A General Framework for Relational Parametricity

Kristina Sojakova
Appalachian State University
Boone, NC, USA
sojakovak@appstate.edu

Patricia Johann
Appalachian State University
Boone, NC, USA
johannp@appstate.edu

Abstract

Reynolds' original theory of *relational parametricity* was intended to capture the observation that polymorphically typed System F programs preserve all relations between inputs. But as Reynolds himself later showed, his theory can only be formulated in a meta-theory with an impredicative universe, such as the Calculus of Inductive Constructions. A number of more abstract treatments of relational parametricity have since appeared; however, as we show, none of these seem to express Reynolds' original theory in a satisfactory way.

To correct this, we develop an abstract framework for relational parametricity that delivers a model expressing Reynolds' theory in a direct and natural way. This framework is uniform with respect to a choice of meta-theory, which allows us to obtain the well-known PER model of Longo and Moggi as a direct instance in a natural way as well. Underlying the framework is the novel notion of a *λ2-fibration with isomorphisms*, which relaxes certain strictness requirements on split λ2-fibrations. Our main theorem is a generalization of Seely's classical construction of sound models of System F from split λ2-fibrations: we prove that the canonical interpretation of System F induced by every λ2-fibration with isomorphisms validates System F's entire equational theory on the nose, independently of the parameterizing meta-theory.

Moreover, we offer a novel relationally parametric model of System F (after Orsanigo), which is *proof-relevant* in the sense that witnesses of relatedness are themselves suitably related. We show that, unlike previous frameworks for parametricity, ours recognizes this new model in a natural way. Our framework is thus *descriptive*, in that it accounts for well-known models, as well as *prescriptive*, in that it identifies abstract properties that good models of relational parametricity should satisfy and suggests new constructions of such models.

CCS Concepts • **Theory of computation → Type theory**;

Keywords System F, categorical semantics, parametricity

ACM Reference format:
Kristina Sojakova and Patricia Johann. 2018. A General Framework for Relational Parametricity. In *Proceedings of LICS '18: 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, Oxford, United Kingdom, July 9–12, 2018 (LICS '18)*, 10 pages.
https://doi.org/10.1145/3209108.3209141

LICS '18, July 9–12, 2018, Oxford, United Kingdom

ACM ISBN 978-1-4503-5583-4/18/07…$15.00
https://doi.org/10.1145/3209108.3209141

1 Introduction

Reynolds [12] introduced the notion of *relational parametricity* to model the extensional behavior of programs in System F [6], the formal calculus at the core of all polymorphic functional languages. His goal was to give a type $\alpha \vdash T(\alpha)$ an *object interpretation* T_0 and a *relational interpretation* T_1, where T_0 takes sets to sets and T_1 takes relations $R \subseteq A \times B$ to relations $T_1(R) \subseteq T_0(A) \times T_0(B)$. A term $\alpha; x : S(\alpha) \vdash t(\alpha, x) : T(\alpha)$ was to be interpreted as a map t_0 associating to each set A a function $t_0(A) : S_0(A) \to T_0(A)$. The interpretations were to be given inductively on the structure of T and t in such a way that they implied two key theorems: the *Identity Extension Lemma*, stating that if R is the equality relation on A then $T_1(R)$ is the equality relation on $T_0(A)$; and the *Abstraction Theorem*, stating that, for any relation $R \subseteq A \times B$, $t_0(A)$ and $t_0(B)$ map arguments related by $S_1(R)$ to results related by $T_1(R)$. A similar result holds for types and terms with any number of free variables.

In Reynolds' treatment of relational parametricity, if $U(\alpha)$ is the type $\alpha \vdash S(\alpha) \to T(\alpha)$, for example, then $U_0(A)$ is the set of functions $f : S_0(A) \to T_0(A)$ and, for $R \subseteq A \times B$, $U_1(R)$ relates $f : S_0(A) \to T_0(A)$ to $g : S_0(B) \to T_0(B)$ iff f and g map arguments related by $S_r(R)$ to results related by $T_1(R)$. Similarly, if V is the type $\cdot \vdash \forall\alpha.S(\alpha)$, then V_0 consists of those polymorphic functions f that take a set A and return an element of $S_0(A)$, and also have the property that for any relation $R \subseteq A \times B$, $f(A)$ and $f(B)$ are related by $S_1(R)$. Two such polymorphic functions f and g are then related by V_1 iff for any relation $R \subseteq A \times B$, $f(A)$ and $g(B)$ are related by $S_1(R)$. These definitions allow us to deduce interesting properties of (interpretations of) terms solely from their types. For example, for any term $t : \forall\alpha.\alpha \to \alpha$, the Abstraction Theorem guarantees that the interpretation t_0 of t is related to itself by the relational interpretation of $\forall\alpha.\alpha \to \alpha$. So if we fix a set A, fix $a \in A$, and define a relation on A by $R \coloneqq \{(a, a)\}$, then $t_0(A)$ must be related to itself by the relational interpretation of $\alpha \vdash \alpha \to \alpha$ applied to R. This means that $t_0(A)$ must carry arguments related by R to results related by R. Since a is related to itself by R, $t_0(A)\,a$ must be related to itself by R, so that $t_0(A)\,a$ must be a. That is, t_0 must be the polymorphic identity function. Such applications of relational parametricity are useful in many different scenarios, *e.g.*, when proving invariance of polymorphic functions under changes of data representation, equivalences of programs, and "free theorems" [16].

The well-known problem with Reynolds' treatment of relational parametricity (see [13]) is that the universe of sets is not impredicative, and hence the aforementioned "set" V_0 cannot be formed. This issue can be resolved if we instead work in a meta-theory that has an impredicative universe; a natural choice is an extensional version of the Calculus of Inductive Constructions (CIC), *i.e.*, a dependent type theory with a cumulative Russell-style hierarchy of universes $\mathbb{U}_0 : \mathbb{U}_1 : \ldots$, where $\mathbb{U}_0$ is impredicative, and extensional identity types. Since there are no set-theoretic models of System F, from now on we will consider Reynolds' definitions as internal to the type theory we just described.

After Reynolds' original paper, more abstract treatments of his ideas were given by, *e.g.*, Robinson and Rosolini [14], O'Hearn and Tennent [10], Dunphy and Reddy [2], and Ghani *et al.* [5]. The approach is to use a categorical structure — reflexive graph categories for [2, 10, 14] and fibrations for [5] — to represent sets and relations, and to interpret types as appropriate functors and terms as natural transformations. In particular, [2] aims to "[address] parametricity in all its incarnations", and similarly for [5]. Surprisingly and significantly, though, Reynolds' notion of parametricity does not appear to be directly addressed by either of these frameworks.

This is because even though sets and relations are easily organized into the required categorical structure, the interpretation of type constructors as suggested by Reynolds does not necessarily preserve this categorical structure *on the nose*, only up to an isomorphism. For example, let $\alpha \vdash S(\alpha)$ and $\alpha \vdash T(\alpha)$ be two types, with object interpretations S_0 and T_0 and relational interpretations S_1 and T_1. The interpretation of the product $\alpha \vdash S(\alpha) \times T(\alpha)$ should be an appropriate product of interpretations; that is, the object interpretation should map a set A to $S_0(A) \times T_0(A)$ and the relational interpretation should map a relation R to $S_1(R) \times T_1(R)$, with the product of two relations defined in the obvious way. For the Identity Extension Lemma to hold, we need $S_1(\mathsf{Eq}(A)) \times T_1(\mathsf{Eq}(A))$ to be the same as $\mathsf{Eq}(S_0(A) \times T_0(A))$. Here, the equality relation $\mathsf{Eq}(A)$ on a set A maps $(a, b) : A \times A$ to the type $\mathsf{Id}(a, b)$ of proofs of equality between a and b, so that a and b are related iff $\mathsf{Id}(a, b)$ is inhabited, *i.e.*, iff a is identical to b. By the induction hypothesis, $S_1(\mathsf{Eq}(A))$ is $\mathsf{Eq}(S_0(A))$, and similarly for T, so we need to show that $\mathsf{Eq}(S_0(A)) \times \mathsf{Eq}(T_0(A))$ is $\mathsf{Eq}(S_0(A) \times T_0(A))$. But this is not necessarily the case since the identity type on a product is in general not *identical* to the product of identity types, but rather just suitably *isomorphic*. So the interpretation of $\alpha \vdash S(\alpha) \times T(\alpha)$ is not necessarily an indexed or fibered functor (in the settings of [2] and [5], respectively).

Three ways to fix this problem come to mind. Firstly, we can attempt to change the meta-theory, by, *e.g.*, imposing an additional axiom asserting that two logically equivalent propositions are definitionally equal. We do not pursue this approach here since we prefer to stay more general rather than rely on *ad hoc* axioms that make the shoe fit in this particular case (and that may be insufficient for a more involved example). The second possibility is to use (a syntactic analogue of) strictification, pursued in, *e.g.*, [1]. The idea is that instead of interpreting a closed type as a set A (on the object level), we interpret it as a set A endowed with a relation R_A that is *isomorphic*, but not necessarily identical, to the canonical discrete relation Eq_A. The chosen equality relation on the set A — more precisely, on the entire structure $(A, (R_A, i : R_A \simeq \mathsf{Eq}_A))$ — will then be R_A rather than Eq_A. This allows us to construct R_A in a way that respects all type constructors on the nose, so that the aforementioned issue with $\mathsf{Eq}(S_0(A)) \times \mathsf{Eq}(T_0(A))$ not being identical to $\mathsf{Eq}(S_0(A) \times T_0(A))$ is avoided. The problem here, however, is that there can be many different ways to endow A with a discrete relation (R_A, i); in other words, the type of discrete relations on A is not contractible, and the type of sets endowed with a discrete relation is not canonically equivalent to the intended type of sets.

Here we suggest a third approach: we record the isomorphisms witnessing the preservation of the Identity Extension Lemma for each type constructor, and propagate them throughout the construction. Constructing a sound model directly from Reynolds' theory, however, poses two challenges. Firstly, we cannot interpret a type $\alpha \vdash T(\alpha)$ in a discrete fashion, as a pair of maps $T_0 : |\mathsf{Set}| \to \mathsf{Set}$ and $T_1 : |\mathsf{Rel}| \to \mathsf{Rel}$; indeed, since the domain of T_1 is the discrete category $|\mathsf{Rel}|$, T_1 is not required to preserve isomorphisms in Rel. As a result, even if we know that the pair (T_0, T_1) satisfies the Identity Extension Lemma, its reindexing, defined by precomposition, might not. Secondly, even on the set level, the interpretation of the $\forall$-type constructor as suggested by Reynolds does not commute with reindexing *on the nose* but only up to an isomorphism in Set.

We solve the first problem by specifying subcategories $\mathcal{M}(0) \subseteq \mathsf{Set}$ and $\mathcal{M}(1) \subseteq \mathsf{Rel}$ of *relevant isomorphisms* that form a *reflexive graph category with isomorphisms*. Abstractly, this structure gives us two *face maps* (called ∂_0 and ∂_1 in [2]), which represent the domain and codomain projections, and a *degeneracy* (called I in [2]), which represents the equality functor. We interpret a type $\alpha \vdash T(\alpha)$ as a pair of functors $T_0 : \mathcal{M}(0) \to \mathcal{M}(0)$ and $T_1 : \mathcal{M}(1) \to \mathcal{M}(1)$ that together comprise a *face map- and degeneracy-preserving reflexive graph functor*, and interpret each term as a *face map- and degeneracy-preserving reflexive graph natural transformation*. Since the domain of T_1 is $\mathcal{M}(1)$, T_1 preserves all relevant isomorphisms between relations, so the reindexing of (T_0, T_1) is now well-defined. Choosing $\mathcal{M}(1)$ to contain the isomorphism between the two relations $\mathsf{Eq}(S_0(A)) \times \mathsf{Eq}(T_0(A))$ and $\mathsf{Eq}(S_0(A) \times T_0(A))$ yields the satisfaction of the Identity Extension Lemma for products; other type constructors follow the same pattern.

To solve the second problem, we generalize the notion of a split $\lambda 2$-fibration to the new notion of a *$\lambda 2$-fibration with isomorphisms*, which allows type formers to commute with reindexing up to a suitable relevant isomorphism. Our main theorem shows that every such structure gives a sound model of System F in a natural way. By choosing $\mathcal{M}(0)$ to contain the isomorphism witnessing the commutativity of the $\forall$-type constructor with reindexing, we can use the main result to turn Reynolds' theory of relational parametricity into a sound model of System F, where, additionally, interpretations of types are endowed with a functorial action on isomorphisms and all polymorphic functions respect this action. These natural categorical features go beyond Reynolds' original theory and are essentially forced upon us in the process of making Reynolds' informal theory of parametricity into a sound model of System F.

Finally, we go one level higher and use the ideas of Orsanigo [11] (and Ghani *et al.* [3], which it supersedes) to define a *proof-relevant* model, in which witnesses of relatedness are themselves suitably related via a yet higher relation. This "2-parametric" model of course does not arise as an instance of our framework since it requires additional structure — *e.g.*, the concept of a *2-relation* — pertaining to the higher notion of parametricity. Nevertheless, we would still like to be able to recognize it as a model parametric in the ordinary sense. Various definitions of parametricity for models of System F exist: [2, 5] are examples of "internal" approaches to parametricity, where a model is considered parametric if it is produced via a specified procedure that bakes in desired features of parametricity such as the Identity Extension Lemma. On the other hand, [4, 7, 9, 14] are examples of "external" approaches to parametricity, in which reflexive graphs of models are used to endow models of interest with enough additional structure that they can reasonably be considered parametric. Surprisingly though, the proof-relevant model we give does not appear to satisfy any of these definitions, and in particular does not satisfy any of the external ones. The ability to construct a suitable reflexive graph seems to rely on an implicit assumption of proof-irrelevance, which we elaborate on in Section 6. However, we propose a new definition of a *relationally parametric model of System F* in Section 5 and show that it recognizes all the models we discuss, including the proof-relevant one, in a natural way.

The main contributions of this paper are as follows:

- *We turn Reynolds' original theory of parametricity into a sound model of System F, where interpretations of types are endowed with a functorial action on isomorphisms and all polymorphic functions respect this action. Adapting ideas from [11], we further extend this model one level up, to a model where witnesses of relatedness are themselves suitably related via a yet higher relation.*
- *We develop an abstract framework for relational parametricity that allows a choice of meta-theory (*e.g.*, the Calculus of Constructions, ω-sets and realizable functions,* etc.*), delivers the aforementioned model of Reynolds' parametricity as well as the canonical PER model of Longo and Moggi [8] as direct instances in a uniform way, and exposes properties that good models of System F parametricity should be expected to satisfy,* e.g.*, guaranteeing that interpretations of terms, not just types, suitably commute with the degeneracy.*
- *We construct our interpretation by first introducing the novel notion of a* $\lambda 2$-fibration with isomorphisms*, which allows type formers to commute with reindexing only up to isomorphism, and then proving a generalization of Seely's result ensuring that the canonical model induced by any such fibration validates System F's equational theory on the nose.*
- *We give a novel definition of a* parametric model of System F, *which is a hybrid of the external and internal approaches, and show that it subsumes both the PER model and the Reynolds model (expressed as instances of our framework), as well as the proof-relevant model.*

A technical report with detailed proofs is available on arxiv.org as arXiv:1805.00067.

2 Reflexive Graph Categories

Although Reynolds himself showed that his original approach to relational parametricity does not work in set theory, we can still use it as a guide for designing an abstract framework for parametricity. Instead of sets and relations, we consider abstract notions of "sets" and "relations", and require them to be related as follows: *i)* for any relation R, there are two canonical ways of projecting an object out of R, corresponding to the domain and codomain operations, *ii)* for any object A, there is a canonical way of turning it into a relation, corresponding to the equality relation on A, and *iii)* if we start with an object A, turn it into a relation according to *ii)*, and then project out an object according to *i)*, we get A back. This suggests that our abstract relations and the canonical operations on them can be organized into a reflexive graph structure: categories $\mathcal{X}_0$, $\mathcal{X}_1$ and functors $\mathbf{f}_\top, \mathbf{f}_\bot : \mathcal{X}_1 \to \mathcal{X}_0$, $\mathbf{d} : \mathcal{X}_0 \to \mathcal{X}_1$ such that $\mathbf{f}_\top \circ \mathbf{d} = \mathsf{id} = \mathbf{f}_\bot \circ \mathbf{d}$, as is done in [2].

Since there are no set-theoretic models of System F ([13]), all of the reflexive graph structure identified above must to be internal to some ambient category C. In particular, $\mathcal{X}_0$ and $\mathcal{X}_1$ must be categories internal in C, and $\mathbf{f}_\top, \mathbf{f}_\bot$, and $\mathbf{d}$ must be functors internal in C. For Reynolds' original model, the ambient category has types $A : \mathbb{U}_1$ as objects and terms $f : \Sigma_{A,B:\mathbb{U}_1} A \to B$ as morphisms. Here, $\mathbb{U}_1$ is the universe one level above the impredicative universe $\mathbb{U}_0$; we will denote $\mathbb{U}_0$ simply by $\mathbb{U}$ below. This ensures that $\mathbb{U}$ is an object in C. To model relations, we introduce:

$$\mathsf{isProp}(A) \coloneqq \Pi_{a,b:A} \mathsf{Id}(a,b)$$
$$\mathsf{Prop} \coloneqq \Sigma_{A:\mathbb{U}} \mathsf{isProp}(A)$$

The type Prop of *propositions* singles out those types in $\mathbb{U}$ with the property that any two inhabitants, if they exist, are equal. Propositions can be used to model relations as follows: in Reynolds' original model, $a : A$ is related to $b : B$ in at most one way under any relation R (either $(a,b) \in R$ or not), so the type of proofs that $(a,b) \in R$ is a proposition. Conversely, given $R : A \times B \to \mathsf{Prop}$, we consider a and b to be related by R iff $R(a,b)$ is inhabited.

To see the universe $\mathbb{U}$ as a category Set internal to C we take its object of objects Set_0 to be $\mathbb{U}$ and define its object of morphisms by $\mathsf{Set}_1 \coloneqq \Sigma_{A,B:\mathbb{U}} A \to B$. We define the category R of relations by giving its objects R_0 and R_1 of objects and morphisms, respectively:

$$\mathsf{R}_0 \coloneqq \Sigma_{A,B:\mathsf{Set}} A \times B \to \mathsf{Prop}$$
$$\mathsf{R}_1 \coloneqq \Sigma_{((A_1,A_2),R_A),((B_1,B_2),R_B):\mathsf{R}_0} \Sigma_{(f,g):(A_1 \to B_1)\times(A_2 \to B_2)} \Pi_{(a_1,a_2):A_1\times A_2} R_A(a_1,a_2) \to R_B(f(a_1),g(a_2))$$

We clearly have two internal functors from R to Set corresponding to the domain and codomain projections, respectively. We also have an internal functor Eq from Set to R that constructs an equality relation with $\mathsf{Eq}\,A \coloneqq ((A,A),\mathsf{Id}_A)$ and $\mathsf{Eq}\,((A,B),f) \coloneqq \big((\mathsf{Eq}\,A,\mathsf{Eq}\,B),(f,f),\mathsf{ap}_f\big)$. Here, the term ap_f of type $\mathsf{Id}_A(a_1,a_2) \to \mathsf{Id}_B(f(a_1),f(a_2))$ is defined as usual by Id-induction and witnesses the fact that f respects equality.

These observations motivate the next two definitions, in which we denote the category of categories and functors internal to C by $\mathsf{Cat}(C)$, and assume C is locally small and has all finite products. (A category is *locally small* if each of its hom-sets is small, *i.e.*, is a set rather than a proper class.)

Definition 2.1. A *reflexive graph structure* $\mathcal{X}$ on a category C consists of:

- objects $\mathcal{X}(0)$ and $\mathcal{X}(1)$ of C
- distinct arrows $\mathcal{X}(\mathbf{f}_\star) : \mathcal{X}(1) \to \mathcal{X}(0)$ for $\star$: **Bool**
- an arrow $\mathcal{X}(\mathbf{d}) : \mathcal{X}(0) \to \mathcal{X}(1)$

such that $\mathcal{X}(\mathbf{f}_\star) \circ \mathcal{X}(\mathbf{d}) = \mathsf{id}$.

The requirement that the two face maps $\mathcal{X}(\mathbf{f}_\top)$ and $\mathcal{X}(\mathbf{f}_\bot)$ are distinct is to ensure that there are enough relations for the notion of relation-preservation to be meaningful. Otherwise, as also observed in [2], we could see *any* category C as supporting a trivial reflexive graph structure whose only relations are the equality ones. For readers familiar with [7], the condition $\mathcal{X}(\mathbf{f}_\top) \neq \mathcal{X}(\mathbf{f}_\bot)$ serves a purpose similar to that of the requirement in Definition 8.6.2 of [7] that the fiber category $\mathbb{F}_1$ over the terminal object in $\mathbb{C}$ is the category of relations in the preorder fibration $\mathbb{D} \to \mathbb{E}$ on the fiber category $\mathbb{E}_1$ over the terminal object in $\mathbb{B}$. Both conditions imply that some relations must be *heterogeneous*. But while in [7] relations are obtained in a standard way as predicates (given by a preorder fibration) over a product, we do not assume that relations are constructed in any specific way, but rather only that the abstract operations on relations suitably interact. Moreover, since the two face maps $\mathcal{X}(\mathbf{f}_\top)$ and $\mathcal{X}(\mathbf{f}_\bot)$ are distinct, any morphism generated by the face maps and the degeneracy $\mathcal{X}(\mathbf{d})$ must be one of the seven distinct maps $\mathsf{id}_{\mathcal{X}(0)}, \mathsf{id}_{\mathcal{X}(1)}, \mathcal{X}(\mathbf{f}_\star), \mathcal{X}(\mathbf{d})$, and $\mathcal{X}(\mathbf{d}) \circ \mathcal{X}(\mathbf{f}_\star)$ for $\star$: **Bool**. Every such morphism thus has a canonical representation.

Definition 2.2. A *reflexive graph category (on C)* is a reflexive graph structure on $\mathrm{Cat}(C)$.

Example 2.3 (PER model)**.** We take the ambient category C to be the category of ω-sets, given in Definition 6.3 of [8]. We construct a reflexive graph category, which we call $\mathcal{R}_{PER}$, as follows. The internal category $\mathcal{R}_{PER}(0)$ of "sets" is the category $\mathbf{M}'$ given in Definition 8.4 of [8]. Informally, the objects of $\mathbf{M}'$ are partial equivalence relations on $\mathbb{N}$, and the morphisms are realizable functions that respect such relations. To define the internal category $\mathcal{R}_{PER}(1)$ of "relations", we first construct its object of objects. The carrier of this ω-set is the set of pairs of the form $R := ((A_\mathrm{d}, A_\mathrm{c}), R_A)$, where A_d and A_c are partial equivalence relations and R_A is a saturated predicate on the product PER $A_\mathrm{d} \times A_\mathrm{c}$. A saturated predicate on a PER A is a predicate on $\mathbb{N}$ such that $a_1 \sim_A a_2$ and $R(a_1)$ imply $R(a_2)$. To finish the construction of our object of objects for $\mathcal{R}_{PER}(1)$ we take any pair $((A_\mathrm{d}, A_\mathrm{c}), R_A)$ as above to be realized by any natural number.

The carrier of the object of morphisms for $\mathcal{R}_{PER}(1)$ comprises all pairs of the form

$$\big(\big(((A_\mathrm{d}, A_\mathrm{c}), R_A), ((B_\mathrm{d}, B_\mathrm{c}), R_B)\big), \big(\{m_1\}_{A_\mathrm{d}\to B_\mathrm{d}}, \{m_2\}_{A_\mathrm{c}\to B_\mathrm{c}}\big)\big)$$

satisfying the condition that, for any k, l such that $k \sim_{A_\mathrm{d}} k$, $l \sim_{A_\mathrm{c}} l$, and $R_A(\langle k, l\rangle)$ holds, $R_B\big(\langle m_1 \cdot k, m_2 \cdot l\rangle\big)$ holds as well. The first component records the domain and codomain of the morphism and the second component is a pair of equivalence classes under the specified exponential PERs. As in [8], we denote the application of the n^{th} partial recursive function to a natural number a in its domain by $n \cdot a$. To finish the construction of the object of morphisms for $\mathcal{R}_{PER}(1)$, we take a pair of pairs as above to be realized by a natural number k iff $\mathsf{fst}(k) \sim_{A_\mathrm{d}\to B_\mathrm{d}} m_1$ and $\mathsf{snd}(k) \sim_{A_\mathrm{c}\to B_\mathrm{c}} m_2$.

We again have two internal functors $\mathcal{R}_{PER}(\mathbf{f}_\top)$ and $\mathcal{R}_{PER}(\mathbf{f}_\bot)$ from $\mathcal{R}_{PER}(1)$ to $\mathcal{R}_{PER}(0)$ corresponding to the two projections. We also have an equality functor Eq from $\mathcal{R}_{PER}(0)$ to $\mathcal{R}_{PER}(1)$ whose action on objects is given by $\mathsf{Eq}\,A := ((A, A), \Delta_A)$, where $\Delta_A(k)$ iff $\mathsf{fst}(k) \sim_A \mathsf{snd}(k)$, and whose action on morphisms is given by

$$\mathsf{Eq}\,((A, B), \{m\}_{A\to B}) := \big((\mathsf{Eq}\,A, \mathsf{Eq}\,B), (\{m\}_{A\to B}, \{m\}_{A\to B})\big)$$

Example 2.4 (Reynolds' model)**.** We obtain a reflexive graph category $\mathcal{R}_{REY}$ by taking $\mathcal{R}_{REY}(0) := \mathsf{Set}$, $\mathcal{R}_{REY}(1) := \mathsf{R}$, and $\mathcal{R}_{REY}(\mathbf{d}) := \mathsf{Eq}$, and letting $\mathcal{R}_{REY}(\mathbf{f}_\top)$ and $\mathcal{R}_{REY}(\mathbf{f}_\bot)$ be the functors corresponding to the domain and codomain projections, respectively.

If $\mathcal{X}$ is a reflexive graph category, then the discrete graph category $|\mathcal{X}|$ and the product reflexive graph category $\mathcal{X}^n$ for $n \in \mathbb{N}$ are defined in the obvious ways: $|\mathcal{X}(l)|$ has the same objects as $\mathcal{X}(l)$ but only the identity morphisms, and $(\mathcal{X} \times \mathcal{Y})(l) = \mathcal{X}(l) \times \mathcal{Y}(l)$ for $l \in \{0, 1\}$. For the latter, the product on the right-hand side is a product of internal categories, which exists because C has finite products by assumption. To simplify the presentation, we will omit explicit mentions of the category C, and treat definitions and constructions internal to C as though they were external.

Given a reflexive graph category $\mathcal{X}$ axiomatizing the sets and relations, an obvious first attempt at pushing Reynolds' original idea through is to take the interpretation $[\![T]\!]$ of a type $\overline{\alpha} \vdash T$ with n free type variables to be a pair $([\![T]\!](0), [\![T]\!](1))$, where $[\![T]\!](0) : |\mathcal{X}(0)|^n \to \mathcal{X}(0)$ and $[\![T]\!](1) : |\mathcal{X}(1)|^n \to \mathcal{X}(1)$ are functions giving the "set" and "relation" interpretations of the type T. Although as explained in the introduction, this approach will need some tweaking — we will need to endow $[\![T]\!](0)$ and $[\![T]\!](1)$ with actions on *some* morphisms — it suggests:

Definition 2.5. Let $\mathcal{X}$ and $\mathcal{Y}$ be reflexive graph categories. A *reflexive graph functor* $\mathcal{F} : \mathcal{X} \to \mathcal{Y}$ is a pair $(\mathcal{F}(0), \mathcal{F}(1))$ of functors such that $\mathcal{F}(0) : \mathcal{X}(0) \to \mathcal{Y}(0)$ and $\mathcal{F}(1) : \mathcal{X}(1) \to \mathcal{Y}(1)$.

Writing T_0 for $[\![T]\!](0)$ and T_1 for $[\![T]\!](1)$, we recall from the introduction that T_0 and T_1 should be appropriately related via the domain and codomain projections and the equality functor. Since the two face maps $\mathcal{X}(\mathbf{f}_\star)$ now model the projections, and the degeneracy $\mathcal{X}(\mathbf{d})$ models the equality functor, we end up with the following conditions: *i)* for each object $\overline{R}$ in $\mathcal{X}(1)^n$, we have $\mathcal{X}(\mathbf{f}_\star)\,T_1(\overline{R}) = T_0(\mathcal{X}(\mathbf{f}_\star)^n\,\overline{R})$, and *ii)* for each object $\overline{A}$ in $\mathcal{X}(0)^n$, we have $\mathcal{X}(\mathbf{d})\,T_0(\overline{A}) = T_1(\mathcal{X}(\mathbf{d})^n\,\overline{A})$. We examine what these conditions imply for Reynolds' model by considering the product $\alpha \vdash S(\alpha) \times T(\alpha)$ of two types $\alpha \vdash S(\alpha)$ and $\alpha \vdash T(\alpha)$. By the induction hypothesis, S and T are interpreted as pairs (S_0, S_1) and (T_0, T_1), where $S_0, T_0 : \mathsf{Set}_0 \to \mathsf{Set}_0$ and $S_1, T_1 : \mathsf{R}_0 \to \mathsf{R}_0$ satisfy *i)* and *ii)*. The interpretation of a product should be a product of interpretations, *i.e.*, $(S\times T)_0\,A := S_0(A)\times T_0(A)$ and $(S\times T)_1\,R := S_1(R)\times T_1(R)$. It remains to be seen that this interpretation satisfies *i)* and *ii)*.

Fix a relation R on A and B. Condition *i)* entails that $S_1(R) := ((S_0(A), S_0(B)), R_S)$ and $T_1(R) := ((T_0(A), T_0(B)), R_T)$ for some R_S and R_T. Thus $S_1(R) \times T_1(R)$ has the form $\big((S_0(A) \times T_0(A), S_0(B) \times T_0(B)), R_{S\times T}\big)$, where $R_{S\times T}$ maps a pair of pairs $((a, b), (c, d))$ to $R_S(a, c)\times R_T(b, d)$. Thus *i)* is satisfied simply by construction, which leads us to define:

Definition 2.6. A reflexive graph functor $\mathcal{F} : \mathcal{X} \to \mathcal{Y}$ is *face map-preserving* if $\mathcal{Y}(\mathbf{f}_\star) \circ \mathcal{F}(1) = \mathcal{F}(0) \circ \mathcal{X}(\mathbf{f}_\star)$ for every $\star \in \mathbf{Bool}$.

In Reynolds' model, condition *ii)* gives that $S_1(\mathsf{Eq}(A))$ is $\mathsf{Eq}(S_0(A))$ for any set A, and similarly for T. We thus need to show that $\mathsf{Eq}(S_0(A)) \times \mathsf{Eq}(T_0(A))$ is $\mathsf{Eq}(S_0(A) \times T_0(A))$. But while the domains and codomains of these two relations agree (all are $S_0(A) \times T_0(A)$), the former maps $((a, b), (c, d))$ to $\mathsf{Id}(a, c) \times \mathsf{Id}(b, d)$, while the latter maps it to $\mathsf{Id}((a, b), (c, d))$. These two types are not necessarily identical, but they are *isomorphic* (*i.e.*, there are functions back and forth that compose to identity on both sides).

We thus relax condition *ii)* to allow an isomorphism $\varepsilon_T(\overline{A}) : \mathcal{X}(\mathbf{d})\,T_0(\overline{A}) \cong T_1(\mathcal{X}(\mathbf{d})^n\,\overline{A})$. In fact, we can require more: since the domains and codomains of $\mathcal{X}(\mathbf{d})\,T_0(\overline{A})$ and $T_1(\mathcal{X}(\mathbf{d})^n\,\overline{A})$ coincide by condition *i)*, we can insist that both projections map the isomorphism $\varepsilon_T(\overline{A})$ to the identity morphism on $T_0(\overline{A})$. This coherence condition is a natural counterpart to the equation $\mathcal{X}(\mathbf{f}_\star)\circ\mathcal{X}(\mathbf{d}) = \mathrm{id}$, and turns out to be not just a design choice but a necessary requirement: in Reynolds' model, for instance, the proof that the interpretations of $\forall$-types (as defined later) suitably commute with the functor Eq depends precisely on the morphisms underlying the maps $\varepsilon_T(\overline{A})$ being identities. This suggests:

Definition 2.7. A reflexive graph functor $\mathcal{F} : \mathcal{X} \to \mathcal{Y}$ is said to be *degeneracy-preserving* if there is a natural isomorphism $\varepsilon_\mathcal{F} : \mathcal{Y}(\mathbf{d}) \circ \mathcal{F}(0) \to \mathcal{F}(1) \circ \mathcal{X}(\mathbf{d})$ satisfying the coherence condition $\mathcal{Y}(\mathbf{f}_\star) \circ \varepsilon_\mathcal{F} = \mathrm{id}$.

As a first approximation, we can try to interpret a type $\overline{\alpha} \vdash T$ with n free type variables as a face map- and degeneracy-preserving reflexive graph functor $(T_0, T_1) : |\mathcal{X}|^n \to \mathcal{X}$. Reynolds' original idea for interpreting terms suggests that the interpretation of a term $\overline{\alpha}; x : S \vdash t : T$ should be a (vacuously) natural transformation $t_0 : S_0 \to T_0$. As observed in [5], the Abstraction Theorem can then be formulated as follows: there is a (vacuously) natural transformation $t_1 : S_1 \to T_1$ such that, for any object $\overline{R}$ in $\mathcal{X}(1)^n$, we have

$\mathcal{X}(\mathbf{f}_\star)\, t_1(\overline{R}) = t_0(\mathcal{X}(\mathbf{f}_\star)^n\, \overline{R})$. To see that this does indeed give what we want, we revisit Reynolds' model. There, the face maps are the domain and codomain projections and an object $\overline{R}$ in $\mathcal{X}(1)^n$ is an n-tuple of relations. Denote $\mathcal{X}(\mathbf{f}_\top)^n\, \overline{R}$ by $\overline{A}$ and $\mathcal{X}(\mathbf{f}_\perp)^n\, \overline{R}$ by $\overline{B}$. Then $t_1(\overline{R})$ is a morphism of relations from $S_1(\overline{R})$ to $T_1(\overline{R})$ and, since S_1 and T_1 are face map-preserving, $S_1(\overline{R}) := \big((S_0(\overline{A}), S_0(\overline{B})), R_S\big)$ and $T_1(\overline{R}) := \big((T_0(\overline{A}), T_0(\overline{B})), R_T\big)$ for some R_S and R_T. By definition, $t_1(\overline{R})$ gives maps $f : S_0(\overline{A}) \to T_0(\overline{A})$, $g : S_0(\overline{B}) \to T_0(\overline{B})\big)$, together with a map $h : \Pi_{(a_1, a_2): S_0(\overline{A}) \times S_0(\overline{B})} R_S(a_1, a_2) \to R_T(f(a_1), g(a_2))$ stating precisely that f and g map related inputs to related outputs. By definition, $\mathcal{X}(\mathbf{f}_\top)\, t_1(\overline{R})$ is $\big((S_0(\overline{A}), T_0(\overline{A})), f\big)$ and $\mathcal{X}(\mathbf{f}_\perp)\, t_1(\overline{R})$ is $\big((S_0(\overline{B}), T_0(\overline{B}), g\big)$, so the condition that $\mathcal{X}(\mathbf{f}_\star)\, t_1(\overline{R})$ is $t_0(\mathcal{X}(\mathbf{f}_\star)^n\, \overline{R})$ implies that the maps underlying $t_0(\overline{A})$ and $t_0(\overline{B})$ must be f and g, respectively, and so must indeed map related inputs to related outputs, as witnessed by h. Pairing the natural transformations t_0 and t_1 motivates:

Definition 2.8. Let $\mathcal{F}, \mathcal{G} : \mathcal{X} \to \mathcal{Y}$ be reflexive graph functors. A *reflexive graph natural transformation* $\eta : \mathcal{F} \to \mathcal{G}$ is a pair $(\eta(0), \eta(1))$ of natural transformations $\eta(0) : \mathcal{F}(0) \to \mathcal{G}(0)$ and $\eta(1) : \mathcal{F}(1) \to \mathcal{G}(1)$.

The Abstraction Theorem then further suggests defining:

Definition 2.9. A reflexive graph natural transformation $\eta : \mathcal{F} \to \mathcal{G}$ is *face map-preserving* if $\mathcal{F}$ and $\mathcal{G}$ are face map-preserving and, for each $\star \in \mathbf{Bool}$, we have $\mathcal{Y}(\mathbf{f}_\star) \circ \eta(1) = \eta(0) \circ \mathcal{X}(\mathbf{f}_\star)$.

The interpretation of a term $\overline{\alpha}; x : S \vdash t : T$ should then be a face map-preserving natural transformation from (S_0, S_1) to (T_0, T_1). We also have the dual notion:

Definition 2.10. A reflexive graph natural transformation $\eta : \mathcal{F} \to \mathcal{G}$ is *degeneracy-preserving* if $\mathcal{F}$ and $\mathcal{G}$ are degeneracy-preserving, as witnessed by the natural isomorphisms $\varepsilon_{\mathcal{F}}$ and $\varepsilon_{\mathcal{G}}$, respectively, and, for every X in $\mathcal{X}(0)$, we have $\big(\eta(1)\ (\mathcal{X}(\mathbf{d})\ X)\big) \circ \varepsilon_{\mathcal{F}}(X) = \varepsilon_{\mathcal{G}}(X) \circ \big(\mathcal{Y}(\mathbf{d})\ (\eta(0)\ X)\big)$.

There is no explicit analogue of Definition 2.10 in Reynolds' model for the following reason: Reynolds' model (as well as the PER model) is proof-irrelevant, in the precise sense that the functor $\langle \mathcal{X}(\mathbf{f}_\perp), \mathcal{X}(\mathbf{f}_\top) \rangle$ is faithful, and this condition is sufficient to guarantee that *any* face map-preserving natural transformation is automatically degeneracy-preserving as well. This may or may not be the case in proof-relevant models (although in the model from Section 6 it is), so we explicitly restrict attention below only to those natural transformations that are face map- *and* degeneracy-preserving (as also done in [2]), and omit further mention of these properties.

Identity and composition for natural transformations between reflexive graph functors are defined levelwise. Identity for reflexive graph functors is also obvious, but composition requires some care:

Definition 2.11. For reflexive graph functors $\mathcal{F} : \mathcal{X} \to \mathcal{Y}$ and $\mathcal{G} : \mathcal{Y} \to \mathcal{Z}$, the reflexive graph functor $\mathcal{G} \circ \mathcal{F} : \mathcal{X} \to \mathcal{Z}$ is defined as follows:

- $(\mathcal{G} \circ \mathcal{F})(l) := \mathcal{G}(l) \circ \mathcal{F}(l)$
- $\varepsilon_{\mathcal{G} \circ \mathcal{F}}(X) := (\mathcal{G}(1)\ \varepsilon_{\mathcal{F}}) \circ \varepsilon_{\mathcal{G}}(\mathcal{F}(0)\ X)$

Here, the first composition is a composition of functors and the second is a composition of morphisms in the category $\mathcal{Z}(1)$.

Given reflexive graph functors $\mathcal{F}_1, \mathcal{F}_2 : \mathcal{X} \to \mathcal{Y}$ and $\mathcal{G}_1, \mathcal{G}_2 : \mathcal{Y} \to \mathcal{Z}$, and natural transformations $\varepsilon : \mathcal{F}_1 \to \mathcal{F}_2$ and $\eta : \mathcal{G}_1 \to \mathcal{G}_2$, the compositions $\eta \circ \mathcal{F}_1 : \mathcal{G}_1 \circ \mathcal{F}_1 \to \mathcal{G}_2 \circ \mathcal{F}_1$ and $\mathcal{G}_1 \circ \varepsilon : \mathcal{G}_1 \circ \mathcal{F}_1 \to \mathcal{G}_1 \circ \mathcal{F}_2$ are defined levelwise in the obvious way. A particular reflexive graph functor of interest, which we will use to interpret type variables, is projection:

Definition 2.12. Given a reflexive graph category $\mathcal{X}$ and $i \in \{1, ..., n\}$, the i^{th} *reflexive graph projection functor* is the reflexive graph functor $\mathrm{pr}^n_i : \mathcal{X}^n \to \mathcal{X}$, where $\mathrm{pr}^n_i(l) : \mathcal{X}(l)^n \to \mathcal{X}(l)$ is the usual i^{th} projection functor and $\varepsilon_{\mathrm{pr}^n_i}(X) := \mathrm{id}$.

Dually, we have the following:

Definition 2.13. For reflexive graph functors $\mathcal{F}_1, \ldots, \mathcal{F}_m : \mathcal{X} \to \mathcal{Y}$, the reflexive graph functor $\langle \mathcal{F}_1, \ldots, \mathcal{F}_m \rangle : \mathcal{X} \to \mathcal{Y}^m$ is defined as follows:

- $\langle \mathcal{F}_1, \ldots, \mathcal{F}_m \rangle(l) := \langle \mathcal{F}_1(l), \ldots, \mathcal{F}_m(l) \rangle$
- $\varepsilon_{\langle \mathcal{F}_1, \ldots, \mathcal{F}_m \rangle}(X) := \langle \varepsilon_{\mathcal{F}_1}(X), \ldots, \varepsilon_{\mathcal{F}_m}(X) \rangle$

Similarly, given reflexive graph natural transformations $\eta_1 : \mathcal{F}_1 \to \mathcal{G}_1, \ldots, \eta_m : \mathcal{F}_m \to \mathcal{G}_m$, the reflexive graph natural transformation $\langle \eta_1, \ldots, \eta_m \rangle : \langle \mathcal{F}_1, \ldots, \mathcal{F}_m \rangle \to \langle \mathcal{G}_1, \ldots, \mathcal{G}_m \rangle$ is defined in the obvious way.

3 Reflexive Graph Categories with Isomorphisms

As noted above, if we try to interpret a type $\overline{\alpha} \vdash T$ as a reflexive graph functor $[\![T]\!] : \mathcal{X}^n \to \mathcal{X}$ we encounter a problem with contravariance. Specifically, if $\alpha \vdash A$ and $\alpha \vdash B$ are types, then to interpret the function type $\alpha \vdash A \to B$ as the exponential of $[\![A]\!]$ and $[\![B]\!]$, $[\![A \to B]\!](0)$ must map each object X to the exponential $([\![A]\!](0)\, X) \Rightarrow ([\![B]\!](0)\, X)$ and each morphism $f : X \to Y$ to a morphism from $([\![A]\!](0)\, X) \Rightarrow ([\![B]\!](0)\, X)$ to $([\![A]\!](0)\, Y) \Rightarrow ([\![B]\!](0)\, Y)$. But there is no canonical way to construct a morphism of this type because $[\![A]\!](0)\, f$ goes in the wrong direction. This is a well-known problem that is unrelated to parametricity.

The usual solution is to require the domains of the functors interpreting types to be discrete, so that $[\![T]\!] : |\mathcal{X}|^n \to \mathcal{X}$. However, as noted in the introduction, this will not work in our setting. Consider types $\alpha \vdash S(\alpha)$ and $\cdot \vdash T$. By the induction hypothesis, $[\![S]\!] : |\mathcal{X}| \to \mathcal{X}$ and $[\![T]\!] : 1 \to \mathcal{X}$ are face map- and degeneracy-preserving reflexive graph functors. The interpretation of the type $\cdot \vdash S[\alpha := T]$ should be given by the composition $[\![S]\!] \circ [\![T]\!] : 1 \to \mathcal{X}$, which should be a face map- and degeneracy-preserving functor. While preservation of face maps is easy to prove, preservation of degeneracies poses a problem: writing S_0 and S_1 for $[\![S]\!](0)$ and $[\![S]\!](1)$, and similarly for T, we need $S_1(T_1)$ to be isomorphic to the degeneracy $\mathbf{d}(S_0(T_0))$. By assumption, T_1 is isomorphic to the degeneracy $\mathbf{d}(T_0)$, and $S_1(\mathbf{d}(T_0))$ is isomorphic to $\mathbf{d}(S_0(T_0))$, so if we knew that S_1 mapped isomorphic relations to isomorphic relations we would be done. But since the domain of S_1 is $|\mathcal{X}(1)|$, there is no reason that it should preserve non-identity isomorphisms of $\mathcal{X}(1)$.

In this paper we solve this contravariance problem in a different way. We first note that the issue does not arise if $[\![A]\!](0)\, f$ is an isomorphism, even if that isomorphism is not the identity. This leads us to require, for each $l \in \{0, 1\}$, a wide subcategory $\mathcal{M}(l) \subseteq \mathcal{X}(l)$ such that every morphism in $\mathcal{M}(l)$ is in fact an isomorphism.

Definition 3.1. Given a reflexive graph category $\mathcal{X}$, a *reflexive graph subcategory* of $\mathcal{X}$ is a reflexive graph category $\mathcal{M}$ together with a reflexive graph functor $\mathcal{I} : \mathcal{M} \to \mathcal{X}$ such that:

- The object and morphism parts of $\mathcal{I}$ are monomorphisms.

- $\mathcal{I}(0) \circ \mathcal{M}(\mathbf{f}_\star) = \mathcal{X}(\mathbf{f}_\star) \circ \mathcal{I}(1)$ for $\star \in \mathbf{Bool}$.
- $\mathcal{I}(1) \circ \mathcal{M}(\mathbf{d}) = \mathcal{X}(\mathbf{d}) \circ \mathcal{I}(1)$.

The subcategory $(\mathcal{M}, \mathcal{I})$ is *wide* if the object parts of $\mathcal{I}(0)$ and $\mathcal{I}(1)$ are isomorphisms.

The last two conditions in Definition 3.1 guarantee that $\mathcal{I}$ preserves face maps and degeneracies on the nose.

Definition 3.2. A *reflexive graph category with isomorphisms* is a reflexive graph category $\mathcal{X}$ together with a wide reflexive graph subcategory $(\mathcal{M}, \mathcal{I})$ such that every morphism in $\mathcal{M}(l)$, $l \in \{0, 1\}$, is an isomorphism.

We view $\mathcal{M}(l)$ as selecting the *relevant isomorphisms* of $\mathcal{X}(l)$, in the sense that a morphism of $\mathcal{X}(l)$ is relevant iff it lies in the image of $\mathcal{I}(l)$. Given a reflexive graph category with isomorphisms $(\mathcal{X}, (\mathcal{M}, \mathcal{I}))$ we can now interpret a type $\overline{\alpha} \vdash T$ with n free type variables as a reflexive graph functor $[\![T]\!] : \mathcal{M}^n \to \mathcal{M}$. It is important that $[\![T]\!]$ carries (tuples of) relevant isomorphisms to relevant isomorphisms: if $[\![T]\!]$ were instead a functor from $\mathcal{M}^n$ to $\mathcal{X}$, then it would not be possible to define substitution (see Definition 4.2).

A trivial choice is to take $\mathcal{M} := |\mathcal{X}|$. Then $[\![T]\!] : |\mathcal{X}|^n \to |\mathcal{X}|$ and $\varepsilon_{[\![T]\!]}$ is necessarily the identity natural transformation, so $[\![T]\!]$ preserves degeneracies on the nose. This instantiation shows that, despite being motivated by Reynolds' model, for which the Identity Extension Lemma holds only up to isomorphism, our framework can also uniformly subsume strict models of parametricity, for which the Identity Extension Lemma holds on the nose.

Example 3.3 (PER model, continued)**.** We take $\mathcal{M} := |\mathcal{R}_{PER}|$.

Example 3.4 (Reynolds' model, continued)**.** For each l, we take the objects of $\mathcal{M}(l)$ to be the objects of $\mathcal{R}_{REY}(l)$, and the morphisms of $\mathcal{M}(l)$ to be *all* isomorphisms of $\mathcal{R}_{REY}(l)$. For example, the morphisms of $\mathcal{M}(0)$ are

$$\{(i, j) : \mathsf{Set}_1 \times \mathsf{Set}_1 \ \& \\ i_\mathrm{d} = j_\mathrm{c} \times i_\mathrm{c} = j_\mathrm{d} \times j \circ i = \mathrm{id} \times i \circ j = \mathrm{id}\}$$

Here and at several places below we write $a = b$ for $\mathsf{Id}(a, b)$ and $\{x : A \ \& \ B(x)\}$ for $\Sigma_{x:A} B(x)$ to enhance readability. Moreover, $\circ$ and id are composition and identity in the category Set, and we use the subscripts $(\cdot)_\mathrm{d}$ and $(\cdot)_\mathrm{c}$ to denote the domain and codomain of a morphism. The first (or second) projection gives the required mono from $\mathcal{M}(0)$ to Set_1.

With this infrastructure in place we can now interpret a term $\overline{\alpha}; x : S \vdash t : T$ as a natural transformation from $\mathcal{I} \circ [\![S]\!]$ to $\mathcal{I} \circ [\![T]\!]$. Importantly, the components of such a natural transformation are drawn from $\mathcal{X}(l)$ (as witnessed by post-composition with $\mathcal{I}$), rather than just $\mathcal{M}(l)$, as would be the case if we interpreted t as a natural transformation from $[\![S]\!]$ to $[\![T]\!]$. In fact, this latter interpretation would not even be sensible, since not every term gives rise to an isomorphism (most do not).

4 Cartesian Closed Reflexive Graph Categories with Isomorphisms

We want to interpret a type context of length n as the natural number n, types with n free type variables as reflexive graph functors from $\mathcal{M}^n$ to $\mathcal{M}$, and terms with n free type variables as natural transformations between reflexive graph functors with codomain $\mathcal{X}$. Following the standard procedure, we first define, for each n, a category $\mathcal{M}^n \to \mathcal{M}$ to interpret expressions with n free type variables, and then combine these categories using the usual Grothendieck construction. This gives a fibration whose fiber over n is $\mathcal{M}^n \to \mathcal{M}$.

Definition 4.1. The category $\mathcal{M}^n \to \mathcal{M}$ has as its objects the face map- and degeneracy-preserving reflexive graph functors from $\mathcal{M}^n$ to $\mathcal{M}$, and as its morphisms from $\mathcal{F}$ to $\mathcal{G}$ the face map- and degeneracy-preserving reflexive graph natural transformations from $\mathcal{I} \circ \mathcal{F}$ to $\mathcal{I} \circ \mathcal{G}$.

If $\mathcal{F}$ and $\mathcal{G}$ are degeneracy-preserving then $\mathcal{I} \circ \mathcal{F}$ and $\mathcal{I} \circ \mathcal{G}$ are as well, and it is therefore sensible to require natural transformations between the latter two to be degeneracy-preserving. To move between the fibers we need a notion of substitution:

Definition 4.2. For any m-tuple $\mathbf{F} := (F_1, \ldots, F_m)$ of objects in $\mathcal{M}^n \to \mathcal{M}$, the functor $\mathbf{F}^*$ from $\mathcal{M}^m \to \mathcal{M}$ to $\mathcal{M}^n \to \mathcal{M}$ is defined by $\mathbf{F}^*(G) := G \circ \langle F_1, \ldots, F_m \rangle$ for objects and $\mathbf{F}^*(\eta) := \eta \circ \langle F_1, \ldots, F_m \rangle$ for morphisms.

When giving a categorical interpretation of System F, a category for interpreting type contexts is also required. Writing $\mathcal{R}$ for the tuple $(\mathcal{X}, (\mathcal{M}, \mathcal{I}))$, we define:

Definition 4.3. The *category of contexts* $\mathsf{Ctx}(\mathcal{R})$ is given by:

- objects are natural numbers
- morphisms from n to m are m-tuples of objects in $\mathcal{M}^n \to \mathcal{M}$
- the identity $\mathsf{id}_n : n \to n$ has as its i^{th} component the i^{th} projection functor pr_i^n
- given morphisms $\mathbf{F} : n \to m$ and $\mathbf{G} = (G_1, \ldots, G_k) : m \to k$, the i^{th} component of the composition $\mathbf{G} \circ \mathbf{F} : n \to k$ is $\mathbf{F}^*(G_i)$

Defining the product $n \times 1$ in $\mathsf{Ctx}(\mathcal{R})$ to be the natural number sum $n + 1$ shows that $\mathsf{Ctx}(\mathcal{R})$ can model System F type contexts:

Lemma 4.4. *The category* $\mathsf{Ctx}(\mathcal{R})$ *has a terminal object 0 and products* $(-) \times 1$.

The categories $\mathsf{Ctx}(\mathcal{R})$ and $\mathcal{M}^n \to \mathcal{M}$ can be combined to give:

Definition 4.5. The category $\int_n \mathcal{M}^n \to \mathcal{M}$ is defined as follows:

- objects are pairs (n, F), where F is an object in $\mathcal{M}^n \to \mathcal{M}$
- morphisms from (n, F) to (m, G) are pairs $(\mathbf{F}, \eta)$, where $\mathbf{F} : n \to m$ is a morphism in $\mathsf{Ctx}(\mathcal{R})$ and $\eta : F \to \mathbf{F}^*(G)$ is a morphism in $\mathcal{M}^n \to \mathcal{M}$
- the identity on (n, F) is the pair $(\mathsf{id}_n, \mathsf{id}_F)$, where $\mathsf{id}_n : n \to n$ is the identity in $\mathsf{Ctx}(\mathcal{R})$ and $\mathsf{id}_F : F \to F$ is the identity in $\mathcal{M}^n \to \mathcal{M}$
- the composition of two morphisms $(\mathbf{F}, \eta_1) : (n, F) \to (m, G)$ and $(\mathbf{G}, \eta_2) : (m, G) \to (k, H)$ is the pair $(\mathbf{G} \circ \mathbf{F}, \mathbf{F}^*(\eta_2) \circ \eta_1)$, where the first composition is in $\mathsf{Ctx}(\mathcal{R})$ and the second composition is in $\mathcal{M}^n \to \mathcal{M}$

This is a standard (op)Grothendieck construction, and resuts in a category whose objects can be understood as pairing a kinding context and a typing context over it, and whose morphisms can be understood as simultaneous substitutions.

To appropriately interpret arrow types will we need to know that the category $\mathcal{M}^n \to \mathcal{M}$ is cartesian closed. We define:

Definition 4.6. A reflexive graph category with isomorphisms $\mathcal{R}$ *has terminal objects* if each $\mathcal{X}(l)$ has a terminal object $1_{\mathcal{X}(l)}$. The terminal objects are *stable under face maps* if, for all $\star \in \mathbf{Bool}$, the canonical morphism from $\mathcal{X}(\mathbf{f}_\star)\, 1_{\mathcal{X}(1)}$ to $1_{\mathcal{X}(0)}$ is the identity. The terminal objects are *stable under degeneracies* if the canonical morphism from $\mathcal{X}(\mathbf{d})\, 1_{\mathcal{X}(0)}$ to $1_{\mathcal{X}(1)}$ is in $\mathcal{M}(1)$.

Definition 4.7. A reflexive graph category with isomorphisms $\mathcal{R}$ *has products* if each $\mathcal{X}(l)$ has products $\times_l$ that preserve membership in $\mathcal{M}(l)$. The products are *stable under face maps* if, for all $\star \in \mathbf{Bool}$ and objects A, B in $\mathcal{X}(1)$, the canonical morphism from $\mathcal{X}(\mathbf{f}_\star)\,(A \times_1 B)$ to $(\mathcal{X}(\mathbf{f}_\star)\,A) \times_0 (\mathcal{X}(\mathbf{f}_\star)\,B)$ is the identity. It has products *stable under degeneracies* if, for all objects A, B in $\mathcal{X}(0)$, the canonical morphism from $\mathcal{X}(\mathbf{d})\,(A \times_0 B)$ to $(\mathcal{X}(\mathbf{d})\,A) \times_1 (\mathcal{X}(\mathbf{d})\,B)$ is in $\mathcal{M}(1)$.

Definition 4.8. A reflexive graph category with isomorphisms $\mathcal{R}$ that has products also *has exponentials* if each $\mathcal{X}(l)$ has exponentials $\Rightarrow_l$ that preserve membership in $\mathcal{M}(l)$. The exponentials are *stable under face maps* if, for all $\star \in \mathbf{Bool}$ and objects A, B in $\mathcal{X}(1)$, the canonical morphism from $\mathcal{X}(\mathbf{f}_\star)\,(A \Rightarrow_1 B)$ to $(\mathcal{X}(\mathbf{f}_\star)\,A) \Rightarrow_0 (\mathcal{X}(\mathbf{f}_\star)\,B)$ is the identity. It has exponentials *stable under degeneracies* if, for all objects A, B in $\mathcal{X}(0)$, the canonical morphism from $\mathcal{X}(\mathbf{d})\,(A \Rightarrow_0 B)$ to $(\mathcal{X}(\mathbf{d})\,A) \Rightarrow_1 (\mathcal{X}(\mathbf{d})\,B)$ is in $\mathcal{M}(1)$.

We combine the above to obtain the main definition of this section:

Definition 4.9. A reflexive graph category with isomorphisms is *cartesian closed* if it has terminal objects, products, and exponentials, all stable under face maps and degeneracies.

Example 4.10. [PER model, continued] Terminal objects, products, and exponentials are defined for $\mathcal{R}_{PER}$ in the obvious ways, inheriting from the corresponding constructs on PERs. It is not hard to check that all of these constructs are preserved on the nose by the two face maps (projections) and the degeneracy (equality functor), and thus, in our terminology, are stable under face maps and degeneracies.

Example 4.11. [Reynolds' model, continued] Here, too, terminal objects, products, and exponentials are defined for $\mathcal{R}_{REY}$ in the obvious ways, relating two pairs iff their first and second components are related, and two functions iff they map related arguments to related results. It is easy to see that all of these constructs are preserved on the nose (*i.e.*, up to *definitional* equality) by the projections, and thus are stable under face maps. Unlike in the PER model though, they are only preserved by the equality functor Eq up to (the canonical) isomorphism. For example, as discussed just after Definition 2.6, the two types $\mathsf{Id}((a,b),(c,d))$ and $\mathsf{Id}(a,c)\times\mathsf{Id}(b,d)$ for $(a,b),(c,d) : A \times B$ are not necessarily identical, although they are isomorphic under the canonical (iso)morphism from $\mathsf{Eq}(A \times B)$ to $\mathsf{Eq}(A) \times \mathsf{Eq}(B)$. A similar situation arises for function types $A \to B$: by function extensionality, $\mathsf{Id}(f,g)$ and $\Pi_{a,\,a':A}\mathsf{Id}(f(a), g(a'))$ are isomorphic, but not necessarily identical, via the canonical isomorphism. Nevertheless, we still get stability under degeneracies since we explicitly allowed for this possibility in Definition 4.8.

5 Reflexive Graph Models of Parametricity

As Examples 4.10 and 4.11 show, cartesian closed reflexive graph categories with isomorphisms suitably generalize the structure of sets and relations. Moreover, they allow us to interpret unit, product, and function types in a natural way:

Lemma 5.1. *If $\mathcal{R}$ is a cartesian closed reflexive graph category with isomorphisms, then the forgetful functor from $\int_n \mathcal{M}^n \to \mathcal{M}$ to $\mathsf{Ctx}(\mathcal{R})$ is a split cartesian closed fibration with a split generic object $\Omega \coloneqq 1$.*

To interpret $\forall$-types we need to know that, in the forgetful fibration from Lemma 5.1, each weakening functor induced by the first projection from $n+1$ to n for $n \in \mathbb{N}$ has a right adjoint $\forall_n$. Here we differ from [2], where only $\forall_0$ is required, with the intention that $\forall_n$ can be derived from $\forall_0$ using partial application. We observe that this approach does not appear to work since a partial application of an indexed functor is not necessarily an indexed functor. Hence we require an entire family of adjoints $\forall_n$.

Example 5.2 (PER model, continued)**.** Define the adjoint $\forall_n$ by

$$\begin{aligned}\forall_n\,\mathcal{F}(0)\,\overline{A} \;&\coloneqq \big\{(m,k) \mid \text{ for all } A,\ (m,k) \in \mathcal{F}(0)\,(\overline{A}, A), \\ &\qquad \text{and for all } R,\ \langle m,k\rangle \in \mathcal{F}(1)(\overline{\mathsf{Eq}\,A}, R)\big\} \\ \forall_n\mathcal{F}(1)\,\overline{R} \;&\coloneqq \Big(\big(\forall_n\mathcal{F}(0)\,\overline{R_{\mathrm{d}}}, \forall_n\mathcal{F}(0)\,\overline{R_{\mathrm{c}}}\big), \\ &\qquad \big\{m \mid \text{for all } R,\ m \in \mathcal{F}(1)\,(\overline{R}, R)\big\}\Big)\end{aligned}$$

where for any relation $R \coloneqq ((A_{\mathrm{d}}, A_{\mathrm{c}}), R_A)$ we write R_{d} for A_{d} and R_{c} for A_{c}. We will employ a similar convention for Reynolds' model. To define $\forall_n$ on a morphism $\eta : \mathcal{F} \to \mathcal{G}$, we put

$$\begin{aligned}\forall_n\,\eta(0)\,\overline{A} \;\coloneqq \Big(&\big(\forall_n\mathcal{F}(0)\,\overline{A}, \forall_n\mathcal{G}(0)\,\overline{A}\big), \\ &\{m \cdot 0\}_{(\forall_n\mathcal{F}(0)\,\overline{A})\to(\forall_n\mathcal{G}(0)\,\overline{A})}\Big)\end{aligned}$$

where m is any natural number realizing $\eta(0)\,\overline{A}$.

Example 5.3 (Reynolds' model, continued)**.** On sets, the adjoint $\forall_n$ is defined as follows:

$$\begin{aligned}\forall_n\,\mathcal{F}(0)\,\overline{A} \quad \coloneqq \quad &\{f_0 : \Pi_{A:\mathbb{U}}\mathcal{F}(0)\,(\overline{A}, A)\ \& \\ &\ f_1 : \Pi_{R:\mathrm{R}_0}\mathcal{F}(1)\,(\overline{\mathsf{Eq}\,A}, R)\,(f_0(R_{\mathrm{d}}), f_0(R_{\mathrm{c}}))\ \& \\ &\ \Pi_{i:M(0)}\mathcal{F}(0)\big(\overline{\mathsf{id}_{\mathcal{M}(0)}(A)}, i\big)\ f_0(i_{\mathrm{d}}) = f_0(i_{\mathrm{c}})\}\end{aligned}$$

The last condition says that f_0 is functorial in its argument, in the sense that if i is an isomorphism between two types $A, B : \mathrm{Set}_0$, then $f_0(A)$ and $f_0(B)$ are suitably related via the obvious isomorphism between $\mathcal{F}(0)\,(\overline{A}, A)$ and $\mathcal{F}(0)\,(\overline{A}, B)$. This condition, which does not have an analogue in the set-theoretic presentation of Reynolds' model, is needed because we do not work with discrete domains (*e.g.*, we use $\mathcal{F} : \mathcal{M}^n \to \mathcal{M}$ rather than $\mathcal{F} : |\mathcal{M}|^n \to \mathcal{M}$), as is common in other presentations of parametricity. A very similar condition does appear, *e.g.*, in the definition of parametric limits for the category of sets in [2]. The analogous condition asserting the functoriality of f_1 is automatically satisfied since the codomain of f_1 is a proposition. Finally, we define $\forall_n\mathcal{F}(1)\,\overline{R}$ to be the relation with domain $\forall_n\mathcal{F}(0)\,\overline{R_0}$ and codomain $\forall_n\mathcal{F}(0)\,\overline{R_1}$ mapping $((f_0, f_1), (g_0, g_1))$ to $\Pi_{R:\mathrm{R}_0}\mathcal{F}(1)\,(\overline{R}, R)\,(f_0(R_{\mathrm{c}}), g_0(R_{\mathrm{c}}))$.

With a family of adjoints $\forall_n$ in hand, we are almost in a position to use Seely's result [15], which constructs a model of System F from a split $\lambda 2$-fibration. The only missing piece is showing that the adjoints are natural, *i.e.*, that they satisfy the following Beck-Chevalley condition: for any morphism $\mathbf{F} : n \to m$ in $\mathsf{Ctx}(\mathcal{R})$ and object $\mathcal{G} : \mathcal{M}^{m+1} \to \mathcal{M}$, the canonical morphism $\theta_\forall(\mathbf{F}, \mathcal{G})$ from $\mathbf{F}^*(\forall_m(\mathcal{G}))$ to $\forall_n((\mathbf{F} \times \mathsf{id})^*(\mathcal{G}))$ is the identity. But here we hit a snag. In Reynolds' model, the type $\mathcal{F}^*(\forall_1(\mathcal{G}))$ for $\mathcal{F} : 1 \to \mathcal{M}$ and $\mathcal{G} : \mathcal{M}^2 \to \mathcal{M}$ has the form

$$\begin{aligned}\{&f_0 : \Pi_{A:\mathbb{U}}\mathcal{G}(0)\,(\mathcal{F}(0), A)\ \& \\ &f_1 : \Pi_{R:\mathrm{R}_0}\mathcal{G}(1)\,(\mathsf{Eq}\,\mathcal{F}(0), R)\,(f_0(R_{\mathrm{d}}), f_0(R_{\mathrm{c}}))\ \& \ldots\}\end{aligned}$$

whereas the type $\forall_0((\mathcal{F} \times \mathsf{id})^*(\mathcal{G}))$ has the form

$$\begin{aligned}\{&f_0 : \Pi_{A:\mathbb{U}}\mathcal{G}(0)\,(\mathcal{F}(0), A)\ \& \\ &f_1 : \Pi_{R:\mathrm{R}_0}\mathcal{G}(1)\,(\mathcal{F}(1), R)\,(f_0(R_{\mathrm{d}}), f_0(R_{\mathrm{c}}))\ \& \ldots\}\end{aligned}$$

Since $\mathcal{F}$ is degeneracy-preserving, $\mathcal{F}(1)$ is isomorphic to $\mathrm{Eq}\,\mathcal{F}(0)$. But these are not necessarily identical, so $\theta_\forall(\mathcal{F},\mathcal{G})$ is not necessarily an identity, and we cannot directly invoke Seely's result. However, we can still show that $\theta_\forall(\mathcal{F},\mathcal{G})$ is an *isomorphism*, which, together with a few other observations, will be enough to construct a sound model of System F using our main theorem to come. For this, the following notation will be useful. Let $\mathcal{B}$ be a category whose objects are in a bijection with the natural numbers, with $n+1$ serving as a product of n and 1. For $n,k \in \mathbb{N}$ and a morphism $A : n \to 1$, let $\mathsf{q}_n : n \times 1 \to 1$ denote the second projection, let $\mathsf{p}_n(k) : n+k+1 \to n+k$ be the "weakening morphism" that drops the k^{th} variable in the context, counting from the right, and let $\mathsf{s}_n(k,A) : n+k \to n+k+1$ be the "substitution morphism" that substitutes A for the k^{th} variable.

The main obstacle to verifying the equational theory of System F in the more general setting when type formers are allowed to commute with substitution only up to canonical isomorphisms is that the substitution of isomorphic types may yield non-isomorphic results. Consider, for instance, types $\cdot \vdash S$ and $\alpha,\beta \vdash T$. By assumption, weakening is modeled by the weakening functor $\mathsf{p}_n(k)^*$, so $[\![\alpha \vdash S]\!]$ is isomorphic to $\mathsf{p}_0(0)^*\,[\![S]\!]$. Since substitution (of X) is modeled by the substitution functor $\mathsf{s}_n(k,X)^*$, both $\mathsf{s}_1\big(0,[\![\alpha \vdash S]\!]\big)^*\,[\![T]\!]$ and $\mathsf{s}_1\big(0,\mathsf{p}_0(0)^*\,[\![S]\!]\big)^*\,[\![T]\!]$ should model the substitution $T[\beta := S]$, up to isomorphism. But there is no *a priori* reason that these two types *should* be isomorphic: the split generic object ensures that any object A in the fiber over n can be identified with a morphism $A : n \to 1$ but it gives no guarantee that if A and B are isomorphic then A^* and B^* will be naturally isomorphic, or even related in any way whatsoever (the PER model furnishes a counterexample). If X arises as an interpretation of a System F type, then we can construct an isomorphism between $\mathsf{s}_n(k,A)^*\,X$ and $\mathsf{s}_n(k,B)^*\,X$ by induction on the structure of X but, of course, this property does not extend to arbitrary objects.

We solve this problem by *requiring* that, for *relevant* isomorphisms from A to B, the functors $\mathsf{s}_n(k,A)^*$ and $\mathsf{s}_n(k,B)^*$ are naturally isomorphic. To select the relevant isomorphisms, we introduce:

Definition 5.4. A *wide subfibration* of a split fibration $U : \mathcal{E} \to \mathcal{B}$ is a restriction $U' : \mathcal{E}' \to \mathcal{B}$ of U, where $\mathcal{E}'$ is a wide subcategory of $\mathcal{E}$ with the property that, for any object X of $\mathcal{E}$ and $f : Y \to UX$ in $\mathcal{B}$, the cartesian lifting of f with respect to U is cartesian with respect to U'.

Definition 5.5. *A fibration with isomorphisms* is a split fibration $U : \mathcal{E} \to \mathcal{B}$ (the *underlying fibration*), together with a wide subfibration $U' : \mathcal{E}' \to \mathcal{B}$ of U (the *fibration of isomorphisms*), satisfying the following properties:

1. The objects of $\mathcal{B}$ are in bijection with the natural numbers, with 0 serving as a terminal object in $\mathcal{B}$, 1 serving as a split generic object for U, and $n+1$ serving as a product of n and 1.
2. For $n \in \mathbb{N}$, every morphism in $\mathcal{E}'_n$ is an isomorphism.
3. For $n,k \in \mathbb{N}$ and morphism $i : A \to B$ in $\mathcal{E}'_n$, there is a natural transformation $\phi_n(k,i)$ between $\mathsf{s}(k,A)^*$, $\mathsf{s}_n(k,B)^* : \mathcal{E}_{n+k+1} \to \mathcal{E}_{n+k}$ with components in $\mathcal{E}'$ such that:
   a. $\phi_n(0,i)\,\mathsf{q}_n = i$
   b. $\phi_n(k+1,i)\,\mathsf{q}_{n+k+1} = \mathrm{id}_{\mathsf{q}_{n+k}}$
   c. $\phi_n(0,i)\,(\mathsf{p}_n(0)^*\,X) = \mathrm{id}_X$ for every object X in $\mathcal{E}_n$
   d. $\phi_n(k+1,i)\,(\mathsf{p}_{n+k+1}(0)^*\,X) = \mathsf{p}_{n+k}(0)^*\,(\phi_n(k,i)\,X)$ for every object X in $\mathcal{E}_{n+k+1}$

Condition 2 justifies our choice of terminology in Definition 5.5. Conditions 3a through 3d ensure that when X is the interpretation of a System F type, $\phi_n(k,i)\,X$ is precisely the isomorphism determined by induction on the structure of X. Similar conditions are needed any time we impose more structure on the fibration. For example:

Definition 5.6. A *cartesian closed fibration with isomorphisms* is a fibration with isomorphisms such that:

1. For $n \in \mathbb{N}$, the fiber $\mathcal{E}_n$ is cartesian closed, with a terminal object 1_n, products $\times_n$, and exponentials $\Rightarrow_n$, and products and exponentials preserve membership in $\mathcal{E}'$.
2. Beck-Chevalley: for a morphism $f : n \to m$ in $\mathcal{B}$ and objects X, Y in $\mathcal{E}_m$, the canonical morphisms below are in $\mathcal{E}'$:
   $$\theta_1(f) : f^*(1_m) \to 1_n$$
   $$\theta_\times(f,X,Y) : f^*(X \times_m Y) \to \big(f^*(X) \times_n f^*(Y)\big)$$
   $$\theta_\Rightarrow(f,X,Y) : f^*(X \Rightarrow_m Y) \to \big(f^*(X) \Rightarrow_n f^*(Y)\big)$$
3. For $n,k \in \mathbb{N}$, morphism $i : A \to B$ in $\mathcal{E}'_n$, and objects X, Y in $\mathcal{E}_{n+k+1}$, we have
   $$\theta_1(\mathsf{s}_n(k,B)) \circ \phi_n(k,i)(1_{n+k+1}) = \theta_1(\mathsf{s}_n(k,A))$$
   $$\theta_\times\big(\mathsf{s}_n(k,B),X,Y\big) \circ \phi_n(k,i)\,(X \times_{n+k+1} Y) = \big((\phi_n(k,i)\,X) \times_{n+k} (\phi_n(k,i)\,Y)\big) \circ \theta_\times\big(\mathsf{s}_n(k,A),X,Y\big)$$
   $$\theta_\Rightarrow\big(\mathsf{s}_n(k,B),X,Y\big) \circ \phi_n(k,i)\,(X \Rightarrow_{n+k+1} Y) = \big((\phi_n(k,i)\,X)^{-1} \Rightarrow_{n+k} (\phi_n(k,i)\,Y)\big) \circ \theta_\Rightarrow\big(\mathsf{s}_n(k,A),X,Y\big)$$

We can now state the key definition of this section:

Definition 5.7. A *λ2-fibration with isomorphisms* is a cartesian closed fibration with isomorphisms such that:

1. For $n \in \mathbb{N}$, the weakening functor $\mathsf{p}_n(0)^* : \mathcal{E}_n \to \mathcal{E}_{n+1}$ has a right adjoint $\forall_n$, and these adjoints preserve membership in $\mathcal{E}'$.
2. Beck-Chevalley: for a morphism $f : n \to m$ in $\mathcal{B}$ and object X in $\mathcal{E}_{m+1}$, the canonical morphism below is in $\mathcal{E}'$:
   $$\theta_\forall(f,X) : f^*(\forall_m(X)) \to \forall_n((f \times \mathrm{id})^*(X))$$
3. For $n,k \in \mathbb{N}$, morphism $i : A \to B$ in $\mathcal{E}'_n$, and object X in $\mathcal{E}_{n+k+2}$, we have
   $$\theta_\forall\big(\mathsf{s}_n(k,B),X\big) \circ \phi_n(k,i)\,(\forall_{n+k+1}(X)) = \forall_{n+k}(\phi_n(k+1,i)\,X) \circ \theta_\forall\big(\mathsf{s}_n(k,A),X\big)$$

Any split λ2-fibration trivially gives rise to a λ2-fibration with isomorphisms by taking the total category $\mathcal{E}'$ of the fibration of isomorphisms to consist of only the chosen cartesian morphisms in $\mathcal{E}$, which forces every vertical morphism in $\mathcal{E}'$ to be an identity. Our first main result generalizes Seely's [15] well-known one:

Theorem 5.8. *Every λ2-fibration $U : \mathcal{E} \to \mathcal{B}$ with isomorphisms gives a sound model of System F in which:*

- *every type context Γ is interpreted as an object $[\![\Gamma]\!]$ in $\mathcal{B}$*
- *every type $\Gamma \vdash T$ is interpreted as an object $[\![\Gamma \vdash T]\!]$ in the fiber $\mathcal{E}_{[\![\Gamma]\!]}$*
- *every term context $\Gamma;\Delta$ is interpreted as an object $[\![\Gamma \vdash \Delta]\!]$ in the fiber $\mathcal{E}_{[\![\Gamma]\!]}$*
- *every term $\Gamma;\Delta \vdash t : T$ is interpreted as a morphism $[\![\Gamma;\Delta \vdash t : T]\!]$ from $[\![\Gamma;\Delta]\!]$ to $[\![\Gamma \vdash T]\!]$ in the fiber $\mathcal{E}_{[\![\Gamma]\!]}$*

We now want to specify when a model of System F given by a $\lambda 2$-fibration with isomorphisms according to Theorem 5.8 is relationally parametric. For this we will use our second main result, which shows that every cartesian closed reflexive graph category with isomorphisms naturally gives rise to a cartesian closed fibration with isomorphisms.

Theorem 5.9. *Given a cartesian closed reflexive graph category $\mathcal{R}$ with isomorphisms, there is a canonical cartesian closed fibration with isomorphisms whose underlying fibration is the forgetful functor from $\int_n \mathcal{M}^n \to \mathcal{M}$ to* $\mathsf{Ctx}(\mathcal{R})$.

If $\mathcal{R}$ is a cartesian closed reflexive graph category with isomorphisms, we denote by $\mathsf{F}(\mathcal{R})$ the canonical cartesian closed fibration with isomorphisms whose existence is guaranteed by Theorem 5.9. To formulate an abstract definition of a parametric model, we will appropriately relate a $\lambda 2$-fibration with isomorphisms U to $\mathsf{F}(\mathcal{R})$. To see how, we revisit the simplest model, namely the System F term model. In the split $\lambda 2$-fibration U_{term} corresponding to the term model, the fiber over $n \in \mathbb{N}$ consists of types and terms with n free type variables. Let $\mathcal{U}$ be the category consisting of closed System F types and terms between them. Then $\mathcal{U}$ induces a split cartesian closed fibration, U_{set}, whose fiber over n consists of functors $|\mathcal{U}|^n \to \mathcal{U}$ and natural transformations between them.

A type $\overline{\alpha} \vdash T$ with n free variables can now be seen as functor $|\mathcal{U}|^n \to \mathcal{U}$, and a term $\overline{\alpha}; x : S \vdash t : T$ as a natural transformation between S and T. We thus have a morphism of split cartesian closed fibrations $\mu : U_{term} \to U_{set}$. However, unlike U_{term}, U_{set} does not admit the family of adjoints required to make it a $\lambda 2$-fibration. Still, we can view U_{term} as a version of U_{set} that "enriches" the functors and natural transformations with enough extra information to ensure that the desired adjoints exist: in this example, the information that the maps involved are not *ad hoc*, but come from syntax. Since these adjunctions are only applicable to non-empty contexts, no such "enrichment" should be necessary for objects and morphisms over the *terminal* object. And indeed, the restriction of μ to the fibers over the respective terminal objects is clearly an equivalence. These observations echo those immediately following Definition 2.1, and motivate our main definition:

Definition 5.10. Let $\mathcal{R}$ be a cartesian closed reflexive graph category with isomorphisms. A *parametric model of System F over $\mathcal{R}$* is a $\lambda 2$-fibration U with isomorphisms together with a morphism $\mu : U \to \mathsf{F}(\mathcal{R})$ of cartesian closed fibrations with isomorphisms whose restriction to the fibers of U and $\mathsf{F}(\mathcal{R})$ over the terminal objects is full, faithful, and essentially surjective.

Our main theorem shows that the definition of a parametric model is indeed sensible:

Theorem 5.11. *Every parametric model of System F over a cartesian closed reflexive graph category $(\mathcal{X}, (\mathcal{M}, \mathcal{I}))$ with isomorphisms, as specified in Definition 5.10, is a sound model in which:*

- *every type $\Gamma \vdash T$ can be seen as a face map- and degeneracy-preserving reflexive graph functor $[\![\Gamma \vdash T]\!] : \mathcal{M}^{|\Gamma|} \to \mathcal{M}$*
- *every term $\Gamma; \Delta \vdash t : T$ can be seen as a face map- and degeneracy-preserving reflexive graph natural transformation $[\![\Gamma; \Delta \vdash t : T]\!] : [\![\Gamma \vdash \Delta]\!] \to [\![\Gamma \vdash T]\!]$, with the domain and codomain seen as reflexive graph functors into $\mathcal{X}$*

Theorem 5.12 (PER model)**.** *Let $\mathcal{R}_{PER}$ be the cartesian closed reflexive graph category with isomorphisms defined in Examples 2.3, 3.3, and 4.10. The family of adjoints defined in Example 5.2 makes $\mathsf{F}(\mathcal{R}_{PER})$ into a $\lambda 2$-fibration with isomorphisms, and hence into a parametric model of System F over $\mathcal{R}_{PER}$.*

Theorem 5.13 (Reynolds' model)**.** *Let $\mathcal{R}_{REY}$ be the reflexive graph category with isomorphisms defined in Examples 2.4, 3.4, and 4.11. The family of adjoints defined in Example 5.3 makes $\mathsf{F}(\mathcal{R}_{REY})$ into a $\lambda 2$-fibration with isomorphisms, and hence into a parametric model of System F over $\mathcal{R}_{REY}$.*

6 A Proof-Relevant Model of Parametricity

We now describe a proof-relevant version of Reynolds' model, in which witnesses of relatedness are themselves related. The construction of such a model is the subject of [11], but the development there seems to contain a major technical gap. Specifically, it is unclear how to prove the $\forall$-case in Lemma 9.4: when types are interpreted as discrete functors $|\mathcal{X}|^n \to \mathcal{X}$, the reindexing of a degeneracy-preserving functor might not be degeneracy-preserving. We already observed this in the introduction, but this issue is not addressed in [11] and the proof of the lemma is not given there. Since this lemma is crucial to the soundness of the interpretation, it is unknown whether the result of [11] can be salvaged as-is. For this reason, we only reuse the main ideas of [11] for handling the higher dimensional structure and otherwise proceed independently.

Example 6.1. We use the same ambient category as in Example 2.4 and reuse the (internal) category Set of types. The category R of relations is almost the same as in Example 2.4, except that relations are now proof-relevant, *i.e.*, $\mathsf{R}_0 := \Sigma_{A,B:\mathsf{Set}} A \times B \to \mathbb{U}$. Given relations R on A, B and S on C, D, to relate two witnesses $p : R(a,b)$ and $q : S(c,d)$ we should know *a priori* how a relates to c and b to d. This motivates defining the category 2R of *2-relations*, whose objects Q are tuples (Q^0, Q^1, Q^2, Q^3) of relations forming a square

$$\begin{array}{ccc} A & \xrightarrow{Q^0} & B \\ {\scriptstyle Q^1}\downarrow & & \downarrow{\scriptstyle Q^3} \\ C & \xrightarrow[Q^2]{} & D \end{array}$$

together with a Prop-valued predicate (also denoted Q) on the type of tuples of the form $((a,b,c,d),(p,q,r,s))$, where $p : Q^0(a,b)$, $q : Q^1(a,c)$, $r : Q^2(c,d)$, and $s : Q^3(b,d)$. This gives four face maps from 2R to R, one for each edge. We also have four functors in the other direction: *e.g.*, given R, we obtain the 2-relation $\mathsf{Eq}_=(R)$ by placing R on top and bottom, with equalities as vertical edges, and mapping $\big((a,b,a,b),(p,-,r,-)\big)$ to $\mathsf{Id}(p,r)$. Similarly, $\mathsf{Eq}_{\|}(R)$ places R on left and right, $\mathsf{C}_\top(R)$ places R on top and left, and $\mathsf{C}_\bot(R)$ places R on bottom and right, all filling the remaining edges with equalities. The functors $\mathsf{Eq}_=$, $\mathsf{Eq}_{\|}$ are called *degeneracies* and $\mathsf{C}_\top, \mathsf{C}_\bot$ are called *connections*. We define terminal objects, products, exponentials, and isomorphisms in the obvious way.

The above structure induces two cartesian closed fibrations with isomorphisms of interest: the first one is induced by combining the first two levels, the categories Set and R, into a cartesian closed reflexive graph category with isomorphisms R_{PREY}; this is the fibration $\mathsf{F}(\mathsf{R}_{PREY})$. We recall that the objects of $\mathsf{F}(\mathsf{R}_{PREY})$ over n are pairs $\{\mathcal{F}(l) : \mathcal{M}(l)^n \to \mathcal{M}(l)\}_{l \in \{0,1\}}$ of functors that commute with the two face maps from R to Set on the nose, as well as with the degeneracy Eq up to a suitably coherent natural isomorphism. The morphisms are pairs $\{\eta(l) : \mathcal{F}(l) \to \mathcal{G}(l)\}_{l \in \{0,1\}}$ of natural transformations that respect both face maps from R to Set and the degeneracy

Eq. The second fibration, which we call F_{2D}, is induced in much the same way, but taking into account all three levels. This means that the objects over n are triples $\{\mathcal{F}(l) : \mathcal{M}(l)^n \to \mathcal{M}(l)\}_{l \in \{0,1,2\}}$ of functors that commute with *all* face maps — the two from from R to Set as well as the four from 2R to R — on the nose and *all* degeneracies Eq, $\mathsf{Eq}_{=}$, $\mathsf{Eq}_{\|}$ and connections $\mathbf{C}_\top$, $\mathbf{C}_\bot$ up to suitably coherent natural isomorphisms. Analogously, the morphisms are triples $\{\eta(l) : \mathcal{F}(l) \to \mathcal{G}(l)\}_{l \in \{0,1,2\}}$ of natural transformations that respect all face maps, degeneracies, and connections. We have the obvious forgetful morphism of $\lambda^{\to}$-fibrations from F_{2D} to $\mathsf{F}(\mathsf{R}_{PREY})$ that only retains the structure pertaining to levels 0 and 1.

The fibration F_{2D} admits a family of adjoints to weakening functors as follows. The adjoint $\forall_n \mathcal{F}(0)\,\overline{A}$ is the type

$$\begin{aligned}
&\{f_0 : \Pi_{A:\mathbb{U}}\mathcal{F}(0)(\overline{A}, A)\,\& \\
&f_1 : \Pi_{R:\mathsf{R}_0}\mathcal{F}(1)(\overline{\mathsf{Eq}\,A}, R)\,(f_0(R_\mathrm{d}), f_0(R_\mathrm{d}))\,\& \\
&f_2 : \Pi_{Q:2\mathsf{R}_0}\mathcal{F}(2)(\overline{\mathsf{Eq}_{=}(\mathsf{Eq}(A))}, Q)\,\big((f_0\,Q^0_\mathrm{d}, f_0\,Q^0_\mathrm{c}, f_0\,Q^1_\mathrm{c}, f_0\,Q^2_\mathrm{c}), \\
&\qquad\qquad\qquad (f_1\,Q^0, f_1\,Q^1, f_1\,Q^2, f_1\,Q^3)\big)\,\& \\
&\Pi_{i:M(0)}\mathcal{F}(0)(\overline{\mathsf{id}_{\mathcal{M}(0)}(A)}, i)\, f_0(i_\mathrm{c}) = f_0(i_\mathrm{d})\,\& \\
&\Pi_{i:M(1)}\mathcal{F}(1)\big(\overline{\mathsf{id}_{\mathcal{M}(1)}(\mathsf{Eq}\,A)}, i\big)\,\big(f_0\,(i_\mathrm{d})_\mathrm{d}, f_0\,(i_\mathrm{d})_\mathrm{c}\big)\, f_1(i_\mathrm{d}) = f_1(i_\mathrm{c})\}
\end{aligned}$$

In the type of f_2, we could have just as well used any of the other functors $\mathsf{Eq}_{\|}$, $\mathbf{C}_\top$, $\mathbf{C}_\bot$ instead of $\mathsf{Eq}_{=}$ since their compositions with Eq are all naturally isomorphic. We next define $\forall_n \mathcal{F}(1)\,\overline{R}$ to be the relation with domain $\forall_n \mathcal{F}(0)\,\overline{R_\mathrm{d}}$ and codomain $\forall_n \mathcal{F}(0)\,\overline{R_\mathrm{c}}$ mapping $((f_0, f_1, f_2), (g_0, g_1, g_2))$ to

$$\begin{aligned}
&\{\phi : \Pi_{R:\mathsf{R}_0}\mathcal{F}(1)\,(\overline{R}, R)\,(f_0(R_\mathrm{d}), g_0(R_\mathrm{c}))\,\& \\
&\phi_{=} : \Pi_{Q:2\mathsf{R}_0}\mathcal{F}(2)\,(\overline{\mathsf{Eq}_{=}\,R}, Q)\,\big((f_0\,Q^0_\mathrm{d}, f_0\,Q^0_\mathrm{c}, g_0\,Q^1_\mathrm{c}, g_0\,Q^2_\mathrm{c}), \\
&\qquad\qquad\qquad (f_1\,Q^0, \phi\,Q^1, g_1\,Q^2, \phi\,Q^3)\big)\,\&\ldots \\
&\Pi_{i:M(1)}\mathcal{F}(1)(\overline{\mathsf{id}_{\mathcal{M}(1)}(R)}, i)\,\big(f_0\,(i_\mathrm{d})_\mathrm{d}, g_0\,(i_\mathrm{d})_\mathrm{c}\big)\,\phi_1(i_\mathrm{d}) = \phi_1(i_\mathrm{c})\,\}
\end{aligned}$$

The component $\phi_{=}$ asserts that ϕ appropriately interacts with the degeneracy $\mathsf{Eq}_{=}$. The analogous components $\phi_{\|}$, $\phi_{\mathbf{C}_\top}$, $\phi_{\mathbf{C}_\bot}$ for $\mathsf{Eq}_{\|}$, $\mathbf{C}_\top$, $\mathbf{C}_\bot$ are omitted for space reasons.

We define $\forall_n \mathcal{F}(2)\,\overline{Q}$ to be the 2-relation with underlying tuple of relations $\big(\forall_n\mathcal{F}(1)\,\overline{Q^0}, \forall_n \mathcal{F}(1)\,\overline{Q^1}, \forall_n \mathcal{F}(1)\,\overline{Q^2}, \forall_n \mathcal{F}(1)\,\overline{Q^3}\big)$, and mapping a tuple of the form $\big(((f_0,\ldots),(g_0,\ldots),(h_0,\ldots),(l_0,\ldots)), ((\phi_0,\ldots),(\phi_1,\ldots),(\phi_2,\ldots),(\phi_3,\ldots))\big)$ to

$$\begin{aligned}
\Pi_{Q:2\mathsf{R}_0}\mathcal{F}(2)\,(\overline{Q}, Q)\,\big((&f_0\,Q^0_\mathrm{d}, g_0\,Q^0_\mathrm{c}, h_0, Q^1_\mathrm{d}, l_0\,Q^2_\mathrm{c}), \\
&(\phi_0\,Q^0, \phi_1\,Q^1, \phi_2\,Q^2, \phi_3\,Q^3)\big)
\end{aligned}$$

Finally, unlike the frameworks [2, 4, 5, 7, 9, 14], our definition of a parametric model recognizes the above proof-relevant model:

Theorem 6.2 (Proof-relevant model). *The family of adjoints defined in Example 6.1 makes* F_{2D} *into a* $\lambda 2$*-fibration with isomorphisms, and hence into a parametric model of System F over* R_{PREY}.

7 Discussion

We can now be more specific about how our approach compares to the external approaches in [4, 7, 9, 14], all of which are based on a reflexive graph of split $\lambda 2$-fibrations. We already noted that the fibration corresponding to Reynolds' model is not a split $\lambda 2$-fibration, but only a $\lambda 2$-fibration *with isomorphisms*. It thus is not a direct instance of the definitions found in [4, 7, 9, 14]. The same is true for the fibration corresponding to the proof-relevant model, but there we have an even bigger problem: it is unclear how to define the family of adjoints for the second fibration (called r in [7]) of "heterogeneous" reflexive graph functors in a way that is compatible with the adjoint structure on the original fibration. This is because unlike in the proof-irrelevant case, the definition of $\forall_n \mathcal{F}(1)$ now has conditions, such as the one witnessed by $\phi_{=}$, which are only meaningful for "homogeneous" reflexive graph functors, *i.e.*, those where the domain and codomain of $\mathcal{F}(1)(\overline{R})$ are given by the *same* functor $\mathcal{F}(1)$, albeit applied to different arguments ($\overline{R_\mathrm{d}}$ vs. $\overline{R}_\mathrm{c}$).

We indicate three directions for future work. Readers interested with *applications of parametricity* will notice that we do not require conditions such as fullness or (op)cartesianness of certain maps or well-pointedness of certain categories. This follows the spirit of [7], where the notion of *parametricity* pertains to the suitable interaction with (what we call) face maps and degeneracies. Specific *applications* such as establishing the Graph Lemma and the existence of initial algebras are left for another occasion. Readers fond of *type theory* might wonder about possible models expressed in the *intensional* type theory. Although currently there are no well-known models for which the latter would be the right choice of meta-theory, that might change with more research into higher notions of parametricity. Finally, readers familiar with *cubical sets* no doubt recognized the structure of sets, relations, and 2-relations with face maps, degeneracies, and connections from the last section as the first few levels of the cubical hierarchy, and wonder whether one can formulate the analogous notions of 2-parametricity, 3-parametricity, . . . using this hierarchy. We conjecture the answer to be a *YES!* and plan to pursue this question in future work.

Acknowledgments This research is supported by NSF awards 1420175 and 1545197. We thank Steve Awodey and Peter Dybjer for helpful discussions.

References

[1] R. Atkey, N. Ghani, and Patricia Johann. 2014. A Relationally Parametric Model of Dependent Type Theory. In *Principles of Programming Languages*. 503–516.
[2] B. Dunphy and U. Reddy. 2004. Parametric Limits. In *Logic in Computer Science*. 242–251.
[3] N. Ghani, F. Nordvall Forsberg, and F. Orsanigo. 2016. Proof-Relevant Parametricity. In *A List of Successes That Can Change the World - Essays Dedicated to Philip Wadler on the Occasion of His 60th Birthday (Lecture Notes in Computer Science 9600)*. 109–131.
[4] N. Ghani, F. Nordvall Forsberg, and A. Simpson. 2016. Comprehensive Parametric Polymorphism: Categorical Models and Type Theory. In *Foundations of Software Science and Computation Structures*. 3–19.
[5] N. Ghani, P. Johann, F. Nordvall Forsberg, F. Orsanigo, and T. Revell. 2015. Bifibrational Functorial Semantics for Parametric Polymorphism. In *Mathematical Foundations of Program Semantics*. 165–181.
[6] J.-Y. Girard, P. Taylor, and Y. Lafont. 1989. *Proofs and Types*. Cambridge University Press.
[7] B. Jacobs. 1999. *Categorical Logic and Type Theory*. Elsevier.
[8] G. Longo and E. Moggi. 2009. Constructive Natural Deduction and its ω-set Interpretation. *Mathematical Structures in Computer Science* 1(2) (2009), 215–254.
[9] Q. Ma and J. Reynolds. 1992. Types, Abstraction, and Parametric Polymorphism. In *Mathematical Foundations of Program Semantics*. 1–40.
[10] P. W. O'Hearn and R. D. Tennent. 1995. Parametricity and Local Variables. *J. ACM* 42(3) (1995), 658–709.
[11] F. Orsanigo. 2017. *Bifibrational Parametricity: From Zero to Two Dimensions*. Ph.D. Dissertation. University of Strathclyde.
[12] J. Reynolds. 1983. Types, Abstraction, and Parametric Polymorphism. In *International Federation for Information Processing*. 513–523.
[13] J. C. Reynolds. 1984. Polymorphism is not set-theoretic. *Semantics of Data Types* (1984), 145–156.
[14] E. Robinson and G. Rosolini. 1994. Reflexive Graphs and Parametric Polymorphism. In *Logic in Computer Science*. 364–371.
[15] R. A. G. Seely. 1987. Categorical Semantics for Higher Order Polymorphic Lambda Calculus. *Journal of Symbolic Logic* 52 (1987), 969–989.
[16] P. Wadler. 1987. Theorems for Free!. In *Functional Programming and Computer Architecture*. 347–359.