial type theory, 2025. doi:10.48550/arXiv.2501.13229.

15 Martin Hofmann. A simple model for quotient types. In *International Conference on Typed Lambda Calculi and Applications*, pages 216–234. Springer, 1995. doi:10.1007/BFb0014055.

16 Martin Hofmann. Syntax and semantics of dependent types. In *Extensional Constructs in Intensional Type Theory*, pages 13–54. Springer, 1997.

17 Martin Hofmann and Thomas Streicher. The groupoid interpretation of type theory. *Twenty-five years of constructive type theory (Venice, 1995)*, 36:83–111, 1995.

18 Ambrus Kaposi, András Kovács, and Thorsten Altenkirch. Constructing quotient inductive-inductive types. *Proc. ACM Program. Lang.*, 3(POPL), January 2019. doi:10.1145/3290315.

19 Ambrus Kaposi and Szumi Xie. Second-order generalised algebraic theories: Signatures and first-order semantics. In Jakob Rehof, editor, *9th International Conference on Formal Structures for Computation and Deduction, FSCD 2024, July 10-13, 2024, Tallinn, Estonia*, volume 299 of *LIPIcs*, pages 10:1–10:24. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi:10.4230/LIPIcs.FSCD.2024.10.

20 András Kovács. *Type-Theoretic Signatures for Algebraic Theories and Inductive Types.* PhD thesis, Eötvös Loránd University, 2022.

21 Nikolai Kudasov, Emily Riehl, and Jonathan Weinberger. Formalizing the $\infty$-categorical yoneda lemma, 2023. doi:10.48550/arXiv.2309.08340.

22 Daniel R. Licata and Robert Harper. 2-dimensional directed type theory. In Michael W. Mislove and Joël Ouaknine, editors, *Twenty-seventh Conference on the Mathematical Foundations of Programming Semantics, MFPS 2011, Pittsburgh, PA, USA, May 25-28, 2011*, volume 276 of *Electronic Notes in Theoretical Computer Science*, pages 263–289. Elsevier, 2011. doi:10.1016/j.entcs.2011.09.026.

23 Per Martin-Löf. An intuitionistic theory of types: Predicative part. In H.E. Rose and J.C. Shepherdson, editors, *Logic Colloquium '73*, volume 80 of *Studies in Logic and the Foundations of Mathematics*, pages 73–118. Elsevier, 1975.

24 Per Martin-Löf. Constructive mathematics and computer programming. In L. Jonathan Cohen, Jerzy Łoś, Helmut Pfeiffer, and Klaus-Peter Podewski, editors, *Logic, Methodology and Philosophy of Science VI*, volume 104 of *Studies in Logic and the Foundations of Mathematics*, pages 153–175. Elsevier, 1982.

25 Jacob Neumann. *A Generalized Algebraic Theory of Directed Equality.* PhD thesis, University of Nottingham, 2025.

**26** Jacob Neumann and Thorsten Altenkirch. Synthetic 1-categories in directed type theory. *CoRR*, abs/2410.19520, 2024. doi:10.48550/arXiv.2410.19520.

**27** Max S. New and Daniel R. Licata. A formal logic for formal category theory. In Orna Kupferman and Pawel Sobocinski, editors, *Foundations of Software Science and Computation Structures - 26th International Conference, FoSSaCS 2023, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2023, Paris, France, April 22-27, 2023, Proceedings*, volume 13992 of *Lecture Notes in Computer Science*, pages 113–134. Springer, 2023. doi:10.1007/978-3-031-30829-1_6.

**28** Paige Randall North. Towards a directed homotopy type theory. *Electronic Notes in Theoretical Computer Science*, 347:223–239, 2019. doi:10.1016/j.entcs.2019.09.012.

**29** Andreas Nuyts. Towards a directed homotopy type theory based on 4 kinds of variance. *Mém. de mast. Katholieke Universiteit Leuven*, 2015.

**30** E. Riehl and M. Shulman. A type theory for synthetic $\infty$-categories. *Higher Structures*, 1(1):116–193, 2017. doi:10.2140/hs.2017.1.116.

**31** Michael Shulman. Towards a third-generation HOTT, 2022. Carnegie Mellon University HoTT Seminar. URL: https://www.cmu.edu/dietrich/philosophy/hott/seminars/index.html.

**32** Taichi Uemura. *Abstract and concrete type theories*. PhD thesis, University of Amsterdam, 2021.

**33** Taichi Uemura. A general framework for the semantics of type theory. *Mathematical Structures in Computer Science*, 33(3), March 2023. doi:10.1017/s0960129523000208.

**34** The Univalent Foundations Program. *Homotopy Type Theory: Univalent Foundations of Mathematics*. Institute for Advanced Study, Institute for Advanced Study, 2013. URL: https://homotopytypetheory.org/book.

**35** Benno Van Den Berg and Richard Garner. Types are weak $\omega$-groupoids. *Proceedings of the london mathematical society*, 102(2):370–394, 2011.

**36** Matthew Z. Weaver and Daniel R. Licata. A constructive model of directed univalence in bicubical sets. In Holger Hermanns, Lijun Zhang, Naoki Kobayashi, and Dale Miller, editors, *LICS '20: 35th Annual ACM/IEEE Symposium on Logic in Computer Science, Saarbrücken, Germany, July 8-11, 2020*, LICS '20, pages 915–928, New York, NY, USA, 2020. ACM. doi:10.1145/3373718.3394794.

**37** Jonathan Weinberger and Ulrik Buchholtz. Type-theoretic modalities for synthetic $(\infty, 1)$-categories, 2019. International Conference on Homotopy Type Theory (HoTT 2019).