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, $\lambda_2$, (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 $\lambda_2$, 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 $\lambda_2$ were developed many years ago by Seely [25]. They are studied systematically as $\lambda_2$ 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 \(\lambda^2\), it is natural
to combine the structure of \(\lambda^2\) fibrations with that of parametricity graphs. This
results in the notion of \(\lambda^2\) parametricity graph, which we define in Section 3.
Sadly, \(\lambda^2\) 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-
thoretic structure, which overcomes the limitation by modelling Plotkin and
Abadi’s logic for parametricity \[21\]. This method of modelling the combination of
\(\lambda^2\) 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 \(\lambda^2\) with that of parametricity graphs.

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

Sections 2 and 3 define comprehensive \(\lambda^2\) fibrations and comprehensive \(\lambda^2\)
parametricity graphs respectively. In Section 4 we present a type theory \(\lambda^{2R}\),
corresponding to our category-theoretic structure, which provides a simple system
for reasoning about parametricity. The type theory \(\lambda^{2R}\) 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 \(\lambda^{2R}\). 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 \(\lambda^{2R}\), and hence always present in comprehensive \(\lambda^2\) parametricity
graphs. In Section 6, we finally apply the technical machinery and establish that
the expected consequences of relational parametricity are indeed derivable in
\(\lambda^{2R}\), and hence hold in comprehensive \(\lambda^2\) parametricity graphs.

In summary, the main contributions of this work are:
(i) The definition of comprehensive $\lambda 2$ fibrations as models of $\lambda 2$.
(ii) The definition of a new category-theoretic notion of model of relational parametricity, obtained by combining parametricity graphs and comprehensive $\lambda 2$ fibrations into comprehensive $\lambda 2$ parametricity graphs.
(iii) The extraction of $\lambda 2R$ as the type theory intrinsic to comprehensive $\lambda 2$ parametricity graphs.
(iv) The derivation of the expected consequences of parametricity in $\lambda 2R$, and hence in comprehensive $\lambda 2$ parametricity graphs. This requires novel techniques: establishing the opfibration property of comprehensive $\lambda 2$ 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 $\lambda 2$ Fibrations

In Fig. 1 we recall the polymorphic $\lambda$-calculus $\lambda 2$. We use $x, y, \ldots$ to range over term variables, and $\alpha, \beta, \ldots$ to range over type variables. Our presentation has four judgements: $\Gamma \text{ ctxt}$, stating that $\Gamma$ is a well-formed context; $\Gamma \vdash A \text{ type}$, stating that $A$ is a well-formed type in context $\Gamma$; $\Gamma \vdash t : A$, stating that the term $t$ has type $A$ in context $\Gamma$; and judgemental equality $\Gamma \vdash t_1 = t_2 : A$. We assume $\beta$ and $\eta$-equalities for both term abstraction, $\lambda$, and type abstraction, $\Lambda$. 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 $\lambda 2$ is that we interleave type variables and term variables in a single context. This approach is not only natural, but indeed standard when $\lambda 2$ is considered in the context of dependent type theory; for example, when derived as an instance of a pure type system [9]. Since there is no dependency of $\lambda 2$ 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 $\lambda 2$ fibration, which models $\lambda 2$. We directly restrict the definition to the split case to circumvent coherence issues that would otherwise arise, cf. [15].

**Definition 1 ($\lambda 2$ fibration).** A $\lambda 2$ fibration is a split fibration $p : T \to C$, where the base category $C$ 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 $\omega$ for $pU$;
(iii) and has fibred-products along projections $X \times \omega \longrightarrow X$ in $C$.

Moreover, the reindexing functors given by the splitting are required to preserve the above-specified structure in fibres on the nose.
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 λ2, and their existence is anyway derivable in parametric models.

In a λ2 fibration, we write \( T_X \) for the fibre category over \( X \). We also use \( X \) as a subscript when referring to structure in \( T_X \); e.g., \( 1_X \) is the specified terminal object in \( T_X \), and \( \Rightarrow_X \) is the exponential structure in \( T_X \). Given \( f : X \rightarrow Y \) in \( \mathbb{C} \), we write \( f^* \) for the reindexing functor \( T_Y \rightarrow T_X \), and \( A^* f : f^* A \rightarrow A \) 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 \( \pi_i^* : T_X \rightarrow T_{X \times \Omega} \).

We recall in outline the semantic interpretation of λ2 in a λ2 fibration \( T \rightarrow \mathbb{C} \). A context \( \Theta = \alpha_1, \ldots, \alpha_n \) of type variables is interpreted as the \( n \)-fold product \( [\Theta] = \Omega^n \) in \( \mathbb{C} \). A type \( A \) in type-variable context \( \Theta \) is then interpreted as an object \( [A]_{\Theta} \) of \( T \) 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 \( \pi_i : \Omega^n \rightarrow \Omega \) to interpret \( \alpha_i \) over \( [\Theta] \). Finally, the interpretation of a term \( \Gamma \vdash t : A \) is obtained by splitting \( \Gamma \) into its component contexts: \( \Theta \) of type variables, and \( \Delta \) of term variables. Then \( \Delta = x_1 : A_1, \ldots, x_m : A_m \) 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 \( [t]_\Gamma : [\Delta]_{\Theta} \rightarrow [A]_{\Theta} \) in \( T_{[\Theta]} \).

![Fig. 1: The type system λ2](image)
In the above outline, one sees that the structure of a $\lambda 2$ 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 \cite{Law} to model the process of context extension \cite{13}. It is natural to apply the same idea to $\lambda 2$.

**Definition 2 (Comprehensive $\lambda 2$ fibration).** A $\lambda 2$ fibration $p : T \to C$ is **comprehensive** if it enjoys the comprehension property \cite{13} Def. 10.4.7: the terminal-object functor $X \mapsto 1_X : C \to T$ has a specified right adjoint $K : T \to C$.

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

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

\[
\begin{align*}
\sem{\cdot} &= 1 \\
\sem{\Gamma, \alpha} &= \sem{\Gamma} \times \Omega \\
\sem{\Gamma, x : A} &= K\sem{A}_{\Gamma} \\
\sem{\forall \alpha. A} &= \prod_{\Omega} \sem{A}_{\Gamma, \alpha} \\
\sem{\exists \alpha. A} &= \prod_{\Omega} \sem{A}_{\Gamma, \alpha} \\
\end{align*}
\]

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

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

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

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

(i) for every type $\Gamma \vdash A$ type, every global point $1_{\sem{\Gamma}} \longrightarrow \sem{A}_{\Gamma}$ is the denotation $\sem{\Gamma}$ of some term $\Gamma \vdash t : A$; and

(ii) for all terms $\Gamma \vdash t_1, t_2 : A$ satisfying $\sem{t_1}_{\Gamma} = \sem{t_2}_{\Gamma}$, 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 $\lambda 2$ Parametricity Graphs

Reflexive graph categories are studied in \cite{19, 21, 16, 78} 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 \( \nabla_1, \nabla_2 : \mathbb{E} \to \mathbb{V} \) and \( \Delta : \mathbb{V} \to \mathbb{E} \) satisfying \( \nabla_1 \Delta = \text{id}_\mathbb{V} = \nabla_2 \Delta \). Informally, one thinks of \( \mathbb{E} \) as a category whose objects are binary ‘relations’ between objects of \( \mathbb{V} \). Then \( \nabla_1, \nabla_2 \) are ‘projection’ functors, and \( \Delta \) 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 \( \nabla_1, \nabla_2 \) means that the relational interpretation commutes with the usual non-relational interpretation of types. The preservation of structure by \( \Delta \), 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) \( \lambda 2 \) fibrations. A morphism from one (comprehensive) \( \lambda 2 \) fibration \( p' : \mathbb{T}' \to \mathbb{C}' \) to another \( p : \mathbb{T} \to \mathbb{C} \) is given by a pair of functors, \( H : \mathbb{T}' \to \mathbb{T} \) and \( L : \mathbb{C}' \to \mathbb{C} \) such that \( pH = Lp' \), 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) \( \lambda 2 \) fibrations, we thus mean a pair of (comprehensive) \( \lambda 2 \) fibrations with functors between them:

\[
\begin{array}{ccc}
\nabla_1^\mathbb{T}, & \Delta^\mathbb{T}, & \nabla_2^\mathbb{T} \\
R(\mathbb{T}) & \leftrightarrow & \mathbb{T} \\
p^R & | & p \\
R(\mathbb{C}) & \leftrightarrow & \mathbb{C} \\
\nabla_1^\mathbb{C}, & \Delta^\mathbb{C}, & \nabla_2^\mathbb{C}
\end{array}
\]

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) \( \lambda 2 \) fibrations, and where each of the triples \( \nabla_1^\mathbb{T}, \Delta^\mathbb{T}, \nabla_2^\mathbb{T} \) and \( \nabla_1^\mathbb{C}, \Delta^\mathbb{C}, \nabla_2^\mathbb{C} \) is a reflexive graph category. We emphasise that \( p^R : R(\mathbb{T}) \to R(\mathbb{C}) \), in \([1]\), is an arbitrary (comprehensive) \( \lambda 2 \) fibration fitting into the diagram. The notation \( R(\cdot) \) is merely mnemonic, and does not imply that \( R(\mathbb{T}) \) 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 \( R(\mathbb{T}) \) behave sufficiently like relations. In \([19]\), this was addressed by requiring the fibre category \( R(\mathbb{T})_{1_{R(\mathbb{C})}} \), over the terminal object, to coincide with a particular category of logical relations over \( \mathbb{T}_{1_{\mathbb{C}}} \). As well as only being applicable if \( \mathbb{T}_{1_{\mathbb{C}}} \) has (sufficient) finite limits, this requirement also has the weakness that it says nothing about other fibres of \( p^R \). 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 \( p^R \). This is provided by Dunphy and Reddy’s notion of parametricity graph \([7,8]\), which we now recall.

A reflexive graph category \( \nabla_1, \nabla_2 : \mathbb{E} \to \mathbb{V}, \Delta : \mathbb{V} \to \mathbb{E} \) is said to be relational if the functor \( \langle \nabla_1, \nabla_2 \rangle : \mathbb{E} \to \mathbb{V} \times \mathbb{V} \) 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 \( R : A \leftrightarrow B \) to mean an object \( R \) of \( \mathbb{E} \) with \( \nabla_1 R = A \) and \( \nabla_2 R = B \), and we write \( f \times g : R \to S \) to mean that there is a (necessarily unique) map \( h : R \to S \) in \( \mathbb{E} \) with \( \nabla_1 h = f \) and \( \nabla_2 h = g \). A reflexive graph category satisfies the identity property if, for every \( h : \Delta A \to \Delta B \) in \( \mathbb{E} \), it holds that \( \nabla_1 h = \nabla_2 h \). This allows one to think of \( \Delta A \) 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 \( \Delta \). A parametricity graph is a relational reflexive graph category with the identity property, for which the functor \( \langle \nabla_1, \nabla_2 \rangle : \mathbb{E} \to \mathbb{V} \times \mathbb{V} \) 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 \( f : A' \to A \) and \( g : B' \to B \) in \( \mathbb{V} \), reindexing produces an inverse image relation \( [f \times g]^{-1} R : A' \leftrightarrow B' \).

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

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

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

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

\[
\begin{array}{ccc}
\mathcal{R}(\mathbb{T})_W & \xrightarrow{(p^R)^* \phi} & \mathcal{R}(\mathbb{T})_{W'} \\
\langle \nabla_1^T, \nabla_2^T \rangle |_{\mathcal{R}(\mathbb{T})_W} \downarrow & & \downarrow \langle \nabla_1^T, \nabla_2^T \rangle |_{\mathcal{R}(\mathbb{T})_{W'}} \\
\mathbb{T}_{\nabla_1^T W} \times \mathbb{T}_{\nabla_2^T W} & \xrightarrow{p^*(\nabla_1^T \phi) \times p^*(\nabla_2^T \phi)} & \mathbb{T}_{\nabla_1^{T'} W'} \times \mathbb{T}_{\nabla_2^{T'} W'}
\end{array}
\]

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

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

\footnote{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 $\lambda 2R$ 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 $\lambda 2R$, suggested by the structure of comprehensive $\lambda 2$ 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 $\lambda 2R$ are given by Fig. 1 (it extends $\lambda 2$) in combination with Fig. 2. The latter adds three new judgements: $[\Theta\ rctxt]$ says that $\Theta$ is a well-defined relational context; $[\Theta \vdash A_1 RA_2 \ rel]$ says that $R$ is a relation between types $A_1$ and $A_2$, in relational context $\Theta$; and $[\Theta \vdash (t_1: A_1) R (t_2: A_2)]$ 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 $\rho, \ldots$, 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 $\rho^\alpha$ to every type variable $\alpha$. 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 $\Theta$, all variables $\alpha, \beta, \rho$ need to be sufficiently fresh. However, the formulation of $\lambda 2R$ is such that variables on the left-hand side of relations are always manipulated separately from variables on the right. Thus, for example, $\alpha$ is sufficiently fresh in $\alpha \rho \beta$, as long as $\alpha$ does not already occur on the left side $(\Theta)_1$ of $\Theta$. 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 $(\cdot)$ 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
Related context formation rules:

\[ \begin{align*}
\Theta \text{ ctx} & \quad \Theta \text{ ctx} \quad \Theta \vdash A_1 R A_2 \text{ rel} \\
\Theta, (x_1 : A_1) R (x_2 : A_2) \text{ ctx} & \quad (x_1 \notin \Theta, x_2 \notin \Theta)
\end{align*} \]

Relation formation rules:

\[ \begin{align*}
\Theta \vdash \alpha \rho \beta \text{ rel} & \quad (\alpha \rho \beta \in \Theta) \\
\Theta \vdash A_1 R A_2 \text{ rel} & \quad \Theta \vdash B_1 S B_2 \text{ rel} \\
\Theta, (\forall \alpha. A_1) (\forall \alpha \rho \beta. R) (\forall \beta. A_2) \text{ rel} & \quad \Theta \vdash B_1 R B_2 \text{ rel} \\
\Theta, \alpha \rho \beta \vdash (t_1 : A_1) R (t_2 : A_2) & \quad \Theta, \alpha \rho \beta \vdash (t_1 : A_1) S (t_2 : B_2) \\
\Theta, (s_1 : A_1) \vdash (R \rightarrow S) (s_2 : A_2) \text{ rel} & \quad \Theta \vdash (s_1 : A_1) R (t_2 : A_2) \\
\Theta, (s_1 : A_1) \vdash B_1 (R \rightarrow S) (s_2 : A_2) \text{ rel} & \quad \Theta \vdash (s_1 : A_1) R (s_2 : A_2) \\
\Theta, (\forall \alpha. A_1) (\forall \alpha \rho \beta. R) (\forall \beta. A_2) \text{ rel} & \quad \Theta \vdash (t_1 : A_1) R (t_2 : A_2) \\
\Theta \vdash B_1 S B_2 \text{ rel} & \quad \Theta \vdash (t_1 : A_1) R (t_2 : A_2) \\
\Theta \vdash (t_1 \in \Theta) R (t_2 : B_2) & \quad \Theta \vdash (t_1 \in \Theta) S (t_2 : B_2)
\end{align*} \]

Relatedness rules:

\[ \begin{align*}
\Theta \vdash (x_1 : A_1) R (x_2 : A_2) & \quad ((x_1 : A_1) R (x_2 : A_2) \in \Theta) \\
\Theta \vdash (x_1 : A_1) R (x_2 : A_2) & \quad \vdash (t_1 : B_1) S (t_2 : B_2) \\
\Theta \vdash (\lambda x_1. t_1 : A_1) \vdash B_1 (R \rightarrow S) (\lambda x_2. t_2 : A_2) \text{ rel} & \quad \vdash (\lambda x_1. t_1 : A_1) R (t_2 : A_2) \\
\Theta \vdash (s_1 : A_1) \vdash B_1 (R \rightarrow S) (s_2 : A_2) \text{ rel} & \quad \vdash (s_1 : A_1) R (t_2 : A_2) \\
\Theta \vdash (s_1 : A_1) \vdash B_1 (R \rightarrow S) (s_2 : A_2) \text{ rel} & \quad \vdash (s_1 : A_1) R (s_2 : A_2)
\end{align*} \]

Parametricity rule:

\[ \begin{align*}
(\Gamma) \vdash (s : A) (A) (t : A) & \quad \Gamma \vdash s = t : A
\end{align*} \]

Fig. 2: The type system \( \lambda 2R \)
The operations $(-)_i$ (for $i \in \{0, 1\}$) on relational contexts:

$$
\langle \_ \rangle_i = \cdot
$$

$(\Theta, \alpha_1 \rho_2) = (\Theta)_i, \alpha_i$

$(\Theta, (x_1 : A_1) R(x_2 : A_2))_i = (\Theta)_i, x_i : A_i$

The operation $(-)$ on contexts and types:

$$
\langle \_ \rangle = \cdot
$$

$(\Gamma, \alpha) = (\Gamma), \alpha \rho^\alpha$

$(\Gamma \to B) = (\langle \Gamma \to \rangle (B))$

$(\Gamma, x : A) = (\Gamma) \langle x : A \rangle (A \langle x : A \rangle)$

$(\forall \alpha. A) = \forall \alpha \rho^\alpha. (A)$

Fig. 3: Syntactic reflexive graph structure

two such operations: $R[\alpha \rho \beta \to ASB]$ substitutes, in the relation $R$, the type $A$ for all left occurrences of $\alpha$, the type $B$ for all right occurrences of $\beta$ (which may itself be $\alpha$), and the relation $S$ for all occurrences of $\rho$; 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 $\Theta \vdash A_1 RA_2$ rel and $\Theta, \alpha_1 \rho_2 \alpha_2 \vdash (t_1 : B_1) S(t_2 : B_2)$ then $\Theta \vdash (t_1[\alpha_1 \mapsto A_1] : B_1[\alpha_1 \mapsto A_1]) S[\alpha_1 \rho_2 \alpha_2 \mapsto A_1 RA_2] (t_2[\alpha_2 \mapsto A_2] : B_2[\alpha_2 \mapsto A_2]).$

(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 $\Theta \vdash (s_1[x_1 \mapsto t_1] : B_1) S[x_1 \mapsto t_1, x_2 \mapsto t_2] (s_2[x_2 \mapsto t_2] : B_2)$.

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 $\lambda 2$. 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 $(\Gamma) \vdash (t : A)(A)(t : A)$.

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 $(\Gamma) \vdash (s : A)(A)(t : A)$. 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: \( \vdash (s : A)(t : A) \) 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. [8].

We outline the semantic interpretation of \( \text{\text{\L}2R} \). Given a comprehensive \( \text{\text{\L}2} \) parametricity graph, the contexts, types and terms of \( \text{\text{\L}2} \) are interpreted in the comprehensive \( \text{\text{\L}2} \) fibration \( p : T \rightarrow C \), as in Sect. [2]. In addition, we interpret a relational context \( \Theta \) as an object \( \llbracket \Theta \rrbracket \) of \( R(C) \), and a syntactic relation \( \Theta \vdash A \) \( B \) rel as a semantic relation \( \llbracket R \rrbracket_{\Theta} : \llbracket A \rrbracket_{\Theta_1} \leftrightarrow \llbracket B \rrbracket_{\Theta_2} \) in \( R(T)_{\Theta} \). The definitions of \( \llbracket \Theta \rrbracket \) and \( \llbracket R \rrbracket_{\Theta} \) interpret context extension, function space and universal quantification using the structure of the comprehensive \( \text{\text{\L}2} \) fibration \( p^R : R(T) \rightarrow R(C) \), where relation variables \( \alpha \beta \gamma \) are interpreted using the generic object of \( p^R \). For the inverse-image relation \( \Theta \vdash A \) \( (t_1 \times t_2)^{-1} ) R \) \( A \) \( B \) rel, we have that \( \llbracket t_1 \rrbracket_{\Theta_1} \) and \( \llbracket t_2 \rrbracket_{\Theta_2} \) determine maps \( \llbracket A \rrbracket_{\Theta_1} \rightarrow \llbracket B \rrbracket_{\Theta_1} \), and \( \llbracket A \rrbracket_{\Theta_2} \leftrightarrow \llbracket B \rrbracket_{\Theta_2} \) in \( T_{\Theta_1} \) and \( T_{\Theta_2} \) respectively. The fibration property of \( ( \nabla_1, \nabla_2 ) \) in \( \llbracket R \rrbracket_{\Theta} \) gives \( \llbracket (t_1 \times t_2)^{-1} R \rrbracket : \llbracket A \rrbracket_{\Theta_1} \leftrightarrow \llbracket A \rrbracket_{\Theta_2} \) as the inverse image of \( \llbracket R \rrbracket : \llbracket B \rrbracket_{\Theta_1} \rightarrow \llbracket B \rrbracket_{\Theta_2} \) along these maps.

In the above semantic interpretation, the comprehension property is needed in order to interpret a relational context \( \Theta \) as an object \( \llbracket \Theta \rrbracket \) of \( R(C) \), and essential use is made of this in the definition of \( \llbracket (t_1 \times t_2)^{-1} R \rrbracket \). 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 \( \text{\text{\L}2R} \).

The semantics is supported by soundness and completeness theorems.

**Theorem 8 (Soundness for \( \text{\text{\L}2R} \)).** In every comprehensive \( \text{\text{\L}2} \) parametricity graph:

(i) if \( \Gamma \vdash t_1 = t_2 : A \) then \( \llbracket t_1 \rrbracket_R = \llbracket t_2 \rrbracket_R \); and
(ii) if \( \Theta \vdash (t_1 : A_1) R (t_2 : A_2) \) then \( \llbracket t_1 \rrbracket_{\Theta_1} \times \llbracket t_2 \rrbracket_{\Theta_2} : \llbracket A \rrbracket_{\Theta} \rightarrow \llbracket R \rrbracket_{\Theta} \).

**Theorem 9 (Full completeness for \( \text{\text{\L}2R} \)).** There exists a comprehensive \( \text{\text{\L}2} \) parametricity graph satisfying the following.

(i) For every type \( \Gamma \vdash A \) type, every global point \( \llbracket A \rrbracket_R \) is the denotation \( \llbracket t \rrbracket_R \) of some term \( t : A \).
(ii) For all terms \( \Gamma \vdash t_1, t_2 : A \) satisfying \( \llbracket t_1 \rrbracket_R = \llbracket t_2 \rrbracket_R \), we have \( \Gamma \vdash t_1 = t_2 : A \).
(iii) For every relation \( \Theta \vdash A_1 R A_2 \) type, every global point \( \llbracket R \rrbracket_{\Theta} \) arises as \( \llbracket t_1 \rrbracket_{\Theta_1} \times \llbracket t_2 \rrbracket_{\Theta_2} \) 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 \( \text{\text{\L}2} \) 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 \( (A) \) 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 $\alpha$ in $A$). Then we have $\alpha \rho \alpha \vdash A(\alpha)(\langle A(\rho) \rangle)A(\alpha)$ rel. However, the independent handling of left and right variables in $\lambda 2R$ (forced by the semantic correspondence with comprehensive $\lambda 2$ parametricity graphs), means that the latter relation is equivalent to $\alpha \rho \beta \vdash A(\alpha)(\langle A(\rho) \rangle)A(\beta)$ rel, i.e., it is a family (indexed by relations $\rho$) of relations between different types. Indeed, the distinctness of left and right type variables means $\lambda 2R$ has no facility for formulating relations between open types and themselves. In particular, $\lambda 2R$ 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 $\lambda 2R$. Suppose $\Gamma \vdash f : A \to B$. Define: 

$$gr_{\ast}(f) \equiv [f \times id_B]^{-1}(B)$$

Clearly $\langle \Gamma \rangle \vdash A gr_{\ast}(f)B$ rel. Its defining property is that $(x : A) gr_{\ast}(f)(y : B)$ holds if and only if $(fx : B)(\langle y : B \rangle)(y : B)$. 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 $(x : A)\langle A \rangle(\langle w : A \rangle)$ and $y = fw$. Since, by [iii] of Lemma [7], $f$ maps $\langle A \rangle$-related values to $\langle B \rangle$-related values, $gr(f) \subseteq gr_{\ast}(f)$. However, because $(A)$ and $(B)$ 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 $\lambda 2R$ 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

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

where $i_B$ abbreviates $\lambda b. A_0 t b : B \to \forall \alpha. (B \to \alpha) \to \alpha$ and $(- \circ t_2)$ abbreviates $\lambda(v_j : B_j \to \alpha), \lambda(x_j : A_j), v_j(t_1 x_j)$, $\lambda 2R$ supports the derived rules below.

<table>
<thead>
<tr>
<th>$\Theta \vdash A_1R_{A_2}$ rel</th>
<th>$\Theta_1 \vdash t_1 : A_1 \to B_1$</th>
<th>$\Theta_2 \vdash t_2 : A_2 \to B_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Theta \vdash B_1([t_1 \times t_2]R)B_2$ rel</td>
<td>$\Theta \vdash (u_1 : A_1)R(u_2 : A_2)$</td>
<td>$\Theta \vdash (t_1 u_1 : B_1)\langle [t_1 \times t_2]R \rangle (t_2 u_2 : B_2)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\Theta \vdash C_1QC_2$ rel</th>
<th>$\Theta_1 \vdash v_1 : B_1 \to C_1$</th>
<th>$\Theta_2 \vdash v_2 : B_2 \to C_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Theta \vdash (u_1 : B_1)\langle [t_1 \times t_2]R \rangle (u_2 : B_2)$</td>
<td>$\Theta \vdash (v_1 \circ t_1 : A_1 \to C_1)\langle R \to Q \rangle (v_2 \circ t_2 : A_2 \to C_2)$</td>
<td></td>
</tr>
<tr>
<td>$\Theta \vdash (u_1 v_1 : C_1)Q(v_2 u_2 : C_2)$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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

\[
gr!_\Gamma(f) := [id_A \times f]_!(A).
\]

To understand the relationship between the two pseudograph relations we introduce some notation. Given \( R \) and \( S \) such that \( \Theta \vdash ARB \mathrel{rel} \) and \( \Theta \vdash ASB \mathrel{rel} \), 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) \( \langle \Gamma \rangle \vdash (s : A) gr_!(f)(t : B) \iff \Gamma \vdash f s = t : B \iff \langle \Gamma \rangle \vdash (s : A) gr_!(f)(t : B) \).

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 12.** In any comprehensive \( \lambda^2 \) parametricity graph, for every object \( W \) of \( R(C) \), the functor \( \langle \nabla^T_1, \nabla^T_2 \rangle |_{R(T)} : R(T)W \to T_{T_1W} \times T_{T_2W} \) is an opfibration.

### 6 Consequences of Parametricity

System \( \lambda^2R \) is strong enough to prove the familiar consequences of parametricity.

**Theorem 13 (Consequences of Parametricity).** System \( \lambda^2R \) proves:

(i) The unit (terminal) type can be encoded as \( 1 = \forall \alpha. \alpha \to \alpha \).

(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 \( 0 = \forall \alpha. \alpha \).

(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 \( \lambda^2R \) implies that analogous category-theoretic properties (which we do not state for lack of space) hold for comprehensive \( \lambda^2 \) 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 $\alpha$ in $T$, and $T(A)$ for the substitution $T[\alpha \mapsto A]$. If $\alpha$ 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 $\Gamma'$ extends $\Gamma$, 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., $(T)(\langle \alpha \rangle)$ (by which we mean the substitution $(T)[\alpha \mapsto \langle \alpha \rangle]$) syntactically coincides with $(T)(A))$. Also, the functorial action lifts to relations: if $\Theta \vdash (f : A \to B)(R \to S)(f' : A' \to B')$, where $\Theta$ extends $(\Gamma)$, then:

\[ \Theta \vdash (T(f) : T(A) \to T(B))(\langle T(R) \to (T(S))(T(f') : T(A') \to T(B')) \right) . \]

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

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

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

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

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:

\[
\begin{array}{ccc}
T(A) & \xrightarrow{T(h)} & T(B) \\
\downarrow a & & \downarrow b \\
A & \xrightarrow{h} & B
\end{array}
\quad\quad
\begin{array}{ccc}
T(A) & \xrightarrow{T(h')} & T(B) \\
\downarrow a' & & \downarrow b' \\
A & \xrightarrow{h'} & B
\end{array}
\]

where the diagrams are given by terms, in a context $\Gamma'$ extending $\Gamma$, 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.

\[ \langle \Gamma' \rangle \vdash (a : T(A) \to A)(\langle T(gr_*(h)) \to gr_*(h)⟩)(b : T(B) \to B) \]

\[ \langle \Gamma' \rangle \vdash (a' : A \to T(A))(gr_!(h') \to (T)(gr_!(h')))(b' : B \to T(B)) \]

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 \( \lambda 2R \) the property of having a syntax in which function space and universal quantification are basic constructions on relations. Indeed \( \lambda 2R \) 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 \( \lambda 2R \)
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 \( \lambda 2R \) (and System P) in two main ways. The first is that, in System R, every type \( A \) has an associated identity relation \( A^* \). A key rule of System R (written in our notation) is that \( \Theta \vdash x.A^*x \), \( \Theta \) anywhere in relational context \( \Theta \). This rule breaks the independence between left and right variables in the relational judgements of \( \lambda 2R \). (For example, property [1] 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 \( \lambda 2R \) (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 \( \lambda 2R \).

In fact, the interplay between models and syntax could be pushed much further than in the present paper. By adding primitive product types to \( \lambda 2 \) and \( \lambda 2R \), 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 \( \lambda 2R \) (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 reconstructed (albeit in a nontrivial way) as a comprehensive \( \lambda 2 \) 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 $\lambda_2$, 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 ($\lambda_2$) to the relational version ($\lambda_2R$) 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