Comprehensive Parametric Polymorphism: Categorical Models and Type Theory

Neil Ghani[1], Fredrik Nordvall Forsberg[1], and Alex Simpson[2]

[1] Department of Computer and Information Sciences, University of Strathclyde
[2] Faculty of Mathematics and Physics, University of Ljubljana

Abstract. This paper combines reflexive-graph-category structure for relational parametricity with fibrational models of impredicative polymorphism. To achieve this, we modify the definition of fibrational model of impredicative polymorphism by adding one further ingredient to the structure: *comprehension* in the sense of Lawvere. Our main result is that such *comprehensive* models, once further endowed with reflexive-graph-category structure, enjoy the expected consequences of parametricity. This is proved using a type-theoretic presentation of the category-theoretic structure, within which the desired consequences of parametricity are derived. The formalisation requires new techniques because equality relations are not available, and standard arguments that exploit equality need to be reworked.

1 Introduction

According to Strachey [26], a polymorphic program is *parametric* if it applies the same uniform algorithm at all instantiations of its type parameters. Reynolds [23] proposed *relational parametricity* as a mathematical model of parametric polymorphism. Relational parametricity is a powerful mathematical tool with many useful consequences; see [27,21,13] for numerous examples.

The polymorphic lambda-calculus, $\lambda\mathbf{2}$, (a.k.a. System F) was introduced independently by Girard [11] and Reynolds [22]. It serves as a model type theory for (impredicative) polymorphism, and thus provides a significant testing ground for ideas on relational parametricity. In this paper we address the question:

> *What is the fundamental category-theoretic structure needed to model relational parametricity for* $\lambda\mathbf{2}$*, which is both i) minimal, in assuming as little structure as possible; but ii) strong enough to ensure the expected consequences of parametricity hold?*

It is perhaps surprising that this question does not yet have an established answer. On the one hand, category-theoretic models for $\lambda\mathbf{2}$ were developed many years ago by Seely [25]. They are studied systematically as $\lambda\mathbf{2}$ *fibrations* in Jacobs [15]. On the other, the fundamental category-theoretic structure needed to model relational parametricity is also known. The crucial ingredient is the notion of *reflexive graph category* which appeared implicitly in Ma and Reynolds [19], was used explicitly by O'Hearn and Tennent [20], and Robinson and Rosolini [24], and reached maturity in the *parametricity graphs* of Dunphy and Reddy [7,8].

To obtain minimal structure for relational parametricity for $\lambda\mathbf{2}$, it is natural to combine the structure of $\lambda\mathbf{2}$ fibrations with that of parametricity graphs. This results in the notion of $\lambda\mathbf{2}$ *parametricity graph*, which we define in Section 3. Sadly, $\lambda\mathbf{2}$ parametricity graphs enjoy the expected properties of parametricity only in the special case that the underlying category is *well-pointed*. (Similar observations, for different but related notions of model, are made in [6,7,8].) Since well-pointedness rules out many categories of interest in semantics (e.g., functor categories) this limits the generality of the theory.

One way of circumventing the restriction to well-pointed categories was proposed by Birkedal and Møgelberg [6], who developed a more elaborate category-theoretic structure, which overcomes the limitation by modelling Plotkin and Abadi's logic for parametricity [21]. This method of modelling the combination of $\lambda\mathbf{2}$ with an extraneous logic has been been refined and simplified by Hermida [12]. Nonetheless, it does not enjoy the simplicity in conception of combining the structure of category-theoretic models of $\lambda\mathbf{2}$ with that of parametricity graphs.

To obtain our minimal structure, we retain the original idea of combining parametricity graphs with category-theoretic models of $\lambda\mathbf{2}$. However, we implement this in a perhaps unexpected way. *We modify the notion of $\lambda\mathbf{2}$ model.* We ask for $\lambda\mathbf{2}$ fibrations to additionally satisfy Lawvere's *comprehension* property. Not only are the resulting *comprehensive $\lambda\mathbf{2}$ fibrations* natural in their own right as models of $\lambda\mathbf{2}$, but, when combined with parametricity-graph structure to form *comprehensive $\lambda\mathbf{2}$ parametricity graphs*, they do indeed enjoy all expected consequences of parametricity.

Sections 2 and 3 define comprehensive $\lambda\mathbf{2}$ fibrations and comprehensive $\lambda\mathbf{2}$ parametricity graphs respectively. In Section 4, we present a type theory $\lambda\mathbf{2R}$, corresponding to our category-theoretic structure, which provides a simple system for reasoning about parametricity. The type theory $\lambda\mathbf{2R}$ is similar to Dunphy's System P [7], and Abadi, Cardelli and Curien's System R [1], to which it is compared in Section 7.

In Section 5, we develop the technical machinery needed to reason in $\lambda\mathbf{2R}$. A key obstacle is that the system does not include *equality relations*. This means that graph relations, which are a crucial ingredient in standard arguments involving relational parametricity, are not in general definable. In Section 5, we instead identify two forms of *pseudograph* relations, whose subtle interrelationship allows us to establish the consequences we need. One kind of pseudograph relation is immediately definable using the *fibrational* structure built into the notion of parametricity graph. The other type of pseudograph requires *opfibrational* structure. We use an impredicative encoding to show that opfibrational structure is definable in $\lambda\mathbf{2R}$, and hence always present in comprehensive $\lambda\mathbf{2}$ parametricity graphs. In Section 6, we finally apply the technical machinery and establish that the expected consequences of relational parametricity are indeed derivable in $\lambda\mathbf{2R}$, and hence hold in comprehensive $\lambda\mathbf{2}$ parametricity graphs.

In summary, the main contributions of this work are:

(i) The definition of *comprehensive* $\lambda\mathbf{2}$ *fibrations* as models of $\lambda\mathbf{2}$.
(ii) The definition of a new category-theoretic notion of model of relational parametricity, obtained by combining parametricity graphs and comprehensive $\lambda\mathbf{2}$ fibrations into *comprehensive* $\lambda\mathbf{2}$ *parametricity graphs*.
(iii) The extraction of $\lambda\mathbf{2R}$ as the type theory intrinsic to comprehensive $\lambda\mathbf{2}$ parametricity graphs.
(iv) The derivation of the expected consequences of parametricity in $\lambda\mathbf{2R}$, and hence in comprehensive $\lambda\mathbf{2}$ parametricity graphs. This requires novel techniques: establishing the *opfibration* property of comprehensive $\lambda\mathbf{2}$ parametricity graphs, and the use of *pseudograph* relations.

In the category-theoretic parts of the paper, we assume familiarity with fibred category theory, for which Jacobs [15] is our main reference. Nevertheless, a substantial portion of the paper is presented in purely type-theoretic terms, and may be read without reference to the accompanying category-theoretic material.

2 Comprehensive $\boldsymbol{\lambda}$2 Fibrations

In Fig. 1, we recall the polymorphic λ-calculus $\lambda\mathbf{2}$. We use $x, y, \ldots$ to range over term variables, and $\alpha, \beta, \ldots$ to range over type variables. Our presentation has four judgements: $\boxed{\Gamma\ \mathsf{ctxt}}$, stating that Γ is a well-formed context; $\boxed{\Gamma \vdash A\ \mathsf{type}}$, stating that A is a well-formed type in context Γ; $\boxed{\Gamma \vdash t : A}$, stating that the term t has type A in context Γ; and judgemental equality $\boxed{\Gamma \vdash t_1 = t_2 : A}$. We assume β and η-equalities for both term abstraction, λ, and type abstraction, Λ. Equality is also assumed to be a congruence relation, although the rules guaranteeing this have been omitted from Fig. 1 for brevity.

A minor departure from many presentations of $\lambda\mathbf{2}$ is that we interleave type variables and term variables in a single context. This approach is not only natural, but indeed standard when $\lambda\mathbf{2}$ is considered in the context of dependent type theory; for example, when derived as an instance of a pure type system [3]. Since there is no dependency of $\lambda\mathbf{2}$ types on term variables, such interleaving is syntactically vacuous. Nevertheless, we shall see below that its presence does have semantic implications.

We next recall the standard category-theoretic notion of $\lambda\mathbf{2}$ *fibration*, which models $\lambda\mathbf{2}$. We directly restrict the definition to the *split* case to circumvent coherence issues that would otherwise arise, cf. [15].

Definition 1 (λ2 fibration). A $\lambda\mathbf{2}$ *fibration* is a split fibration $p : \mathbb{T} \to \mathbb{C}$, where the base category $\mathbb{C}$ has finite products, and the fibration:

(i) is fibred cartesian closed;
(ii) has a split generic object U [15, Def. 5.2.1] — we write Ω for pU;
(iii) and has fibred-products along projections $X \times \Omega \longrightarrow X$ in $\mathbb{C}$.

Moreover, the reindexing functors given by the splitting are required to preserve the above-specified structure in fibres on the nose.

Context formation rules:

$$\frac{}{\cdot\ \mathsf{ctxt}} \qquad \frac{\Gamma\ \mathsf{ctxt}}{\Gamma,\ \alpha\ \mathsf{ctxt}}\ (\alpha \notin \Gamma) \qquad \frac{\Gamma\ \mathsf{ctxt} \qquad \Gamma \vdash A\ \mathsf{type}}{\Gamma,\ x : A\ \mathsf{ctxt}}\ (x \notin \Gamma)$$

Type formation rules:

$$\frac{}{\Gamma \vdash \alpha\ \mathsf{type}}\ (\alpha \in \Gamma) \qquad \frac{\Gamma \vdash A\ \mathsf{type} \qquad \Gamma \vdash B\ \mathsf{type}}{\Gamma \vdash A \to B\ \mathsf{type}} \qquad \frac{\Gamma,\ \alpha \vdash A\ \mathsf{type}}{\Gamma \vdash \forall\alpha.\, A\ \mathsf{type}}$$

Term typing rules:

$$\frac{}{\Gamma \vdash x : A}\ (x : A \in \Gamma) \qquad \frac{\Gamma,\ x : A \vdash t : B}{\Gamma \vdash \lambda x.\, t : A \to B} \qquad \frac{\Gamma \vdash s : A \to B \qquad \Gamma \vdash t : A}{\Gamma \vdash s\, t : B}$$

$$\frac{\Gamma,\ \alpha \vdash t : A}{\Gamma \vdash \Lambda\alpha.\, t : \forall\alpha.\, A} \qquad \frac{\Gamma \vdash t : \forall\alpha.\, A \qquad \Gamma \vdash B\ \mathsf{type}}{\Gamma \vdash t[B] : A[\alpha \mapsto B]}$$

Judgemental equality:

$$\frac{}{\Gamma \vdash (\lambda x.\, t)u = t[x \mapsto u] : B} \qquad \frac{}{\Gamma \vdash t = \lambda x.\, tx : A \to B}\ (x \notin \Gamma)$$

$$\frac{}{\Gamma \vdash (\Lambda\alpha.\, t)[U] = t[\alpha \mapsto U] : A[\alpha \mapsto U]} \qquad \frac{}{\Gamma \vdash t = \Lambda\alpha.\, t[\alpha] : \forall\alpha.\, A}\ (\alpha \notin \Gamma)$$

Fig. 1: The type system $\lambda\mathbf{2}$

The above definition differs slightly from [15, Def. 8.4.3(b)] in that we do not include fibred coproducts in condition (iii). These are not needed to model $\lambda\mathbf{2}$, and their existence is anyway derivable in parametric models.

In a $\lambda\mathbf{2}$ fibration, we write $\mathbb{T}_X$ for the fibre category over X. We also use X as a subscript when referring to structure in $\mathbb{T}_X$; e.g., $\mathbf{1}_X$ is the specified terminal object in $\mathbb{T}_X$, and $\Rightarrow_X$ is the exponential structure in $\mathbb{T}_X$. Given $f \colon X \longrightarrow Y$ in $\mathbb{C}$, we write f^* for the reindexing functor $\mathbb{T}_Y \to \mathbb{T}_X$, and $A^*f \colon f^*A \longrightarrow A$ for the specified cartesian lifting of f relative to A. We also write $\prod_\Omega$ for the specified right adjoint, given by (iii), to reindexing functors $\pi_1^* : \mathbb{T}_X \to \mathbb{T}_{X\times\Omega}$.

We recall in outline the semantic interpretation of $\lambda\mathbf{2}$ in a $\lambda\mathbf{2}$ fibration $\mathbb{T} \to \mathbb{C}$. A context $\Theta = \alpha_1, \dots, \alpha_n$ of type variables is interpreted as the n-fold product $[\![\Theta]\!] = \Omega^n$ in $\mathbb{C}$. A type A in type-variable context Θ is then interpreted as an object $[\![A]\!]_\Theta$ of $\mathbb{T}$ over $[\![\Theta]\!]$, defined by induction on the structure of A, using cartesian closure for function types, fibred products for universal types, and the reindexing $(\pi_i)^* U$ of the generic object along the projection $\pi_i \colon \Omega^n \longrightarrow \Omega$ to interpret α_i over $[\![\Theta]\!]$. Finally, the interpretation of a term $\Gamma \vdash t : A$ is obtained by splitting Γ into its component contexts: Θ of type variables, and Δ of term variables. Then $\Delta = x_1 : A_1, \dots, x_m : A_m$ is interpreted as the product $[\![\Delta]\!]_\Theta = [\![A_1]\!]_\Theta \times \cdots \times [\![A_m]\!]_\Theta$ in the fibre over $[\![\Theta]\!]$, and t is interpreted as a morphism $[\![t]\!]_\Gamma \colon [\![\Delta]\!]_\Theta \longrightarrow [\![A]\!]_\Theta$ in $\mathbb{T}_{[\![\Theta]\!]}$.

In the above outline, one sees that the structure of a $\lambda\mathbf{2}$ fibration fits uneasily alongside our mixed contexts of interleaved type and term variables, since these have to be separated to define the semantic interpretation. In *dependent* type theory, where no such separation is possible, a more direct semantic interpretation is achieved using Lawvere's *comprehension* property [18] to model the process of context extension [14]. It is natural to apply the same idea to $\lambda\mathbf{2}$.

Definition 2 (Comprehensive $\lambda\mathbf{2}$ fibration). A $\lambda\mathbf{2}$ fibration $p : \mathbb{T} \to \mathbb{C}$ is *comprehensive* if it enjoys the *comprehension* property [15, Def. 10.4.7]: the terminal-object functor $X \mapsto \mathbf{1}_X : \mathbb{C} \to \mathbb{T}$ has a specified right adjoint $K : \mathbb{T} \to \mathbb{C}$.

Requiring a *specified* right adjoint maintains consistency with our policy of working with split fibrational structure. Given A in $\mathbb{T}_X$, we write $\kappa_A : KA \longrightarrow X$ for the 'projection' map obtained by applying p to the counit $\mathbf{1}_{KA} \longrightarrow A$ in $\mathbb{T}$.

To show that comprehensive $\lambda\mathbf{2}$ fibrations permit a direct, inductive-on-syntax semantic interpretation, we present the interpretation of $\lambda\mathbf{2}$ types in detail. A context Γ ctxt is interpreted as an object $[\![\Gamma]\!]$ of $\mathbb{C}$; and a type $\Gamma \vdash A$ type is interpreted as an object $[\![A]\!]_\Gamma$ in $\mathbb{T}_{[\![\Gamma]\!]}$. These are defined by mutual induction, together with maps $\pi^\alpha_\Gamma \colon [\![\Gamma]\!] \longrightarrow \Omega$ for every context Γ containing α.

$$
\begin{array}{lll}
[\![\cdot]\!] = \mathbf{1} & [\![\alpha]\!]_\Gamma = (\pi^\alpha_\Gamma)^*\, U & \pi^\alpha_{\Gamma,\,\alpha} = \pi_2 \\
[\![\Gamma,\, \alpha]\!] = [\![\Gamma]\!] \times \Omega & [\![A \to B]\!]_\Gamma = [\![A]\!]_\Gamma \Rightarrow_{[\![\Gamma]\!]} [\![B]\!]_\Gamma & \pi^\alpha_{\Gamma,\,\beta} = \pi^\alpha_\Gamma \circ \pi_1 \;(\beta \neq \alpha) \\
[\![\Gamma,\, x : A]\!] = K[\![A]\!]_\Gamma & [\![\forall\alpha.\, A]\!] = \prod_\Omega [\![A]\!]_{\Gamma,\,\alpha} & \pi^\alpha_{\Gamma,\,x:A} = \pi^\alpha_\Gamma \circ \kappa_{[\![A]\!]_\Gamma}
\end{array}
$$

Having made the above definitions, a term $\Gamma \vdash t : A$ is interpreted as a global element $[\![t]\!]_\Gamma \colon \mathbf{1}_{[\![\Gamma]\!]} \longrightarrow [\![A]\!]_\Gamma$ in $\mathbb{T}_{[\![\Gamma]\!]}$. The definition, which we omit, is a straightforward induction on the derivation of $\Gamma \vdash t : A$.

The appropriateness of comprehensive $\lambda\mathbf{2}$ fibrations as a notion of model for $\lambda\mathbf{2}$ is supported by soundness and completeness results.

Theorem 3 (Soundness for $\lambda\mathbf{2}$). *If $\Gamma \vdash t_1 = t_2 : A$ then, in every comprehensive $\lambda\mathbf{2}$ fibration, we have $[\![t_1]\!]_\Gamma = [\![t_2]\!]_\Gamma$.*

Theorem 4 (Full completeness for $\lambda\mathbf{2}$). *There exists a comprehensive $\lambda\mathbf{2}$ fibration satisfying:*

(i) for every type $\Gamma \vdash A$ type, every global point $\mathbf{1}_{[\![\Gamma]\!]} \longrightarrow [\![A]\!]_\Gamma$ is the denotation $[\![t]\!]_\Gamma$ of some term $\Gamma \vdash t : A$; and
(ii) for all terms $\Gamma \vdash t_1, t_2 : A$ satisfying $[\![t_1]\!]_\Gamma = [\![t_2]\!]_\Gamma$, we have $\Gamma \vdash t_1 = t_2 : A$.

Theorem 3 is proved by a routine induction on equality derivations, and Theorem 4 by construction of a syntactic model, which has the requisite properties.

3 Comprehensive $\lambda\mathbf{2}$ Parametricity Graphs

Reflexive graph categories are studied in [19,24,16,7,8] as a simple category-theoretic structure for modelling relational parametricity. A *reflexive graph* *category* consists of a pair of categories, $\mathbb{V}$, the *vertex category*, and $\mathbb{E}$, the *edge category*, together with functors $\nabla_1, \nabla_2 \colon \mathbb{E} \to \mathbb{V}$ and $\Delta \colon \mathbb{V} \to \mathbb{E}$ satisfying $\nabla_1 \Delta = \mathrm{id}_{\mathbb{V}} = \nabla_2 \Delta$. Informally, one thinks of $\mathbb{E}$ as a category whose objects are binary 'relations' between objects of $\mathbb{V}$. Then ∇_1, ∇_2 are 'projection' functors, and Δ maps an object to its 'identity relation'.

We shall be guided by the following general thesis. *A model of relational parametricity, irrespective of the type theory for which it is considered, should form a reflexive graph category, in the (2-)category of structure-preserving functors between models of the type theory in question.* This thesis is supported by the following considerations. Endowing the edge category $\mathbb{E}$ with the categorical structure needed to interpret types corresponds to giving types a *relational interpretation*. The preservation of this structure by the projection functors ∇_1, ∇_2 means that the relational interpretation commutes with the usual non-relational interpretation of types. The preservation of structure by Δ, in combination with the identity property discussed later, corresponds to Reynolds' *identity extension* property [23].

In the context of the present paper, we need to specialise the above recipe to (comprehensive) $\lambda\mathbf{2}$ fibrations. A *morphism* from one (comprehensive) $\lambda\mathbf{2}$ fibration $p' \colon \mathbb{T}' \to \mathbb{C}'$ to another $p \colon \mathbb{T} \to \mathbb{C}$ is given by a pair of functors, $H \colon \mathbb{T}' \to \mathbb{T}$ and $L \colon \mathbb{C}' \to \mathbb{C}$ such that $p\,H = L\,p'$, and such that H, L preserve all other specified structure (including the choice of cartesian morphisms in the splitting) on the nose. By a *reflexive graph of (comprehensive)* $\lambda\mathbf{2}$ *fibrations*, we thus mean a pair of (comprehensive) $\lambda\mathbf{2}$ fibrations with functors between them:

$$\begin{tikzcd}
\mathcal{R}(\mathbb{T}) \arrow[r, shift left=2, "{\nabla_1^{\mathbb{T}},\ \Delta^{\mathbb{T}},\ \nabla_2^{\mathbb{T}}}"] \arrow[r, shift right=2] \arrow[d, "p^{\mathcal{R}}"'] & \mathbb{T} \arrow[l] \arrow[d, "p"] \\
\mathcal{R}(\mathbb{C}) \arrow[r, shift left=2] \arrow[r, shift right=2, "{\nabla_1^{\mathbb{C}},\ \Delta^{\mathbb{C}},\ \nabla_2^{\mathbb{C}}}"'] & \mathbb{C} \arrow[l]
\end{tikzcd} \tag{1}$$

where each of the three pairs $\nabla_1^{\mathbb{T}}, \nabla_1^{\mathbb{C}}$ and $\nabla_2^{\mathbb{T}}, \nabla_2^{\mathbb{C}}$ and $\Delta^{\mathbb{T}}, \Delta^{\mathbb{C}}$ is a morphism of (comprehensive) $\lambda\mathbf{2}$ fibrations, and where each of the triples $\nabla_1^{\mathbb{T}}, \nabla_2^{\mathbb{T}}, \Delta^{\mathbb{T}}$ and $\nabla_1^{\mathbb{C}}, \nabla_2^{\mathbb{C}}, \Delta^{\mathbb{C}}$ is a reflexive graph category. We emphasise that $p^{\mathcal{R}} \colon \mathcal{R}(\mathbb{T}) \to \mathcal{R}(\mathbb{C})$, in (1), is an arbitrary (comprehensive) $\lambda\mathbf{2}$ fibration fitting into the diagram. The notation $\mathcal{R}(\cdot)$ is merely mnemonic, and does not imply that $\mathcal{R}(\mathbb{T})$ is obtained using a particular construction from $\mathbb{T}$.

One needs to add further conditions to the above structure to ensure that the objects of $\mathcal{R}(\mathbb{T})$ behave sufficiently like relations. In [19], this was addressed by requiring the fibre category $\mathcal{R}(\mathbb{T})_{\mathbf{1}_{\mathcal{R}(\mathbb{C})}}$, over the terminal object, to coincide with a particular category of *logical relations* over $\mathbb{T}_{\mathbf{1}_{\mathbb{C}}}$. As well as only being applicable if $\mathbb{T}_{\mathbf{1}_{\mathbb{C}}}$ has (sufficient) finite limits, this requirement also has the weakness that it says nothing about other fibres of $p^{\mathcal{R}}$. As a result, the structure is too weak to imply consequences of parametricity in general, see [24,6] for discussion. To remedy this, we instead need axiomatic structure for a category of relations, in a form that is suitable for being imposed fibrewise on $p^{\mathcal{R}}$. This is provided by Dunphy and Reddy's notion of *parametricity graph* [7,8], which we now recall.

A reflexive graph category $\nabla_1, \nabla_2 \colon \mathbb{E} \to \mathbb{V},\ \Delta \colon \mathbb{V} \to \mathbb{E}$ is said to be *relational* if the functor $\langle \nabla_1, \nabla_2 \rangle \colon \mathbb{E} \to \mathbb{V} \times \mathbb{V}$ is faithful. This property allows one to think of morphisms in $\mathbb{E}$ as pairs of *relation-preserving* maps from $\mathbb{V}$. Accordingly, we call objects of $\mathbb{E}$ *relations*, we write $R \colon A \leftrightarrow B$ to mean an object R of $\mathbb{E}$ with $\nabla_1 R = A$ and $\nabla_2 R = B$, and we write $f \times g \colon R \longrightarrow S$ to mean that there is a (necessarily unique) map $h \colon R \longrightarrow S$ in $\mathbb{E}$ with $\nabla_1 h = f$ and $\nabla_2 h = g$. A reflexive graph category satisfies the *identity property* if, for every $h \colon \Delta A \longrightarrow \Delta B$ in $\mathbb{E}$, it holds that $\nabla_1 h = \nabla_2 h$. This allows one to think of ΔA as an identity relation on A (although, cf. Sect. 5 for caveats). In a relational reflexive graph category, the identity property is equivalent to the fullness of the functor Δ. A *parametricity graph* is a relational reflexive graph category with the identity property, for which the functor $\langle \nabla_1, \nabla_2 \rangle \colon \mathbb{E} \to \mathbb{V} \times \mathbb{V}$ is a fibration. The fibration property supports the following definition mechanism. Let $R : A \leftrightarrow B$ be a relation in $\mathbb{E}$. Then, given morphisms $f \colon A' \longrightarrow A$ and $g \colon B' \longrightarrow B$ in $\mathbb{V}$, reindexing produces an *inverse image* relation $[f \times g]^{-1} R \colon A' \leftrightarrow B'$.[3]

The main category-theoretic definition of this paper is a fibrewise adaptation of parametricity graph to the context of comprehensive $\lambda\mathbf{2}$ fibrations.

Definition 5 ((Comprehensive) $\lambda\mathbf{2}$ parametricity graph). A *(comprehensive) $\lambda\mathbf{2}$ parametricity graph* is a reflexive graph of (comprehensive) $\lambda\mathbf{2}$ fibrations, as in (1), that satisfies, for all objects W of $\mathcal{R}(\mathbb{C})$ and X of $\mathbb{C}$:

(Relational) The functor $\langle \nabla_1^{\mathbb{T}}, \nabla_2^{\mathbb{T}} \rangle \restriction_{\mathcal{R}(\mathbb{T})_W} : \mathcal{R}(\mathbb{T})_W \to \mathbb{T}_{\nabla_1^{\mathbb{C}} W} \times \mathbb{T}_{\nabla_2^{\mathbb{C}} W}$ is faithful.
(Identity property) The functor $\Delta^{\mathbb{T}} \restriction_{\mathbb{T}_X} : \mathbb{T}_X \to \mathcal{R}(\mathbb{T})_{\Delta^{\mathbb{C}} X}$ is full.
(Fibration) $\langle \nabla_1^{\mathbb{T}}, \nabla_2^{\mathbb{T}} \rangle \restriction_{\mathcal{R}(\mathbb{T})_W} : \mathcal{R}(\mathbb{T})_W \to \mathbb{T}_{\nabla_1^{\mathbb{C}} W} \times \mathbb{T}_{\nabla_2^{\mathbb{C}} W}$ is a cloven fibration.

Moreover, for every $\phi : W' \longrightarrow W$ in $\mathcal{R}(\mathbb{C})$, we require the commuting square

$$\begin{tikzcd}
\mathcal{R}(\mathbb{T})_W \arrow[r, "(p^{\mathcal{R}})^* \phi"] \arrow[d, "\langle \nabla_1^{\mathbb{T}}, \nabla_2^{\mathbb{T}} \rangle \restriction_{\mathcal{R}(\mathbb{T})_W}"'] & \mathcal{R}(\mathbb{T})_{W'} \arrow[d, "\langle \nabla_1^{\mathbb{T}}, \nabla_2^{\mathbb{T}} \rangle \restriction_{\mathcal{R}(\mathbb{T})_{W'}}"] \\
\mathbb{T}_{\nabla_1^{\mathbb{C}} W} \times \mathbb{T}_{\nabla_2^{\mathbb{C}} W} \arrow[r, "p^*(\nabla_1^{\mathbb{C}} \phi) \times p^*(\nabla_2^{\mathbb{C}} \phi)"'] & \mathbb{T}_{\nabla_1^{\mathbb{C}} W'} \times \mathbb{T}_{\nabla_2^{\mathbb{C}} W'}
\end{tikzcd}$$

(where the notation distinguishes reindexing functors determined by p and $p^{\mathcal{R}}$) to give a cleavage-preserving fibred functor from $\langle \nabla_1^{\mathbb{T}}, \nabla_2^{\mathbb{T}} \rangle \restriction_{\mathcal{R}(\mathbb{T})_W}$ to $\langle \nabla_1^{\mathbb{T}}, \nabla_2^{\mathbb{T}} \rangle \restriction_{\mathcal{R}(\mathbb{T})_{W'}}$.

This definition could by strengthened by asking for the parametricity-graph fibrations to be split instead of merely cloven. Such a strengthening does not

[3] We use $(\cdot)^{-1}$ rather than $(\cdot)^*$ for reindexing to emphasise that we are in a relational setting: $\langle \nabla_1, \nabla_2 \rangle$ is a preorder fibration since it is faithful.

affect any of the results in the sequel, and may seem natural given our use of split fibrations in all previous definitions. Nevertheless, our choice of definition reflects the fact that the weaker cloven assumption is all that is needed to avoid coherence issues arising in the semantic interpretation of the type theory $\lambda\mathbf{2R}$ introduced in Sect. 4 below.

It is Def. 5, with the comprehension property included, that provides our answer to the question highlighted in the introduction. (The definition without comprehension is included for comparison purposes only.)

4 A Type System for Relational Reasoning

We define a type system $\lambda\mathbf{2R}$, suggested by the structure of comprehensive $\lambda\mathbf{2}$ parametricity graphs. This system is similar, in many respects, to System R of Abadi, Cardelli and Curien [1] and System P of Dunphy [7], to which we shall compare it in Sect. 7.

The rules for $\lambda\mathbf{2R}$ are given by Fig. 1 (it extends $\lambda\mathbf{2}$) in combination with Fig. 2. The latter adds three new judgements: $\boxed{\Theta\ \mathsf{rctxt}}$ says that Θ is a well-defined *relational context*; $\boxed{\Theta \vdash A_1 R A_2\ \mathsf{rel}}$ says that R is a *relation* between types A_1 and A_2, in relational context Θ; and $\boxed{\Theta \vdash (t_1 : A_1) R (t_2 : A_2)}$ is a *relatedness judgement*, asserting that $t_1 : A_1$ is related to $t_2 : A_2$ by the relation R.

Relations, in Fig. 2, are built up from a collection of *relation variables* $\rho, \dots$, which, for clarity, we choose to keep disjoint from type and term variables. In the rules, we make use of three operations $(\cdot)_1$, $(\cdot)_2$ and $\langle\cdot\rangle$, defined in Fig. 3, which implement reflexive graph structure on syntax. The $(\cdot)_i$ operations project a relational context to a typing context, whereas the $\langle\cdot\rangle$ operation acts in the other direction. In the definition of the latter, we associate a distinct relation variable ρ^α to every type variable α. Lemma 7 below states how these operations relate typing and relational judgements.

The rules for building relational contexts and relations, in Fig. 2, require some explanation. In adding an assertion $\alpha\rho\beta$ to a relational context Θ, all variables α, β, ρ need to be sufficiently fresh. However, the formulation of $\lambda\mathbf{2R}$ is such that variables on the left-hand side of relations are always manipulated separately from variables on the right. Thus, for example, α is sufficiently fresh in $\alpha\rho\beta$, as long as α does not already occur on the left side $(\Theta)_1$ of Θ. A similar separation principle applies also with respect to the term variables x_1, x_2 in assertions $(x_1 : A_1) R (x_2 : A_2)$. The separation principle means that one needs to be cautious in interpreting assertions of the form $\alpha\rho\alpha$ and $(x : A) R (x : A)$. In such assertions, even though the same variable appears on the left and right, the correct intuition is that these are really two distinct variables. We have chosen not to underline this distinction by requiring the variables to be syntactically different, since doing so would add unnecessary syntactic clutter to the system; for example, it would complicate the definition of the $\langle\cdot\rangle$ operation. Instead, we rely on left and right positioning to make the necessary distinctions. This is crucial in the definition of the substitution operations on relations. There are

Relational context formation rules:

$$\frac{}{\cdot\ \mathsf{rctxt}} \qquad \frac{\mathit{\Theta}\ \mathsf{rctxt}}{\mathit{\Theta}, \alpha\rho\beta\ \mathsf{rctxt}}\ (\rho \notin \mathit{\Theta}, \alpha \notin \mathit{\Theta}_1, \beta \notin \mathit{\Theta}_2)$$

$$\frac{\mathit{\Theta}\ \mathsf{rctxt} \qquad \mathit{\Theta} \vdash A_1 R A_2\ \mathsf{rel}}{\mathit{\Theta}, (x_1 : A_1) R (x_2 : A_2)\ \mathsf{rctxt}}\ (x_1 \notin \mathit{\Theta}_1, x_2 \notin \mathit{\Theta}_2)$$

Relation formation rules:

$$\frac{}{\mathit{\Theta} \vdash \alpha\rho\beta\ \mathsf{rel}}\ (\alpha\rho\beta \in \mathit{\Theta}) \qquad \frac{\mathit{\Theta} \vdash A_1 R A_2\ \mathsf{rel} \qquad \mathit{\Theta} \vdash B_1 S B_2\ \mathsf{rel}}{\mathit{\Theta} \vdash (A_1 \to B_1)(R \to S)(A_2 \to B_2)\ \mathsf{rel}}$$

$$\frac{\mathit{\Theta},\ \alpha\rho\beta \vdash A_1 R A_2\ \mathsf{rel}}{\mathit{\Theta} \vdash (\forall\alpha.\, A_1)(\forall\alpha\rho\beta.\, R)(\forall\beta.\, A_2)\ \mathsf{rel}}$$

$$\frac{\mathit{\Theta} \vdash B_1 R B_2\ \mathsf{rel} \qquad \mathit{\Theta}_1 \vdash t_1 : A_1 \to B_1 \qquad \mathit{\Theta}_2 \vdash t_2 : A_2 \to B_2}{\mathit{\Theta} \vdash A_1([t_1 \times t_2]^{-1} R) A_2\ \mathsf{rel}}$$

Relatedness rules:

$$\frac{}{\mathit{\Theta} \vdash (x_1 : A_1) R (x_2 : A_2)}\ ((x_1 : A_1) R (x_2 : A_2) \in \mathit{\Theta})$$

$$\frac{\mathit{\Theta},\ (x_1 : A_1) R (x_2 : A_2) \vdash (t_1 : B_1) S (t_2 : B_2)}{\mathit{\Theta} \vdash (\lambda x_1.\, t_1 : A_1 \to B_1)(R \to S)(\lambda x_2.\, t_2 : A_2 \to B_2)}$$

$$\frac{\mathit{\Theta} \vdash (s_1 : A_1 \to B_1)(R \to S)(s_2 : A_2 \to B_2) \qquad \mathit{\Theta} \vdash (t_1 : A_1) R (t_2 : A_2)}{\mathit{\Theta} \vdash (s_1\, t_1 : B_1) S (s_2\, t_2 : B_2)}$$

$$\frac{\mathit{\Theta},\ \alpha\rho\beta \vdash (t_1 : A_1) R (t_2 : A_2)}{\mathit{\Theta} \vdash (\mathit{\Lambda}\alpha.\, t_1 : \forall\alpha.\, A_1)\big(\forall\alpha\rho\beta.\, R\big)(\mathit{\Lambda}\beta.\, t_2 : \forall\beta.\, A_2)}$$

$$\frac{\mathit{\Theta} \vdash (t_1 : \forall\alpha.\, A_1)\big(\forall\alpha\rho\beta.\, R\big)(t_2 : \forall\beta.\, A_2) \qquad \mathit{\Theta} \vdash B_1 S B_2\ \mathsf{rel}}{\mathit{\Theta} \vdash (t_1[B_1] : A_1[\alpha \mapsto B_1]) R[\alpha\rho\beta \mapsto B_1 S B_2] (t_2[B_2] : A_2[\beta \mapsto B_2])}$$

$$\frac{\mathit{\Theta} \vdash (t_1\, u_1 : B_1) R (t_2\, u_2 : B_2)}{\mathit{\Theta} \vdash (u_1 : A_1)([t_1 \times t_2]^{-1} R)(u_2 : A_2)} \qquad \frac{\mathit{\Theta} \vdash (u_1 : A_1)([t_1 \times t_2]^{-1} R)(u_2 : A_2)}{\mathit{\Theta} \vdash (t_1\, u_1 : B_1) R (t_2\, u_2 : B_2)}$$

$$\frac{\mathit{\Theta} \vdash (t_1 : A_1) R (t_2 : A_2) \qquad \mathit{\Theta}_1 \vdash t_1 = s_1 : A_1 \qquad \mathit{\Theta}_2 \vdash t_2 = s_2 : A_2}{\mathit{\Theta} \vdash (s_1 : A_1) R (s_2 : A_2)}$$

Parametricity rule:

$$\frac{\langle \mathit{\Gamma} \rangle \vdash (s : A) \langle A \rangle (t : A)}{\mathit{\Gamma} \vdash s = t : A}$$

Fig. 2: The type system $\boldsymbol{\lambda 2R}$

The operations $(-)_i$ (for $i \in \{0, 1\}$) on relational contexts:

$$(\cdot)_i = \cdot$$
$$(\Theta,\, \alpha_1 \rho \alpha_2)_i = (\Theta)_i,\, \alpha_i$$
$$(\Theta,\, (x_1 \!:\! A_1) R (x_2 \!:\! A_2))_i = (\Theta)_i,\, x_i \!:\! A_i$$

The operation $\langle - \rangle$ on contexts and types:

$$\langle \cdot \rangle = \cdot \qquad\qquad \langle \alpha \rangle = \rho^{\alpha}$$
$$\langle \Gamma,\, \alpha \rangle = \langle \Gamma \rangle,\, \alpha\, \rho^{\alpha} \alpha \qquad\qquad \langle A \to B \rangle = (\langle A \rangle \to \langle B \rangle)$$
$$\langle \Gamma,\, x \!:\! A \rangle = \langle \Gamma \rangle,\, (x \!:\! A) \langle A \rangle (x \!:\! A) \qquad\qquad \langle \forall \alpha.\, A \rangle = \forall \alpha\, \rho^{\alpha} \alpha.\, \langle A \rangle$$

Fig. 3: Syntactic reflexive graph structure

two such operations: $R[\alpha\rho\beta \mapsto ASB]$ substitutes, in the relation R, the type A for all left occurrences of α, the type B for all right occurrences of β (which may itself be α), and the relation S for all occurrences of ρ; similarly, $S[x \mapsto s, y \mapsto t]$ substitutes, in the relation S, the term s for all left occurrences of x, and the term t for all right occurrences of y (which may itself be x). Note that relations can indeed contain terms and (hence) type variables, due to the $[t_1 \times t_2]^{-1} R$ construction, where we consider t_1 as occurring on the left, and t_2 on the right.

Lemma 6 (Substitution lemma).

(i) *If* $\Theta \vdash A_1 R A_2 \;\mathsf{rel}$ *and* $\Theta,\, \alpha_1 \rho\, \alpha_2 \vdash (t_1 : B_1) S (t_2 : B_2)$ *then* $\Theta \vdash (t_1[\alpha_1 \mapsto A_1] \!:\! B_1[\alpha_1 \mapsto A_1]) S[\alpha_1 \rho \alpha_2 \mapsto A_1 R A_2] (t_2[\alpha_2 \mapsto A_2] \!:\! B_2[\alpha_2 \mapsto A_2])$.
(ii) *If* $\Theta \vdash (t_1 : A_1) R (t_2 : A_2)$ *and* $\Theta,\, (x_1 : A_1) R (x_2 : A_2) \vdash (s_1 : B_1) S (s_2 : B_2)$ *then* $\Theta \vdash (s_1[x_1 \mapsto t] : B_1) S[x_1 \mapsto t_1, x_2 \mapsto t_2] (s_2[x_2 \mapsto t_2] : B_2)$.

The relatedness rules of Fig. 2 include the expected rules for relations $R \to S$ and $\forall \alpha\rho\beta.\, R$, which mimic the analogous type constructions in $\lambda\mathbf{2}$. The rules for $[t_1 \times t_2]^{-1} R$ implement its intended interpretation as an inverse image construction. In addition, a further rule expresses an extensionality principle for relations with respect to judgemental equality. Such an intermixing of relatedness judgements with equality judgements is legitimised by statement (i) of the lemma below.

Lemma 7.

(i) *If* $\Theta \vdash (t_1 : A_1) R (t_2 : A_2)$ *then* $(\Theta)_i \vdash t_i : A_i$.
(ii) *If* $\Gamma \vdash t : A$ *then* $\langle \Gamma \rangle \vdash (t : A) \langle A \rangle (t : A)$.

Statement (ii) of the lemma asserts that all terms enjoy the characteristic relation-preservation property of relational parametricity. By the extensionality rule, it follows that $\Gamma \vdash s = t : A$ implies $\langle \Gamma \rangle \vdash (s : A) \langle A \rangle (t : A)$. That is, equal terms are *parametrically related*. Since parametric relatedness captures a form of behavioural equivalence, we can ask also for the converse implication to hold. This is implemented by the *parametricity rule* in Fig. 2. This rule, in the general form given, is derivable from its empty-context version: $\vdash (s : A)\langle A\rangle(t : A)$ implies $\vdash s = t : A$. Thus the parametricity rule is equivalent to asking for the relational interpretation of a closed type to act as an identity relation between closed terms—a weak version of Reynold's *identity extension* property [23]. We discuss the relational interpretation of open types in Sect. 5.

We outline the semantic interpretation of $\lambda\mathbf{2R}$. Given a comprehensive $\lambda\mathbf{2}$ parametricity graph, the contexts, types and terms of $\lambda\mathbf{2}$ are interpreted in the comprehensive $\lambda\mathbf{2}$ fibration $p : \mathbb{T} \to \mathbb{C}$, as in Sect. 2. In addition, we interpret a relational context Θ as an object $[\![\Theta]\!]$ of $\mathcal{R}(\mathbb{C})$, and a syntactic relation $\Theta \vdash ARB\ \mathsf{rel}$ as a semantic relation $[\![R]\!]_\Theta : [\![A]\!]_{(\Theta)_1} \leftrightarrow [\![B]\!]_{(\Theta)_2}$ in $\mathcal{R}(\mathbb{T})_{[\![\Theta]\!]}$. The definitions of $[\![\Theta]\!]$ and $[\![R]\!]_\Theta$ interpret context extension, function space and universal quantification using the structure of the comprehensive $\lambda\mathbf{2}$ fibration $p^{\mathcal{R}} : \mathcal{R}(\mathbb{T}) \to \mathcal{R}(\mathbb{C})$, where relation variables $\alpha\rho\beta$ are interpreted using the generic object of $p^{\mathcal{R}}$. For the inverse-image relation $\Theta \vdash A_1([t_1 \times t_2]^{-1}R)A_2\ \mathsf{rel}$, we have that $[\![t_1]\!]_{(\Theta)_1}$ and $[\![t_2]\!]_{(\Theta)_2}$ determine maps $[\![A_1]\!]_{(\Theta)_1} \longrightarrow [\![B_1]\!]_{(\Theta)_1}$ and $[\![A_2]\!]_{(\Theta)_2} \longrightarrow [\![B_2]\!]_{(\Theta)_2}$ in $\mathbb{T}_{[\![(\Theta)_1]\!]}$ and $\mathbb{T}_{[\![(\Theta)_2]\!]}$ respectively. The fibration property of $\langle \nabla_1^{\mathbb{T}}, \nabla_2^{\mathbb{T}}\rangle\restriction_{\mathcal{R}(\mathbb{T})_{[\![\Theta]\!]}}$ then gives $[\![[t_1 \times t_2]^{-1}R]\!] : [\![A_1]\!]_{(\Theta)_1} \leftrightarrow [\![A_2]\!]_{(\Theta)_2}$ as the inverse image of $[\![R]\!] : [\![B_1]\!]_{(\Theta)_1} \leftrightarrow [\![B_2]\!]_{(\Theta)_2}$ along these maps.

In the above semantic interpretation, the comprehension property is needed in order to interpret a relational context Θ as an object $[\![\Theta]\!]$ of $\mathcal{R}(\mathbb{C})$, and essential use is made of this in the definition of $[\![[t_1 \times t_2]^{-1}R]\!]$. Were the comprehension property of models dropped, it would be possible to rejig the semantics to interpret a restricted calculus with inverse-image relations definable only in relational contexts containing no term variables, but not full $\lambda\mathbf{2R}$.

The semantics is supported by soundness and completeness theorems.

Theorem 8 (Soundness for λ2R). *In every comprehensive $\lambda\mathbf{2}$ parametricity graph:*

(i) *if $\Gamma \vdash t_1 = t_2 : A$ then $[\![t_1]\!]_\Gamma = [\![t_2]\!]_\Gamma$; and*
(ii) *if $\Theta \vdash (t_1 : A_1)R(t_2 : A_2)$ then $[\![t_1]\!]_{(\Theta)_1} \times [\![t_2]\!]_{(\Theta)_2} : \mathbf{1}_{[\![\Theta]\!]} \longrightarrow [\![R]\!]_\Theta$.*

Theorem 9 (Full completeness for λ2R). *There exists a comprehensive $\lambda\mathbf{2}$ parametricity graph satisfying the following.*

(i) *For every type $\Gamma \vdash A$ type, every global point $\mathbf{1}_{[\![\Gamma]\!]} \longrightarrow [\![A]\!]_\Gamma$ is the denotation $[\![t]\!]_\Gamma$ of some term $\Gamma \vdash t : A$.*
(ii) *For all terms $\Gamma \vdash t_1, t_2 : A$ satisfying $[\![t_1]\!]_\Gamma = [\![t_2]\!]_\Gamma$, we have $\Gamma \vdash t_1 = t_2 : A$.*
(iii) *For every relation $\Theta \vdash A_1RA_2$ type, every global point $\mathbf{1}_{[\![\Theta]\!]} \longrightarrow [\![R]\!]_\Theta$ arises as $[\![t_1]\!]_{(\Theta)_1} \times [\![t_2]\!]_{(\Theta)_2}$ for terms t_1, t_2 such that $\Theta \vdash (t_1 : A_1)R(t_2 : A_2)$.*

Theorem 8 is proved by induction on derivations. We highlight that the soundness of the parametricity rule follows from the identity property of comprehensive $\lambda\mathbf{2}$ parametricity graphs. Theorem 9 is proved by a term model construction.

5 Direct-image and Pseudograph Relations

As already discussed, the parametricity rule of Fig. 2 interprets the relation $\langle A\rangle$ as an identity relation when A is a closed type. When A contains type variables, however, this interpretation is not available. Consider an open type $\alpha \vdash A(\alpha)$ type (where we write $A(\alpha)$ to highlight the occurrences of α in A). Then we have $\alpha\rho\alpha \vdash A(\alpha)\big(\langle A\rangle(\rho)\big)A(\alpha)$ rel. However, the independent handling of left and right variables in $\lambda\mathbf{2R}$ (forced by the semantic correspondence with comprehensive $\lambda\mathbf{2}$ parametricity graphs), means that the latter relation is equivalent to $\alpha\rho\beta \vdash A(\alpha)\big(\langle A\rangle(\rho)\big)A(\beta)$ rel; i.e., it is a family (indexed by relations ρ) of relations between different types. Indeed, the distinctness of left and right type variables means $\lambda\mathbf{2R}$ has no facility for formulating relations between open types and themselves. In particular, $\lambda\mathbf{2R}$ contains no mechanism for defining identity relations on open types. Nonetheless, the relation $\langle A\rangle$ can act as a kind of *pseudo-identity* relation for type A where the parametricity rule can establish equalities from $\langle A\rangle$-relatedness *in relational contexts of the form* $\langle\Gamma\rangle$.

Graphs of functions are ubiquitous in standard arguments involving relational parametricity. Since we have only pseudo-identity relations, we correspondingly have only *pseudographs* available in $\lambda\mathbf{2R}$. Suppose $\Gamma \vdash f : A \to B$. Define:

$$gr_*(f) \coloneqq [f \times \mathsf{id}_B]^{-1}\langle B\rangle$$

Clearly $\langle\Gamma\rangle \vdash A\, gr_*(f)\, B$ rel. Its defining property is that $(x : A)\, gr_*(f)\, (y : B)$ holds if and only if $(fx : B)\langle B\rangle(y : B)$. Mathematically, there is, however, another natural pseudograph relation, for f, between A and B. This is the relation $gr_!(f)$ defined by $(x : A) gr_!(f)(y : B)$ if there exists $w : A$ such that $(x : A)\langle A\rangle(w : A)$ and $y = fw$. Since, by (ii) of Lemma 7, f maps $\langle A\rangle$-related values to $\langle B\rangle$-related values, $gr_!(f) \subseteq gr_*(f)$. However, because $\langle A\rangle$ and $\langle B\rangle$ are not identity relations, there is no need for this inclusion to be an equality. We shall need to make use of *both* forms of pseudograph relation to derive the standard consequences of parametricity. In order to do so, we must first provide a definition of $gr_!(f)$ in $\lambda\mathbf{2R}$ itself, and establish formal analogues of the informal observations above.

The main construction we need is that of *direct-image relations* $[t_1 \times t_2]_! R$, dual to inverse-image relations. This is achieved using an impredicative encoding.

Theorem 10 (Direct-image relations). *Using the definition*

$$[t_1 \times t_2]_! R \coloneqq [i_{B_1} \times i_{B_2}]^{-1}(\forall\alpha\rho\alpha.\, ([(- \circ t_1) \times (- \circ t_2)]^{-1}(R \to \rho)) \to \rho)$$

where i_B abbreviates $\lambda b.\, \Lambda\alpha.\, \lambda t.\, t\, b : B \to \forall\alpha.\, (B \to \alpha) \to \alpha$ and $(- \circ t_j)$ abbreviates $\lambda(v_j : B_j \to \alpha).\, \lambda(x_j : A_j).\, v_j(t_1\, x_j)$, $\lambda\mathbf{2R}$ supports the derived rules below.

$$\frac{\Theta \vdash A_1 R A_2 \text{ rel} \qquad \begin{array}{c}\Theta_1 \vdash t_1 : A_1 \to B_1 \\ \Theta_2 \vdash t_2 : A_2 \to B_2\end{array}}{\Theta \vdash B_1([t_1 \times t_2]_! R) B_2 \text{ rel}} \qquad \frac{\Theta \vdash (u_1 : A_1) R (u_2 : A_2)}{\Theta \vdash (t_1 u_1 : B_1)\big([t_1 \times t_2]_! R\big)(t_2 u_2 : B_2)}$$

$$\frac{\Theta \vdash (u_1 : B_1)\big([t_1 \times t_2]_! R\big)(u_2 : B_2) \qquad \begin{array}{c}\Theta \vdash C_1 Q C_2 \text{ rel} \quad \Theta_1 \vdash v_1 : B_1 \to C_1 \quad \Theta_2 \vdash v_2 : B_2 \to C_2 \\ \Theta \vdash (v_1 \circ t_1 : A_1 \to C_1)\big(R \to Q\big)(v_2 \circ t_2 : A_2 \to C_2\end{array}}{\Theta \vdash (v_1\, u_1 \colon C_1) Q (v_2\, u_2 : C_2)}$$

In fact, these rules are derivable without use of the parametricity rule of $\lambda\mathbf{2R}$.

It is now straightforward to define the second form of pseudograph relation discussed above. Suppose that $\Gamma \vdash f \colon A \to B$ and define $\langle\Gamma\rangle \vdash A\, gr_!(f)\, B$ rel by:

$$gr_!(f) \coloneqq [\mathsf{id}_A \times f]_! \langle A\rangle \ .$$

To understand the relationship between the two pseudograph relations we introduce some notation. Given R and S such that $\Theta \vdash ARB$ rel and $\Theta \vdash ASB$ rel, let $\Theta \vdash R \subseteq S$ abbreviate $\Theta, (x : A)R(y : B) \vdash (x : A)S(y : B)$.

Lemma 11. *If $\Gamma \vdash f : A \to B$ then:*

(i) $\langle\Gamma\rangle \vdash gr_!(f) \subseteq gr_*(f)$*; and*
(ii) $\langle\Gamma\rangle \vdash (s\!:\!A)\, gr_*(f)\, (t\!:\!B)$ *iff* $\Gamma \vdash f\, s \,=\, t \colon B$ *iff* $\langle\Gamma\rangle \vdash (s\!:\!A)\, gr_!(f)\, (t\!:\!B)$.

We comment that, in spite of item (ii), the converse inclusion to (i) does not hold in general. Property (ii) applies only in context $\langle\Gamma\rangle$, and thus implies nothing about what happens if further relational assumptions are added.

Theorem 10 has a semantic analogue: direct image relations correspond to opfibrational structure on comprehensive $\lambda\mathbf{2}$ parametricity graphs.

Theorem 12. *In any comprehensive $\lambda\mathbf{2}$ parametricity graph, for every object W of $\mathcal{R}(\mathbb{C})$, the functor $\langle\nabla_1^{\mathbb{T}}, \nabla_2^{\mathbb{T}}\rangle\restriction_{\mathcal{R}(\mathbb{T})_W} : \mathcal{R}(\mathbb{T})_W \to \mathbb{T}_{\nabla_1^{\mathbb{C}} W} \times \mathbb{T}_{\nabla_2^{\mathbb{C}} W}$ is an opfibration.*

6 Consequences of Parametricity

System $\lambda\mathbf{2R}$ is strong enough to prove the familiar consequences of parametricity.

Theorem 13 (Consequences of Parametricity). *System $\lambda\mathbf{2R}$ proves:*

(i) The unit (terminal) type can be encoded as $\mathbf{1} = \forall\alpha.\, \alpha \to \alpha$.
(ii) The product of A and B can be encoded as $A \times B = \forall\alpha.\, (A \to B \to \alpha) \to \alpha$.
(iii) The empty (initial) type can be encoded as $\mathbf{0} = \forall\alpha.\, \alpha$.
(iv) The sum of A and B can be encoded as $A{+}B = \forall\alpha.\, (A{\to}\alpha){\to}(B{\to}\alpha){\to}\alpha$.
(v) Existential types can be encoded as $\exists\alpha.\, T(\alpha) = \forall\alpha.\, (\forall\beta.\, (T(\beta) \to \alpha)) \to \alpha$.
(vi) The type $\forall\alpha.\, (T(\alpha) \to \alpha) \to \alpha$ is the carrier of the initial T-algebra for all functorial type expressions $T(\alpha)$.
(vii) The type $\exists\alpha.\, (\alpha \to T(\alpha)) \times \alpha$ is the carrier of the final T-coalgebra for all functorial type expressions $T(\alpha)$.

This result for $\lambda\mathbf{2R}$ implies that analogous category-theoretic properties (which we do not state for lack of space) hold for comprehensive $\lambda\mathbf{2}$ parametricity graphs.

The proofs of (i)–(vii) follow the usual ones, see, e.g., [21], but with graph relations replaced by pseudographs. Pseudograph relations of the form $gr_*(f)$ suffice in all proofs with the exception of the verification of final coalgebras, where $gr_!(f)$ is used. In this section, we explain how this difference in the treatment of initial algebras and final coalgebras arises. For lack of space, we focus on the use of pseudograph relations only, and omit the (standard) supporting arguments.

Suppose $\Gamma, \alpha \vdash T$ type. We write $T(\alpha)$ to highlight the occurrences of α in T, and $T(A)$ for the substitution $T[\alpha \mapsto A]$. If α occurs only positively in T (i.e., not on the left-hand side of an odd number of arrows) then it is standard that T defines an endofunctor on types. If $\Gamma' \vdash f : A \to B$, where Γ' extends Γ, then we use the notation $\Gamma' \vdash T(f) : T(A) \to T(B)$ for the functorial action of T. This action preserves identities and composition up to judgemental equality. In addition, the corresponding relational substitution preserves pseudo-identity relations; i.e., $\langle T \rangle(\langle A \rangle)$ (by which we mean the substitution $\langle T \rangle[\alpha\rho^{\alpha}\alpha \mapsto \langle A \rangle]$) syntactically coincides with $\langle T(A) \rangle$. Also, the functorial action lifts to relations: if $\Theta \vdash (f\!:\!A \to B)(R \to S)(f'\!:\!A' \to B')$, where Θ extends $\langle \Gamma \rangle$, then:

$$\Theta \vdash (T(f)\!:\!T(A) \to T(B))(\langle T \rangle(R) \to \langle T \rangle(S))(T(f')\!:\!T(A') \to T(B')) \enspace .$$

Using these facts (which assert that T is a reflexive-graph functor [8]) one establishes the following properties of the action of $\langle T \rangle$ on pseudograph relations.

Lemma 14. *Suppose α occurs positively in $\Gamma, \alpha \vdash T$ type and $\Gamma' \vdash f\colon A \to B$, where Γ' extends Γ.*

(i) $\langle \Gamma' \rangle \vdash \langle T \rangle(gr_*(f)) \subseteq gr_*(T(f))$.
(ii) $\langle \Gamma' \rangle \vdash gr_!(T(f)) \subseteq \langle T \rangle(gr_!(f))$.

Our proof of this lemma closely mirrors the proof of the Graph Lemma in [9], which exploits the fact that *graph relations* can be defined either using inverse image, analogously to $gr_*(f)$, or using direct image, analogously to $gr_!(f)$.

We now explain how Lemma 14 bears on the proofs of the universal properties of initial algebras and final coalgebras. Given T as above, standard constructions produce a T-algebra and a T-coalgebra that can be shown to be weakly initial and weakly final respectively, without invoking parametricity. The parametricity rule is used to establish the uniqueness part of the universal property. In the initiality and finality arguments, one is led to consider T-algebra and T-coalgebra homomorphisms respectively:

$$\begin{tikzcd} T(A) \arrow[r, "T(h)"] \arrow[d, "a"'] & T(B) \arrow[d, "b"] \\ A \arrow[r, "h"'] & B \end{tikzcd} \qquad \begin{tikzcd} T(A) \arrow[r, "T(h')"] & T(B) \\ A \arrow[r, "h'"'] \arrow[u, "a'"] & B \arrow[u, "b'"'] \end{tikzcd}$$

where the diagrams are given by terms, in a context Γ' extending Γ, which commute up to judgemental equality. Lemma 14 allows one to prove the following crucial properties as consequences of the commutativity of the above diagrams.

$$\langle \Gamma' \rangle \vdash (a\colon T(A) \to A)\left(\langle T \rangle(gr_*(h)) \to gr_*(h)\right)(b\colon T(B) \to B)$$
$$\langle \Gamma' \rangle \vdash (a'\colon A \to T(A))\left(gr_!(h') \to \langle T \rangle(gr_!(h'))\right)(b'\colon B \to T(B))$$

It is the orientation of the function relations above that necessitates the use of a different type of pseudograph relation in each case. Modulo this subtlety, the remaining proofs of initiality and finality proceed as usual, cf. [21].

7 Related and Further Work

System R of [1] and System P of [7] share with $\lambda\mathbf{2R}$ the property of having a syntax in which function space and universal quantification are basic constructions on relations. Indeed $\lambda\mathbf{2R}$ is especially similar to System P, which also has the inverse-image-relation constructor $[t_1 \times t_2]^{-1} R$. The most significant difference is that, in System P, the formation rule for this construction is restricted: the terms t_1, t_2 are not allowed to contain free term variables. However, they are permitted to contain so-called *indeterminates*, which, in the semantics of System P, range over *global elements* in models. This device allows System P to be used to establish consequences of parametricity in *well-pointed* models [7]. In $\lambda\mathbf{2R}$, our general arguments for consequences of parametricity make essential use of the possibility for t_1 and t_2 to contain free term variables. As already observed in Sect. 4, the comprehension property of our models is crucial to the semantic interpretation of inverse-image relations in such cases.

System R of [1] departs from $\lambda\mathbf{2R}$ (and System P) in two main ways. The first is that, in System R, every type A has an associated *identity relation* A^*.[4] A key rule of System R (written in our notation) is that $\Theta \vdash x \, A^* \, x$, whenever $x : A$ appears anywhere in relational context Θ. This rule breaks the independence between left and right variables in the relational judgements of $\lambda\mathbf{2R}$. (For example, property (i) of Lemma 7 fails.) The second difference is that System R has an explicit syntax for defining graph relations, rather than the inverse-image construct of $\lambda\mathbf{2R}$ (and System P), which would be more general in that context. Due to the presence of both identity and graph relations, the arguments, in System R, for consequences of parametricity proceed along standard lines [1]. However, System R currently lacks a corresponding semantic story of the kind we have used in this paper in justification of $\lambda\mathbf{2R}$.

In fact, the interplay between models and syntax could be pushed much further than in the present paper. By adding primitive product types to $\lambda\mathbf{2}$ and $\lambda\mathbf{2R}$, one can strengthen our full completeness results by obtaining syntactic categories that are initial in an appropriate 2-category of strict structure-preserving morphisms of models. It would be more natural, however, to broaden both the notion of model, by replacing splittings of fibrations with cleavages, and the notion of morphism, by permitting non-strict structure preservation. With such a relaxation, coherence issues arise, but one would expect to obtain (pseudo-)initiality of the syntactic model of $\lambda\mathbf{2R}$ (without any need to extend the syntax with products).

For lack of space we have not presented any concrete models in this paper. In fact, any instance of the more elaborate axiomatic structure from [6] can be reconstrued (albeit in a nontrivial way) as a comprehensive $\lambda\mathbf{2}$ parametricity graph. So our minimal structure at least generalises the known models of parametricity. However, we do not know whether our structure encompasses any genuinely new models of relational parametricity that truly exploit the (potential) added generality of our approach.

[4] In System P, every type A is itself a relation, which, although called an "identity relation" in [7], has the properties of the relation $\langle A \rangle$ in the present paper.

The results of the present paper should be contrasted with those of other recent work by first two authors and colleagues [9,10]. In this paper, we have axiomatised category-theoretic structure modelling relational parametricity for the specific type theory $\lambda\mathbf{2}$, where the resulting structure encompasses both 'syntactic' and 'semantic' models. In contrast, [9,10] axiomatise the category-theoretic structure required on a 'semantic' model for Reynolds' original set-theoretic definition of relational parametricity [23] to generalise to the model. Interestingly, the category-theoretic notion of bifibration occurs both as a central ingredient in the axiomatisation of [9,10], and, in the guise of direct-image relations, as a vital tool in the present paper. A novelty in the present paper is that the bifibrational structure is derived rather than assumed.

From a type-theoretic perspective, one advantage of the approach followed in this paper is that the passage from the original type theory ($\lambda\mathbf{2}$) to the relational version ($\lambda\mathbf{2R}$) appears not to depend on specific properties of the former, other than that essential use is made of judgemental equality in the formulation of the parametricity rule. We believe that this potential flexibility may be useful for transferring our methods to dependent type theories, where parametricity is an active area of study [2,4,5,17].

The proof-relevant setting of dependent type theory, however, requires modifications to our semantic framework. In particular the *relational* property of parametricity graphs must be relaxed. Ongoing work on a higher-dimensional, proof-relevant form of parametricity may show how to remove this requirement.

Acknowledgements. We thank Bob Atkey, Claudio Hermida, Rasmus Møgelberg and the anonymous reviewers for helpful discussions and comments. This research was supported by EPSRC grants GR/A11731/01, EP/E016146/1, EP/K023837/1 and EP/M016951/1.

References

1. Martín Abadi, Luca Cardelli, and Pierre-Louis Curien. Formal parametric polymorphism. *Theoretical Computer Science*, 121(1&2):9–58, 1993.
2. Robert Atkey, Neil Ghani, and Patricia Johann. A relationally parametric model of dependent type theory. In Suresh Jagannathan and Peter Sewell, editors, *POPL*, pages 503–515. ACM, 2014.
3. Henk Barendregt. Introduction to generalized type systems. *Journal of Functional Programming*, 1(2):125–154, 1991.
4. Jean-Philippe Bernardy, Thierry Coquand, and Guilhem Moulin. A presheaf model of parametric type theory. In Dan R. Ghica, editor, *MFPS*, ENTCS, pages 17–33. Elsevier, 2015.
5. Jean-Philippe Bernardy, Patrik Jansson, and Ross Paterson. Proofs for free. *Journal of Functional Programming*, 22:107–152, 2012.
6. Lars Birkedal and Rasmus E. Møgelberg. Categorical models for Abadi and Plotkin's logic for parametricity. *Mathematical Structures in Computer Science*, 15:709–772, 2005.
7. Brian Dunphy. *Parametricity as a notion of uniformity in reflexive graphs*. PhD thesis, University of Illinois, 2002.

8. Brian Dunphy and Uday Reddy. Parametric limits. In *LICS*, pages 242–251, 2004.
9. Neil Ghani, Patricia Johann, Fredrik Nordvall Forsberg, Federico Orsanigo, and Timothy Revell. Bifibrational functorial semantics of parametric polymorphism. In Dan R. Ghica, editor, *MFPS*, ENTCS, pages 67–83. Elsevier, 2015.
10. Neil Ghani, Fredrik Nordvall Forsberg, and Federico Orsanigo. Parametric polymorphism — universally. In Valeria de Paiva, Ruy de Queiroz, Lawrence S. Moss, Daniel Leivant, and Anjolina G. de Oliveira, editors, *WoLLIC*, volume 9160 of *LNCS*, pages 81–92. Springer, 2015.
11. Jean-Yves Girard. *Interprétation fonctionelle et élimination des coupures dans l'arithmétique d'ordre supérieur*. PhD thesis, University of Paris VII, 1972.
12. Claudio Hermida. Fibrational relational polymorphism. Available at `http://maggie.cs.queensu.ca/chermida/papers/FibRelPoly.pdf`, 2006.
13. Claudio Hermida, Uday Reddy, and Edmund Robinson. Logical relations and parametricity — a Reynolds programme for category theory and programming languages. *ENTCS*, 303:149–180, 2014.
14. Bart Jacobs. Comprehension categories and the semantics of type dependency. *Theoretical Computer Science*, 107(2):169–207, 1993.
15. Bart Jacobs. *Categorical Logic and Type Theory*. Elsevier, 1999.
16. Yoshiki Kinoshita, Peter W. O'Hearn, John Power, Makoto Takeyama, and Robert D. Tennent. An axiomatic approach to binary logical relations with applications to data refinement. In Martín Abadi and Takayasu Ito, editors, *TACS*, pages 191–212, 1997.
17. Neelakantan R. Krishnaswami and Derek Dreyer. Internalizing relational parametricity in the extensional calculus of constructions. In Simona Ronchi Della Rocca, editor, *CSL*, pages 432–451, 2013.
18. F William Lawvere. Equality in hyperdoctrines and comprehension schema as an adjoint functor. *Applications of Categorical Algebra*, 17:1–14, 1970.
19. QingMing Ma and John C. Reynolds. Types, abstractions, and parametric polymorphism, part 2. In Stephen D. Brookes, Michael G. Main, Austin Melton, Michael W. Mislove, and David A. Schmidt, editors, *MFPS*, volume 598 of *LNCS*, pages 1–40. Springer, 1991.
20. Peter W O'Hearn and Robert D Tennent. Parametricity and local variables. *Journal of the ACM*, 42(3):658–709, 1995.
21. Gordon Plotkin and Martín Abadi. A logic for parametric polymorphism. In Marc Bezem and Jan Friso Groote, editors, *TLCA*, volume 664 of *LNCS*, pages 361–375. Springer, 1993.
22. John Reynolds. Towards a theory of type structure. In B. Robinet, editor, *Programming Symposium*, volume 19 of *LNCS*, pages 408–425. Springer, 1974.
23. John Reynolds. Types, abstraction and parametric polymorphism. In R. E. A. Mason, editor, *Information Processing*, pages 513–523, 1983.
24. E.P. Robinson and G. Rosolini. Reflexive graphs and parametric polymorphism. In *LICS*, pages 364–371. IEEE Computer Society, 1994.
25. Robert A.G Seely. Categorical semantics for higher order polymorphic lambda calculus. *Journal of Symbolic Logic*, pages 969–989, 1987.
26. Christopher Strachey. Fundamental concepts in programming languages. *Higher Order Symbolic Computation*, 13(1-2):11–49, 2000.
27. Philip Wadler. Theorems for free! In Joseph E. Stoy, editor, *FPCA*, pages 347–359. ACM, 1989.