

Generic Programming with Extensible Data Types

Or, Making Ad Hoc Extensible Data Types Less Ad Hoc

ALEX HUBERS, The University of Iowa, USA

J. GARRETT MORRIS, The University of Iowa, USA

We present a novel approach to generic programming over extensible data types. Row types capture the structure of records and variants, and can be used to express record and variant subtyping, record extension, and modular composition of case branches. We extend row typing to capture generic programming over rows themselves, capturing patterns including lifting operations to records and variations from their component types, and the duality between cases blocks over variants and records of labeled functions, without placing specific requirements on the fields or constructors present in the records and variants. We formalize our approach in System $R\omega$, an extension of $F\omega$ with row types, and give a denotational semantics for (stratified) $R\omega$ in Agda.

CCS Concepts: • **Theory of computation** → **Type theory**; • **Software and its engineering** → **Data types and structures**.

Additional Key Words and Phrases: generic programming, extensible data types, row types, row polymorphism, qualified types.

ACM Reference Format:

Alex Hubers and J. Garrett Morris. 2023. Generic Programming with Extensible Data Types: Or, Making Ad Hoc Extensible Data Types Less Ad Hoc. *Proc. ACM Program. Lang.* 7, ICFP, Article 201 (August 2023), 29 pages. <https://doi.org/10.1145/3607843>

1 INTRODUCTION

The goal of extensible data types is to bring type safety to modular software development. Row types [Rémy 1992; Wand 1987] are one approach to that goal. Rows express the structure of records or variants; row polymorphism captures properties like subtyping while maintaining a purely parametric approach to typing. Row typing was originally designed to model object-oriented inheritance, but its applications include: extensible variants in OCaml [Garrigue 1998]; extensible effects [Lindley and Cheney 2012]; typing algebraic effects and handlers [Hillerström and Lindley 2016; Leijen 2014, 2017]; and, extensible protocols in session types [Lindley and Morris 2017].

This paper explores generic programming over rows. Consider defining equality functions for extensible records. Of course, given a particular set of fields, and knowledge of how to compare the field types, existing row type systems can express the equality function for records of those fields. Even with metaprogramming support, however, having to explicitly define equality functions for each record type (and each extension of a record type) creates a significant burden for programmers—a disadvantage for an approach designed to encourage this style of programming! Moreover, approaches that depend on particular sets of fields cannot extend to row polymorphism, a key

Authors' addresses: Alex Hubers, Department of Computer Science, The University of Iowa, 14 MacLean Hall, Iowa City, Iowa, USA, alexander-hubers@uiowa.edu; J. Garrett Morris, Department of Computer Science, The University of Iowa, 14 MacLean Hall, Iowa City, Iowa, USA, garrett-morris@uiowa.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2023 Copyright held by the owner/author(s).

2475-1421/2023/8-ART201

<https://doi.org/10.1145/3607843>

$$\begin{array}{l|l}
Eq : \star \rightarrow \star & eq_{\Sigma} : \forall z : R^{\star}. \Pi(Eq\ z) \rightarrow Eq(\Sigma z) \\
Eq = \lambda t. t \rightarrow t \rightarrow \text{Bool} & eq_{\Sigma} = \lambda d\ v\ w. \text{ana } l\ l\ y. (\text{case } l\ (\lambda x. \text{sel } d\ l\ x\ y) \vee \text{const False})\ v)\ w
\end{array}$$

Fig. 1. Comparing extensible variants in $R\omega$

contributor to the expressiveness of row types. While we could express the extension of a particular record type, we could not (modularly) express that such an extension supports equality.

We propose novel record and variant operations, generic in the particular labels that appear in those records and variants, and realize these operations in System $R\omega$, a core calculus that extends System $F\omega$ with row types based on ROSE [Morris and McKinnon 2019]. Consider the equality function for extensible variants: if we know how to compare the values at each constructor in two variants, we ought to know how to compare the variants. Figure 1 captures this idea in $R\omega$; we have elided type abstractions, applications, and annotations on bound variables, as they can be inferred from the given type signatures. Type operator Eq maps types to equality operators for those types. Function eq_{Σ} compares two variant values v and w , given a record d of comparison operators for their fields. Suppose that z is instantiated with the row $\{a \triangleright \text{Int}, b \triangleright \text{List Bool}\}$: d will be a record of comparison functions $\Pi\{a \triangleright \text{Int} \rightarrow \text{Int} \rightarrow \text{Bool}, b \triangleright \text{List Bool} \rightarrow \text{List Bool} \rightarrow \text{Bool}\}$, and v and w will each be variants $\Sigma\{a \triangleright \text{Int}, b \triangleright \text{List Bool}\}$. (We follow Pottier and Rémy [2005] in implicitly lifting operators on types, like Eq , to the corresponding operators on rows.) Our key novelty is the *ana* combinator: $\text{ana } f\ w$ analyzes variant w , calling f with its constructor label l and contents y . With these in hand, we can then rely on the variant branching combinator (\vee) of ROSE: in *case* v is constructed with label l , we *select* from d the l -labeled function, and use it to compare the contents x and y of the two variants; otherwise, regardless of the contents of v , we can return False. We will return to each component of this definition in the remainder of the paper.

The generic operations in $R\omega$ build on the row type theory ROSE. ROSE is distinguished from other row type theories by two features. First, ROSE uses qualified types [Jones 1994] to capture the structure of row types, rather than incorporating the structure of rows directly into the types of records and variants. This indirection makes it possible to capture structural invariants in ROSE that are difficult or impossible to capture in other row type systems. We rely on this expressiveness in typing the combinators in $R\omega$: the function argument to *ana*, for example, is typed given the assumption that l labels a value of type u in row z . Second, ROSE builds on a general account of rows as partial monoids, encompassing a variety of different row type theories in the literature. While we will fix a particular theory of rows in our formalization of $R\omega$, we will show how our account would generalize other theories of rows as well.

One way to realize the behavior of eq_{Σ} would be to treat (hashes of) labels as keys at runtime: variants would be labeled by these keys and records would be dictionaries over keys. While direct, this approach is neither practically nor theoretically sound: we would be disappointed to learn that selecting a field from a record was not a constant time operation, and it would be difficult to show that well-typed operations only relied on keys that were dynamically present in records or case blocks. We will show that $R\omega$ has a type-safe implementation with no runtime comparison or manipulation of labels, guided by the use of predicates in the types of the generic operations. To do so, we will give a denotational semantics for universe-stratified $R\omega$ typing derivations in Agda, in which rows are interpreted as functions from finite naturals to types, row inclusions and combinations are witnessed by maps from finite naturals to finite naturals, and we have static guarantees that our indexing of records and variants is well-typed.

To summarize, this paper contributes:

- The extension of ROSE to generic programming over rows (§3), particularly the design of combinators that express generic transformations of row-typed products and sums;
- A formalization of our approach in the $R\omega$ calculus (§4), which extends System $F\omega$ with ROSE-style row typing, first-class labels, and generic programming over rows; and,
- The denotation of $R\omega$ derivations in Agda (§5), showing that $R\omega$ is sound and need not introduce runtime manipulation or comparison of labels.

We begin with a review of extensible datatypes in ROSE (§2) and conclude with discussions of related (§6) and future (§7) work.

2 EXTENSIBLE DATATYPES AND THE ROSE TYPE THEORY

The goal of row typing is to support type-safe extensible data types. This section gives an intuitive overview of row typing, and the ROSE type system in particular, preparatory to its extension in the following section.

2.1 The Need for Extensibility

Existing functional language type theories are remarkably expressive, and further additions are rightly viewed with some suspicion. We begin with two examples of the additional value of extensible data types.

The expression problem. Wadler [1998] describes the expression problem as “a new name for an old problem”. Consider an abstract data type along with several operations. For example, we could have a simple type for arithmetic expressions, consisting of constants and sums, along with an operations to reduce expressions to integer values. The challenge is to extend this in two dimensions—say, by adding a new constructor for products, and a new operation to print expressions as character strings—without rewriting or recompiling existing code, and without compromising type safety. In modern functional languages, adding new operations is easy, but adding new constructors requires changing the original type definition and all the existing definitions. In object-oriented languages, adding new cases is easy but adding new operations requires changing the base class and all of its inheritors. Programmers in either camp must resort to encoding tricks to capture the remaining case, making code more difficult to read and maintain.

Modular transformations. The expression problem may not seem entirely compelling: why artificially restrict a common refactoring operation to preserve existing code? As an alternative view of the same problem, consider desugaring or optimization passes in a compiler. We might hope to limit many of our passes to operating on a subset of the whole language [Keep and Dybvig 2013; Sarkar et al. 2004]: for example, a single pass might resolve infix applications, while not changing the remainder of the syntax tree. In writing these passes, we would like to make them generic over the untouched (or only recursively transformed) parts of the syntax tree. This will both make the compiler more readable and maintainable, and provide type-based guarantees of the limited scope of these passes. This problem is essentially the dual of the expression problem: instead of planning to extend our AST, we hope to write passes without fixing most of the AST.

2.2 Row Types

Consider the type of a function that selects the field x from a record—we might write this function $\lambda r. sel\ r\ x$, where sel is our function for selecting record fields, and we use the teletype font to distinguish label constants from label variables. We can imagine many record types which might contain an x field—points on a plane, or in space; pixels on a screen; nodes for lambda expressions in the AST of a functional programming language—and in each case, x might have a different type.

A general type for this function ought to encompass all its possible arguments, associating each with the corresponding result type. This problem, along with its dual for variants, is the starting point for row type systems.

Rows and row polymorphism. A row is an association of labels to types. For example, we write $\{x \triangleright \text{Double}, y \triangleright \text{Double}\}$ for the row that associates both the labels x and y with the type of double-precision floating point numbers. Record and variant types are constructed from rows; for example, a type for Cartesian coordinates is $\Pi\{x \triangleright \text{Double}, y \triangleright \text{Double}\}$, while a more general type for points is $\Sigma\{\text{Cart} \triangleright \Pi\{x \triangleright \text{Double}, y \triangleright \text{Double}\}, \text{Polar} \triangleright \Pi\{r \triangleright \text{Double}, \text{theta} \triangleright \text{Double}\}\}$.

Just introducing rows gets us little closer to solving our initial problem: we can say that $\lambda r. \text{sel } r \ x$ could have type $\Pi\{x \triangleright \text{Double}\} \rightarrow \text{Double}$ or $\Pi\{x \triangleright \text{Int}, y \triangleright \text{Int}\} \rightarrow \text{Int}$, but these are not a general account of its behavior. Instead, this function should have a polymorphic type. In many row type systems [Rémy 1989; Wand 1987], its type would be written similarly to $\forall t \ z. \Pi\{x \triangleright t \mid z\} \rightarrow t$. The syntax $\{x \triangleright t \mid z\}$ denotes the *extension* of row z with the field $x \triangleright t$. As a whole, the type denotes a function from a record containing any fields z , and also $x \triangleright t$, to a value of type t . Similarly, a function that added a new x field to an existing record could be given a type like $\forall t \ z. t \rightarrow \Pi z \rightarrow \Pi\{x \triangleright t \mid z\}$.

This account of row types leaves several questions. First: in the types above, can the instantiation of z already include an association for x ?

- Wand [1987] allows free instantiation of z ; extension is then interpreted as *overwriting* the existing meaning of fields (in both types and terms).
- Rémy [1989] uses the kind system to preclude conflicting meaning of fields, but must introduce a new kind to capture functions which can either overwrite or extend objects.
- Berthomieu and le Monières de Sagazan [1995] and Leijen [2005] allow free instantiation of z , and interpret extension as *shadowing* the existing meaning of fields, such that the original meaning can be recovered later.

Second: does this account generalize from single field extension to arbitrary concatenation of objects? For example, given two records, one of location data and one of color data, can we combine them to form a single record of colored location (or located color) data?

Polymorphism and predicates. Wand [1989] proposes the following term as a test of row type systems with record concatenation:

$$\lambda m \ n. \text{sel } (m \# n) \ k$$

Here m and n are arbitrary records, and the function projects the field k from their concatenation. (In Wand's original example, m and n were records of method implementations, k is a method name, and the term as a whole models multiple inheritance.) The crux of the problem is that if we have to assign a type to either m or n that already commits to field k , then we have over-specified the behavior of the function. On the other hand, if we do not commit to either m or n containing field k , then how can we be sure the function is well-defined at all?

This is the starting point for the ROSE type theory [Morris and McKinnin 2019]. Instead of capturing the structure of rows directly in the types of records and variants, ROSE captures them using predicates in qualified types. For example, in ROSE, the type of the x -selection function would be expressed as

$$\forall t \ z. \{x \triangleright t\} \lesssim z \Rightarrow \Pi z \rightarrow t$$

That is: this is a function that maps z -shaped records to t results, for any types t and z , *such that* the singleton row $\{x \triangleright t\}$ is contained in z . ROSE supports concatenation of records via predicates as well. The type for Wand's example in ROSE is:

$$\forall t \ z_1 \ z_2 \ z_3. (z_1 \odot z_2 \sim z_3, \{k \triangleright t\} \lesssim z_3) \Rightarrow \Pi z_1 \rightarrow \Pi z_2 \rightarrow t$$

$$\begin{array}{c}
\frac{\Gamma \vdash M : \tau}{\Gamma \vdash \ell \triangleright M : \ell \triangleright \tau} \quad \frac{\Gamma \vdash M : \Pi \rho_1 \quad \Gamma \Vdash \rho_2 \lesssim \rho_1}{\Gamma \vdash \text{prj } M : \Pi \rho_2} \quad \frac{\Gamma \vdash M_1 : \Pi \rho_1 \quad \Gamma \vdash M_2 : \Pi \rho_2 \quad \Gamma \Vdash \rho_1 \odot \rho_2 \sim \rho_3}{\Gamma \vdash M_1 \# M_2 : \Pi \rho_3} \\
\frac{\Gamma \vdash M : \ell \triangleright \tau}{\Gamma \vdash M / \ell : \tau} \quad \frac{\Gamma \vdash M : \Sigma \rho_1 \quad \Gamma \Vdash \rho_1 \lesssim \rho_2}{\Gamma \vdash \text{inj } M : \Sigma \rho_2} \quad \frac{\Gamma \vdash M_1 : \Sigma \rho_1 \rightarrow \tau \quad \Gamma \vdash M_2 : \Sigma \rho_2 \rightarrow \tau \quad \Gamma \Vdash \rho_1 \odot \rho_2 \sim \rho_3}{\Gamma \vdash M_1 \nabla M_2 : \Sigma \rho_3 \rightarrow \tau}
\end{array}$$

Fig. 2. Typing of record and variant operations in ROSE

That is: this is a function that maps a z_1 -shaped record and a z_2 shaped record to a t result, *such that* z_1 and z_2 can be concatenated to give row z_3 , and z_3 contains the singleton row $\{k \triangleright t\}$. This type captures the full generality of Wand’s challenge: we do not overconstrain either m or n to always provide field k , but still guarantee that the projection will always be well-defined.

Figure 2 gives the typing rules for the record and variant operations in ROSE. The projection and injection operators are the generalizations of record selection and variant construction; each relies on being able to prove that one row is contained in another. Branching ($M_1 \nabla M_2$) is dual to record concatenation ($M_1 \# M_2$): it combines eliminators for two variants to give the eliminator for their combination. As with concatenation, it relies on being able to prove that the two smaller rows can be combined.

Theories of rows. As ROSE does not commit directly to the structure of rows, but abstracts their structure via the *containment* ($\rho_1 \lesssim \rho_2$) and *combination* ($\rho_1 \odot \rho_2 \sim \rho_3$) predicates, it can be adapted to any of the different notions to row extension:

- To capture *non-overlapping* rows: we stipulate that $\rho_1 \odot \rho_2 \sim \rho_3$ is only satisfiable when ρ_1 and ρ_2 have no fields in common. For this approach, we can define $\rho_1 \lesssim \rho_2$ to hold either when there is some ρ' such that $\rho_1 \odot \rho' \sim \rho_2$ or when $\rho' \odot \rho_1 \sim \rho_2$.
- To capture *overwriting*: $\rho_1 \odot \rho_2 \sim \rho_3$ is always satisfiable, where ρ_3 reflects ρ_1 for any labels that appear in both. We can define $\rho_1 \lesssim \rho_2$ to hold exactly when there is ρ' such that $\rho_1 \odot \rho' \sim \rho_2$. (On the other side, when $\rho' \odot \rho_1 \sim \rho_2$, we cannot necessarily recover fields in ρ_1 from the combination ρ_2 because they may have been overwritten by fields in ρ' .)
- To capture *shadowing*: $\rho_1 \odot \rho_2 \sim \rho_3$ is always satisfiable, and we get two containment predicates, $\rho_1 \lesssim_L \rho_2 \iff \rho_1 \odot \rho' \sim \rho_2$ and $\rho_1 \lesssim_R \rho_2 \iff \rho' \odot \rho_1 \sim \rho_2$, with corresponding injection and projection functions.

ROSE itself is defined generically over a *row theory*, which defines the underlying structure of rows and interpretation of row predicates. So, ROSE encompasses all of the above cases, as well as both simpler (e.g., unlabeled) and more complex (e.g. modules) cases.

2.3 Open Problems in Extensibility

Despite ROSE’s expressiveness, it is still limited in how it describes individual rows. ROSE can capture the structure of rows, but it has no predicates that capture properties of the types in a row. This limitation has several consequences.

If we know that every type in a variant or record supports equality comparisons, we should expect that the variant or record supports equality comparison as well. However, even expressing this problem is not possible in ROSE—the constraint we need to express is on the types that appear in the row, not on the structure of the row itself. The problem recurs when considering higher-order polymorphism. Recall the example of modular AST transformations (§2.1). To maximize flexibility and readability, the pass that transforms infix to prefix applications should not constrain the remainder of the syntax tree. However, this transformation is not only applied at the top level

of expressions or definitions; it must also be applied recursively, regardless of the other nodes in the AST. This, in turn, implies some constraint (such as functoriality) on the remainder of the AST, which cannot be captured in ROSE.

We know that records and variants enjoy strong duality properties: a case expression eliminating a variant corresponds to a record of functions, containing one (appropriately typed) function for each branch in the case expression. This duality is not just of theoretical interest. For example, in implementing a system of algebraic effects and handlers [Plotkin and Power 2003; Plotkin and Pretnar 2009], we could represent effectful computations as abstract syntax trees over operations, and handlers as records of implementations of those operations. We might then hope to define a general handling combinator, which combines an effectful computation with an appropriate handler. However, we cannot implement this operation in ROSE: while we can use the same row variable to describe both records and variants (i.e., both computations and their handlers), the branching and projection operators all refer to specific labels.

Existing row type theories address some of these problems. Blume et al. [2006] distinguishes case blocks from functions, and realizes case blocks by records of functions in their semantics. However, this step in the semantics is not available to programmers. Pottier and Rémy [2005] implicitly lift operations on types to operations on rows: if z is a row of associations $\ell_i \triangleright \tau_i$, then $z \rightarrow v$ is the row of associations $\ell_i \triangleright \tau_i \rightarrow v$. They further postulate an operation *apply* which applies a record of functions to a record of (identically labeled) arguments, producing a record of results. However, this operation is treated as a primitive extension of their calculus. Chlipala [2010] includes a mapping operator on records in a calculus based on $F\omega$, generalizing the lifting of Pottier and Rémy, and provides type-directed generation of record folding operations. He does not consider variants in his approach; moreover, it is not immediately clear that folds, and their duals for variants, would be sufficient to capture the open problems we identify.

3 GENERIC PROGRAMMING IN $R\omega$

System $R\omega$ generalizes ROSE in two dimensions. ROSE imposes Hindley-Milner constraints on typing; $R\omega$ is based on System $F\omega$ extended with qualified types, and so supports first-class polymorphism and general type operators. More significantly, the record and variant operations in ROSE are all specific to concrete labels or sets of labels; $R\omega$ introduces label-generic combinators. This section introduces $R\omega$ by example.

Through the majority of this section, we will assume *simple rows*: labels are restricted to appear at most once in a given row, row combination is commutative (and so there is a single containment operator), and $\rho_1 \odot \rho_2 \sim \rho_3$ is unsatisfiable if ρ_1 and ρ_2 contain any of the same fields. This is the most common approach to typing records and variants and rules out many unexpected behaviors. At the end of the section (§3.5), we will discuss the specific challenges in extending our development to a non-commutative row theory.

3.1 First-Class Labels

In ROSE, labels exist in types, but not in terms. The construction $(\ell \triangleright M)$ and destruction (M/ℓ) terms, which are overloaded for both singleton records and variants, are each essentially infinite families of terms, one for each label. To support label-generic operations, however, we will need to make labels first-class citizens in the term language as well as the type language.

To do so, we follow the approach used by Gaster and Jones [1996] and Sulzmann [1997]. We have added a singleton type constructor $[-]$ to $R\omega$: if ℓ is a label type, then $[\ell]$ is the corresponding singleton type. (For a label constant L , we also write L for the unique inhabitant of $[L]$.) First-class labels allow us to abstract several common patterns in ROSE. For example, to select an individual field from a record, we first apply *prj* to project a singleton record and then use the singleton

deconstruction operator. ROSE introduced syntactic sugar for this pattern; in contrast, we can define the selection function directly in $R\omega$ by:

$$\begin{aligned} \text{sel} &: \forall l : L, t : \star, z : R^\star. \{l \triangleright t\} \lesssim z \Rightarrow \Pi z \rightarrow [l] \rightarrow t \\ \text{sel} &= \Lambda(l : L) (t : \star) (z : R^\star). \lambda(r : \Pi z) (g : [l]). \text{prj } r/g \end{aligned}$$

(Note that predicate abstraction and application remain implicit in $R\omega$.) The type abstractions and annotations in this example, and most of the following, can be determined from the type signatures alone, so we will generally omit them:

$$\text{sel} = \lambda r l. \text{prj } r/l$$

Row type systems are frequently forced to distinguish between record extension (which adds new fields to existing records) and record update (which changes the value—and possibly type—of an existing field in a record), because their types impose different requirements on the input record type. Rémy [1989] introduces *presence polymorphism*, allowing a single term to play both roles at the cost of additional type system complexity. A single term that captures both in $R\omega$:

$$\begin{aligned} \text{upd} &: \forall l : L, t, u : \star, z_1, z_2 : R^\star. z_1 \lesssim \{l \triangleright t\} \Rightarrow [l] \rightarrow u \rightarrow \Pi(z_1 \odot z_2) \rightarrow \Pi(\{l \triangleright u\} \odot z_2) \\ \text{upd} &= \lambda l u r. (l \triangleright u) \# \text{prj } r \end{aligned}$$

We treat \odot as a *partial* type constructor [Ingle et al. 2022; Jones and Diatchki 2008]: we write $\rho_1 \odot \rho_2$ as a type to denote a fresh type variable z under the constraint $\rho_1 \odot \rho_2 \sim z$. Row z_1 is either the empty row or the singleton row mapping l to t ; row z_2 is constrained to combine with $\{l \triangleright t\}$, so cannot contain label l . The input record, of type $\Pi(z_1 \odot z_2)$ may contain field l (depending on the choice of z_1); the output record definitely contains l , mapped to type u .

First-class labels are also useful for capturing programming patterns with variants. We can define a generic function for constructing variants:

$$\begin{aligned} \text{con} &: \forall l : L, t : \star, z : R^\star. \{l \triangleright t\} \lesssim z \Rightarrow [l] \rightarrow t \rightarrow \Sigma z \\ \text{con} &= \lambda l x. \text{inj } (l \triangleright x) \end{aligned}$$

The base case for the branching operator ∇ is a function that maps a singleton variant to a result. We can capture this pattern as well:

$$\begin{aligned} \text{case} &: \forall l : L, t : \star, u : \star. [l] \rightarrow (t \rightarrow u) \rightarrow \Sigma\{l \triangleright t\} \rightarrow u \\ \text{case} &= \lambda l f x. f (x/l) \end{aligned}$$

Representing Booleans as $\text{Bool} = \Sigma\{\text{True} \triangleright \Pi\}, \{\text{False} \triangleright \Pi\}\}$ (syntactic sugar for $\Sigma(\{\text{True} \triangleright \Pi\}) \odot \{\text{False} \triangleright \Pi\}$)), we could then define the usual conditional by:

$$\begin{aligned} \text{ifte} &: \forall t : \star. \text{Bool} \rightarrow t \rightarrow t \rightarrow t \\ \text{ifte} &= \lambda b t f. (\text{case True } (\lambda u. t) \nabla \text{case False } (\lambda u. f)) b \end{aligned}$$

Perhaps most surprisingly, while ROSE lacked syntax or types for first-class labels, adding them does not require extending its semantics in any non-trivial way. The necessary information for the *sel* function, for example, is already captured entirely by the predicate $\{l \triangleright t\} \lesssim z$. The value of type $[l]$ provides no additional information—as you would expect for a value of a singleton type!

3.2 The Duality of Records and Variants

We begin our exploration of generic programming over rows with the duality between records and variants. This duality is foundational to row type systems in general, and to ROSE in particular. Its introduction rule for variants and the elimination rule for records are clearly dual, and the rules for concatenating variant eliminators and concatenating records are nearly as evidently dual. (To make

$$\begin{array}{l|l}
\text{reify} : \forall z : R^\star, t : \star. (\Sigma z \rightarrow t) \rightarrow \Pi(z \rightarrow t) & \text{reflect} : \forall z : R^\star, t : \star. \Pi(z \rightarrow t) \rightarrow (\Sigma z \rightarrow t) \\
\text{reify} = \lambda f. \text{syn } (\lambda l x. f \text{ (con } l x)) & \text{reflect} = \lambda d w. \text{ana } (\lambda l u. \text{sel } d l u) w
\end{array}$$

Fig. 3. Witnessing the duality of records and variants

the duality more explicit, one could have defined a rule for combining record constructors—from $\tau \rightarrow \Pi z_1$ and $\tau \rightarrow \Pi z_2$, obtain $\tau \rightarrow \Pi(z_1 \odot z_2)$ —but this seems to obtain theoretical elegance at the cost of usability.) In fact, we can witness this duality in *ROSE*, but only for concrete rows. For example, we can define the following operations for the Boolean type:

$$\begin{aligned}
\text{Cases}_B &: \star \rightarrow R^\star \\
\text{Cases}_B &= \lambda t. \{\text{True} \triangleright \Pi\{\} \rightarrow t, \text{False} \triangleright \Pi\{\} \rightarrow t\} \\
\text{reify}_B &: \forall t : \star. (\text{Bool} \rightarrow t) \rightarrow \Pi(\text{Cases}_B t) \\
\text{reify}_B &= \lambda f. (\text{True} \triangleright f \text{ (con True ())}) \# (\text{False} \triangleright f \text{ (con False ())}) \\
\text{reflect}_B &: \forall t : \star. \Pi(\text{Cases}_B t) \rightarrow \text{Bool} \rightarrow t \\
\text{reflect}_B &= \lambda d. (\text{case True (sel } r \text{ True)}) \vee (\text{case False (sel } r \text{ False)})
\end{aligned}$$

The type Cases_B abbreviates operations over the constructors of the Boolean type. The reify_B function transforms a function that scrutinizes a Boolean value into a record of functions, one for the *True* case and one for the *False* case; dually, the reflect_B function uses such a record of functions to scrutinize a Boolean value. (We write $()$ for the unique value of the $\Pi\{\}$ type.) Knowing the constructors of the Boolean type is essential to writing this example; while such functions exist for any variant type in *ROSE*, their definition would have to be repeated for each type.

In *Rω*, we can write generic versions of these operators, applicable to any variant type and the corresponding record of cases, as shown in Figure 3. The types of reify and reflect rely on lifting operations on types to operations on rows: if z is the row of types $\ell_i \triangleright \tau_i$, then $z \rightarrow t$ is the row of types $\ell_i \triangleright v_i \rightarrow t$. In reify_B and reflect_B , we relied on concrete constructors in two places: when deconstructing a Boolean value in reflect_B , and when building the record of constructors in reify_B . *Rω* provides label-generic versions of these two operations, one for *analyzing* variants and a dual operator for *synthesizing* records. Here is our first attempt at their typing rules:

$$\begin{aligned}
(\text{T-ana}_1) \quad & \frac{\Gamma \vdash \rho : R^\star \quad \Gamma \vdash M : \forall l : L, u : \star. \{l \triangleright u\} \lesssim \rho \Rightarrow [l] \rightarrow u \rightarrow \tau}{\Gamma \vdash \text{ana } M : \Sigma \rho \rightarrow \tau} \\
(\text{T-syn}_1) \quad & \frac{\Gamma \vdash \rho : R^\star \quad \Gamma \vdash M : \forall l : L, u : \star. \{l \triangleright u\} \lesssim \rho \Rightarrow [l] \rightarrow u}{\Gamma \vdash \text{syn } M : \Pi \rho}
\end{aligned}$$

We write R^κ for the kind of rows over types of kind κ . To avoid a sea of metavariables, we combine kinding and typing assertions in Γ ; the judgment $\Gamma \vdash \rho : R^\star$ is a kinding assertion on ρ , and $\Gamma \vdash \text{ana } M : \Sigma \rho \rightarrow \tau$ is a typing assertion on $\text{ana } M$.

In $\text{ana } M$, the body M is a label-generic version of the cases in a branch expression: given a label l , a type u , and evidence that $\{l \triangleright u\}$ appears in ρ , M consumes a single case—(a witness for) the constructor, and its contents—and produces a result of type τ . If M can do so for any constructor appearing in ρ , then $\text{ana } M$ can consume a value of $\Sigma \rho$ to produce a result of type τ . We use *ana* in implementing reflect . Given the constructor label l and contents u of an arbitrary variant value w , we invoke the l -labeled entry from the record d with argument u . Again, lifting plays a central role: from $\{l \triangleright u\} \lesssim z$, we can conclude that $\{l \triangleright u \rightarrow t\} \lesssim z \rightarrow t$, and so $\text{sel } d l$ is a $u \rightarrow t$ function.

In $\text{syn } M$, the body M is a label-generic version of the components of a concatenation expression: given a label l , a type u , and evidence that $\{l \triangleright u\}$ appears in ρ , M produces a value of type u . If M can do so for each label appearing in ρ , then $\text{syn } M$ can produce a record of type $\Pi\rho$. We use syn in implementing *reify*. In the body, we have access to $f : \Sigma z \rightarrow t$. We build a new function $u \rightarrow t$, which wraps its argument in constructor l and then invokes f . Lifting plays a similar role to its role in *reflect*: as $\{l \triangleright u\} \lesssim z$, the result type includes $l \triangleright u \rightarrow t$.

3.3 Transformations

Next, we consider generic transformations on extensible types.

Type-preserving maps. We begin with type-preserving mappings, such as reversing each field of a record of lists. Here is the version for records; the version for variants is nearly identical.

$$\begin{aligned} \text{map}'_{\Pi} &: \forall z : R^{\star}. (\forall l : L, u : \star. \{l \triangleright u\} \lesssim z \Rightarrow [l] \rightarrow u \rightarrow u) \rightarrow \Pi z \rightarrow \Pi z \\ \text{map}'_{\Pi} &= \lambda f r. \text{syn } (\lambda l. f l (\text{sel } r l)) \end{aligned}$$

The mapped function is label-generic: for any label l and type u appearing in z , the function transforms the old u value into a new u value. Given such a function f and a record r , we synthesize a new record in which each field l contains the result of f applied to the old field and its label.

Type-transforming maps. The far more interesting case is type-transforming mappings, such as transforming a record of lists into a record of their lengths. The challenge here is not defining the term (in fact, it will turn out to appear identical to the previous term), but rather to find an appropriately expressive type. Suppose that we have type constructors $\text{List} : \star \rightarrow \star$ and $\text{Int} : \star$, such that the *length* function has type $\forall t : \star. \text{List } t \rightarrow \text{Int}$. We might then imagine that the pointwise length function on records would have a type like

$$\forall z : R^{\star}. \Pi(\text{List } z) \rightarrow \Pi(\text{const Int } z)$$

where $\text{const} : \star \rightarrow \star \rightarrow \star$ is the expected constant operator, on types. In the input type, we lift the type constructor List over the row z ; this allows us to capture the idea of a row of list types. In the output type, we lift const Int over z ; this replaces each type in z by Int . Instantiating this type with the concrete row $\{a \triangleright \text{Bool}, b \triangleright \text{Char}\}$ would give

$$\Pi\{a \triangleright \text{List Bool}, b \triangleright \text{List Char}\} \rightarrow \Pi\{a \triangleright \text{Int}, b \triangleright \text{Int}\}$$

Of course, we cannot inhabit this type with a term based on map'_{Π} , as the input and output types are not identical. More seriously, however, it is not clear how we could inhabit it with any term based on our previous typing rule for syn . The only types in the output row are Int , and it is not clear how we could reconstruct an application of *length* to a field of the input row given only the information that $l \triangleright \text{Int}$ appears in the output row.

Our solution is to generalize the types of *ana* and *syn* to incorporate a type operator ϕ :

$$\begin{aligned} (\tau\text{-ana}_2) \quad & \frac{\Gamma \vdash \rho : R^{\kappa} \quad \Gamma \vdash \phi : \kappa \rightarrow \star \quad \Gamma \vdash M : \forall l : L, u : \kappa. \{l \triangleright u\} \lesssim \rho \Rightarrow [l] \rightarrow \phi u \rightarrow \tau}{\Gamma \vdash \text{ana}_{\phi} M : \Sigma(\phi \rho) \rightarrow \tau} \\ (\tau\text{-syn}_2) \quad & \frac{\Gamma \vdash \rho : R^{\kappa} \quad \Gamma \vdash \phi : \kappa \rightarrow \star \quad \Gamma \vdash M : \forall l : L, u : \kappa. \{l \triangleright u\} \lesssim \rho \Rightarrow [l] \rightarrow \phi u}{\Gamma \vdash \text{syn}_{\phi} M : \Pi(\phi \rho)} \end{aligned}$$

Differences from the previous rules are shaded. We now allow ρ to range over rows of arbitrary kind κ —we will make use of this in capturing functoriality later in the section—and require that ϕ be a type operator mapping from κ to \star . We then uniformly introduce ϕ in the uses of ρ , both in

$$\begin{aligned}
\text{Iter}^{(\kappa)} &: (\kappa \rightarrow \star) \rightarrow (\kappa \rightarrow \star) \rightarrow R^\kappa \rightarrow \star \\
\text{Iter}^{(\kappa)} &= \lambda f g z. \forall l : L, u : \kappa. (\{l \triangleright u\} \lesssim z) \Rightarrow [l] \rightarrow f u \rightarrow g u \\
\text{map}_{\Pi}^{(\kappa)} &: \forall z : R^\kappa, f : \kappa \rightarrow \star, g : \kappa \rightarrow \star. \text{Iter } f g z \rightarrow \Pi(f z) \rightarrow \Pi(g z) \\
\text{map}_{\Pi}^{(\kappa)} &= \Lambda z f g. \lambda i r. \text{syn}_g (\lambda l. i l (\text{sel } r l)) \\
\text{map}_{\Sigma}^{(\kappa)} &: \forall z : R^\kappa, f : \kappa \rightarrow \star, g : \kappa \rightarrow \star. \text{Iter } f g z \rightarrow \Sigma(f z) \rightarrow \Sigma(g z) \\
\text{map}_{\Sigma}^{(\kappa)} &= \Lambda z f g. \lambda i v. \text{ana}_f (\lambda l x. \text{con } l (i l x)) v
\end{aligned}$$

Fig. 4. Transforming records and variants

typing results of *ana* and *syn* and in typing their body. Rules (T-ana₁) and (T-syn₁) are special cases of these rules, and going forward we will write *ana* and *syn* for *ana*_{λ_t, t} and *syn*_{λ_t, t}, respectively.

With the generalized typing rules for *ana* and *syn*, we can now define kind-indexed families type-transforming maps for record and variants, shown in Figure 4. We write $X^{(\kappa)}$ for a family of X 's indexed by kind κ . We would expect languages based on $R\omega$ to also include kind-polymorphism; we have omitted it from our formalization simply to avoid an orthogonal source of complexity. The type $\text{Iter}^{(\kappa)} f g z$ captures iterated functions over row z ; type operator f is used to construct the input type, and g is used to construct the output type. We make the type abstractions in $\text{map}_{\Pi}^{(\kappa)}$ and $\text{map}_{\Sigma}^{(\kappa)}$ explicit, as we will need to refer to the abstracted types in the calls to *syn* and *ana*.

The implementation of $\text{map}_{\Pi}^{(\kappa)}$ is almost identical to the implementation of map'_{Π} . The crucial difference is in providing the operator g to *syn*. This means that the body of *syn* has the type

$$\forall l : L, u : \kappa. \{l \triangleright u\} \lesssim z \Rightarrow [l] \rightarrow g u$$

That is to say: knowing that $l \triangleright u$ appears in z , we must produce a value of type $g u$. The assumption is sufficient to conclude that $l \triangleright f u$ appears in $f z$, and so *sel* $r l$ is a suitable input to the iterated function $i : \text{Iter}^{(\kappa)} f g z$.

The implementation of $\text{map}_{\Sigma}^{(\kappa)}$ is the expected dual of the implementation of $\text{map}_{\Pi}^{(\kappa)}$. We annotate *ana* with the input-side operator f , so its body has the type

$$\forall l : L, u : \kappa. \{l \triangleright u\} \lesssim z \Rightarrow [l] \rightarrow f u \rightarrow \Sigma(g z)$$

Here we are immediately sure that the value x is a suitable input for i ; from $\{l \triangleright u\} \lesssim z$ we have $\{l \triangleright g u\} \lesssim g z$, and so *con* $l (i l x)$ can be of type $\Sigma(g x)$.

Pointwise application. Pottier and Rémy [2005] describe a pointwise-application operator for records, which maps a record of functions and record of arguments to a record of results. We can describe a similar family of operators in $R\omega$, as follows:

$$\begin{aligned}
Xf^{(\kappa)} &: (\kappa \rightarrow \star) \rightarrow (\kappa \rightarrow \star) \rightarrow (\kappa \rightarrow \star) \\
Xf^{(\kappa)} &= \lambda f g a. f a \rightarrow g a \\
\text{rapply}^{(\kappa)} &: \forall f : \kappa \rightarrow \star, g : \kappa \rightarrow \star, z : R^\kappa. \Pi(Xf^{(\kappa)} f g z) \rightarrow \Pi(f z) \rightarrow \Pi(g z) \\
\text{rapply}^{(\kappa)} &= \lambda d r. \text{map}_{\Pi}^{(\kappa)} (\lambda l x. \text{sel } d l x) r
\end{aligned}$$

The type $Xf^{(\kappa)}$ describes the individual transformation functions; as we expect to have a record of these functions, suited to their record of arguments, we do not have to describe them in a label-generic way. The $\text{rapply}^{(\kappa)}$ function then takes a record of such transformers (note that we

$$\begin{array}{l}
\text{Functor} : (\star \rightarrow \star) \rightarrow \star \\
\text{Functor} = \lambda f. \forall t : \star, u : \star. (t \rightarrow u) \rightarrow f t \rightarrow f u \\
\text{fmap}_{\Sigma} : \forall z : R^{\star \rightarrow \star}. \Pi(\text{Functor } z) \rightarrow \text{Functor } (\Sigma z) \quad \left| \quad \text{fmap}_{\Pi} : \forall z : R^{\star \rightarrow \star}. \Pi(\text{Functor } z) \rightarrow \text{Functor } (\Pi z) \right. \\
\text{fmap}_{\Sigma} = \lambda d f w. \text{map}_{\Sigma}^{(\star \rightarrow \star)} (\lambda l x. \text{sel } d \text{ l } f x) w \quad \left| \quad \text{fmap}_{\Pi} = \lambda d f r. \text{map}_{\Pi}^{(\star \rightarrow \star)} (\lambda l x. \text{sel } d \text{ l } f x) r
\end{array}$$

Fig. 5. Lifting functoriality to records and variants

lift $Xf^{(\kappa)} f g$ from an operator on κ to an operator on R^{κ} and a record of arguments, and produces a record of results. Its implementation is a direct application of map_{Π} , in which the body need only look up the appropriately labeled function in the input d .

Our *rappl*y is not quite the same as Pottier and Rémy's: where we rely on type applications $f z$ and $g z$ based on a single row, they define a pointwise lifting of the function constructor to rows $z_1 \rightarrow z_2$. However: their rows are infinite, with a default type for all labels not mentioned in the row; correspondingly, their records are infinite, with a default value for all labels not mentioned in building the record. This means that $z_1 \rightarrow z_2$ can always be well-defined, by using z_1 's default type as the domain for any labels not mentioned in z_1 and z_2 's default type as the codomain for any labels not mentioned in z_2 . With finite rows, we do not have the same luxury. Should we interpret $z_1 \rightarrow z_2$ as undefined if the label sets of z_1 and z_2 are not identical? Or restrict it to the intersection of those label sets? The former would introduce additional partiality in the type of *rappl*y, while the latter would seem to make *rappl*y impossible to define. Without a more compelling application of this additional flexibility, we have limited ourselves to lifting type operators over rows.

Lifting functoriality. A more substantial application of the map functions is in lifting functoriality—as realized in languages like Haskell—to records and variants. The idea is that if we have a row of type constructors, where each constructor in the row has a suitable mapping operator, then we can derive mapping operators for record and variant type constructors built from that row. Our implementation is shown in Figure 5.

We begin by defining the *Functor* type operator. This should be read as capturing the evidence that a type operator is a functor: *Functor List*, for example, is $\forall t u. (t \rightarrow u) \rightarrow \text{List } t \rightarrow \text{List } u$.

We turn to the types of fmap_{Σ} and fmap_{Π} . We abstract over a row z of type constructors. Lifting *Functor* over z gives a row of types, so $\Pi(\text{Functor } z)$ is a record of evidence that each constructor in z is functorial. Now, we want to make a claim about record and variant types built from z . To do so, we generalize Π and Σ to families of type constructors, where for $z : \kappa_1 \rightarrow \kappa_2$ we write Σz for the type constructor $\lambda t. \Sigma(z t)$ and similarly for Π . This generalization is not necessary—we could write the constructors out—but this abbreviation seems intuitive, and makes the types of fmap_{Σ} and fmap_{Π} natural.

Finally, we can implement fmap_{Σ} and fmap_{Π} directly using map_{Σ} and map_{Π} ; in each case, the mapped function simply looks up the appropriate evidence in d , then applies it to lift f over x .

3.4 Comparing Records and Variants

We continue exploring component-wise operations on variants. Our goal is to compare values of two variant types, given that we can compare the values at each of their constructors. Our intended code is shown in Figure 6. We begin by defining the type operator *Eq*, which captures equality comparisons (actually, any binary comparison); given a row z , $\Pi(\text{Eq } z)$ is a record of comparison operators for each type in z . To compare two values v, w of type Σz , we begin by analyzing w . We

$$\begin{array}{l|l}
\text{Eq} : \star \rightarrow \star & \text{eq}_\Sigma : \forall z : R^\star. \Pi(\text{Eq } z) \rightarrow \text{Eq}(\Sigma z) \\
\text{Eq} = \lambda t. t \rightarrow t \rightarrow \text{Bool} & \text{eq}_\Sigma = \lambda d \, v \, w. \text{ana}(\lambda l \, y. (\text{case } l \, (\lambda x. \text{sel } d \, l \, x \, y) \, \nabla \, \text{const False}) \, v) \, w
\end{array}$$

Fig. 6. Comparing values of variant type

$$\begin{array}{l}
\text{eq}_\Pi : \forall z : R^\star. \Pi(\text{Eq } z) \rightarrow \text{Eq}(\Pi z) \\
\text{eq}_\Pi = \lambda d \, r. \text{fold}(\lambda l \, x. (\text{sel } d \, l) (\text{sel } r \, l \, x)) (\&\&) \text{True}
\end{array}$$

Fig. 7. Comparing values of record type

can then fall back on the branching combinator of ROSE: if v is also built with constructor l , we can compare their contents using the l field of d ; otherwise, the two are definitely unequal.

The only difficulty with this implementation is that it does not type. Consider the branch expression in the body of *ana*. As $v : \Sigma z$, we must show that the two branches combine to give z . However, all we know is that $\{l \triangleright u\} \lesssim z$; while logically this implies that there must be a “remainder” of z less $\{l \triangleright u\}$, we do not have access to it.

Our solution is to update the typing rules for *ana* and *syn*, generalizing the type of the body.

$$\begin{array}{c}
(\tau\text{-ana}_3) \frac{\Gamma \vdash \rho : R^K \quad \Gamma \vdash \phi : \kappa \rightarrow \star \quad \Gamma \vdash M : \forall l : L, u : \kappa, y : R^K. \{l \triangleright u\} \odot y \sim \rho \Rightarrow [l] \rightarrow \phi \, u \rightarrow \tau}{\Gamma \vdash \text{ana}_\phi M : \Sigma(\phi \, \rho) \rightarrow \tau} \\
(\tau\text{-syn}_3) \frac{\Gamma \vdash \rho : R^K \quad \Gamma \vdash \phi : \kappa \rightarrow \star \quad \Gamma \vdash M : \forall l : L, u : \kappa, y : R^K. \{l \triangleright u\} \odot y \sim \rho \Rightarrow [l] \rightarrow \phi \, u}{\Gamma \vdash \text{syn}_\phi M : \Pi(\phi \, \rho)}
\end{array}$$

The changed components of the rules are shaded. Instead of providing evidence that $\{l \triangleright u\} \lesssim \rho$, we now decompose ρ into $\{l \triangleright u\}$ and a row type y . The previous iteration of the rule is a special case of this one. With this rule, our intended implementation of eq_Σ is well-typed.

Unfortunately, the solution for variants does not obviously dualize to give a comparison operator for records. Again, assume we have comparators for each field. The operators we have discussed so far would allow us to build a record of Booleans. However, for the records to be equal, we must then determine whether those Booleans are all true, and (without knowing the specific fields) we have no tools to do so.

To capture functions like these, we introduce a folding operation over records:

$$(\tau\text{-fold}) \frac{M_1 : \forall l : L, t : \star, y : R^\star. (\{l \triangleright t\} \odot y \sim \rho) \Rightarrow [l] \rightarrow t \rightarrow v \quad \Gamma \vdash M_2 : v \rightarrow v \rightarrow v \quad \Gamma \vdash M_3 : v \quad \Gamma \vdash N : \Pi \rho}{\Gamma \vdash \text{fold } M_1 \, M_2 \, M_3 \, N : v}$$

The term M_1 is a label-generic mapping from the fields of the input record $N : \Pi \rho$ to the result type v ; M_2 combines values of type v , and M_3 is an identity for M_2 , used for folding the empty record. Given this folding operator, we can define equality comparison for records, as shown in Figure 7.

Introducing this operator immediately raises several questions. For example: in what order are the mapped fields passed to the folding function? Is the identity included once? At all? And so forth. Our conclusion is that the values passed to fold must follow the same rules as the underlying row theory. Following [Morris and McKinna \[2019\]](#), row theories must be associative and have the empty row as their unit; thus, M_2 should be associative and have M_3 as its unit. For a commutative row theory (as we have been assuming), M_2 should be commutative as well. For a non-commutative

theory, on the other hand, fold would pass values to M_2 consistent with the ordering of fields in the row. And so forth. Absent these constraints, the exact behavior of fold ought to be unspecified.

The dual operator for variants would be an unfold, generating a variant by unrolling a starting value. Introducing such an operator would raise all the same problems as we have for fold. As we have found no compelling uses for unfolding variants, we do not consider this operator further.

3.5 Generic Programming for Non-Commutative Rows

ROSE encompasses multiple models of rows. Our discussion so far has assumed *simple* rows, a commutative row theory which seems to be the most natural approach to typing records and variants. However, other theories may be more suited to particular applications. For example, scoped rows [Berthomieu and le Moniès de Sagazan 1995; Leijen 2005], a non-commutative row theory, are particularly well suited to capturing algebraic effects and handlers. A language that includes both extensible data types and algebraic effects, then, might want to include both simple rows (for data types) and scoped rows (for effects). Alternatively, a language could support encoding algebraic effects via extensible data types, such as by using free monads. But then, to capture effects naturally, the language could support extensible data types over both simple and scoped rows!

The challenge in adapting our account to non-commutative row theories is that we no longer have a single idea of containment. The same label ℓ , or indeed the same labeled type $\ell \triangleright \tau$, may appear multiple times in a single row. To support non-commutative row theories, ROSE introduced two containment operators: the “left” version, $\rho_1 \lesssim_L \rho_2$, which holds if there is a ρ' such that $\rho_1 \odot \rho' \sim \rho_2$, and the “right” version, $\rho_1 \lesssim_R \rho_2$, which holds if there is a ρ' such that $\rho' \odot \rho_1 \sim \rho_2$.

Unfortunately, neither of these is a drop-in replacement for the predicates in our generic operators, as individual entries need not be at either the beginning or end of the input row. We can apply a similar idea, by replacing the constraint $\{l \triangleright u\} \odot y \sim \rho$ with $y_1 \odot \{l \triangleright u\} \odot y_2 \sim \rho$. (Note that we cannot define a corresponding “containment” predicate: l , u , and ρ do not uniquely determine y_1 and y_2 .) As we only have a binary row combination predicate, we express this by $y_1 \odot \{l \triangleright u\} \sim z, z \odot y_2 \sim \rho$:

$$\begin{array}{c}
 \frac{\Gamma \vdash \rho : R^K \quad \Gamma \vdash \phi : \kappa \rightarrow \star \quad \Gamma \vdash M : \forall l : L, u : \kappa, y_1, z, y_2 : R^K. (y_1 \odot \{l \triangleright u\} \sim z, z \odot y_2 \sim \rho) \Rightarrow [l] \rightarrow \phi u \rightarrow \tau}{\Gamma \vdash \text{ana}_\phi M : \Sigma(\phi \rho) \rightarrow \tau} \text{(T-ana}_{nc}\text{)} \\
 \\
 \frac{\Gamma \vdash \rho : R^K \quad \Gamma \vdash \phi : \kappa \rightarrow \star \quad \Gamma \vdash M : \forall l : L, u : \kappa, y_1, z, y_2 : R^K. (y_1 \odot \{l \triangleright u\} \sim z, z \odot y_2 \sim \rho) \Rightarrow [l] \rightarrow \phi u}{\Gamma \vdash \text{syn}_\phi M : \Pi(\phi \rho)} \text{(T-syn}_{nc}\text{)} \\
 \\
 \frac{M_1 : \forall l : L, t : \star, y_1, z, y_2 : R^K. (y_1 \odot \{l \triangleright u\} \sim z, z \odot y_2 \sim \rho) \Rightarrow [l] \rightarrow t \rightarrow v \quad \Gamma \vdash M_2 : v \rightarrow v \rightarrow v \quad \Gamma \vdash M_3 : v \quad \Gamma \vdash N : \Pi \rho}{\Gamma \vdash \text{fold } M_1 M_2 M_3 N : v} \text{(T-fold}_{nc}\text{)}
 \end{array}$$

These rules generalize those previously presented: in a commutative theory, if $y_1 \odot \{l \triangleright u\} \odot y_2 \sim \rho$, then there is a y such that $\{l \triangleright u\} \odot y \sim \rho$, given by $y_1 \odot y_2 \sim y$, and conversely.

4 THE $R\omega$ CALCULUS

This section provides a formal description of the syntax and type system of $R\omega$. As in Morris and McKinna [2019], $R\omega$ is parameterized by a *row theory*, giving the intended interpretation of rows. A row theory \mathcal{T} is a triple $\langle \vdash_{\mathcal{T}}, \equiv_{\mathcal{T}}, \Vdash_{\mathcal{T}} \rangle$, where

- $\vdash_{\mathcal{T}}$ is a kinding relation, capturing when rows are well-formed;
- $\equiv_{\mathcal{T}}$ is an equivalence relation, identifying rows; and,
- $\Vdash_{\mathcal{T}}$ is an entailment relation, giving the meaning of the \lesssim and \odot predicates.

Term variables	x	Type variables	α	Labels	ℓ	Directions	$d \in \{L, R\}$
Kinds	$\kappa ::= \star \mid L \mid R^\kappa \mid \kappa \rightarrow \kappa$						
Predicates	$\pi, \psi ::= \rho \lesssim_d \rho \mid \rho \odot \rho \sim \rho$						
Types	$\phi, \tau, v, \rho, \xi ::= \alpha \mid (\rightarrow) \mid \pi \Rightarrow \tau \mid \forall \alpha : \kappa. \tau \mid \lambda \alpha : \kappa. \tau \mid \tau \tau$ $\mid \ell \mid [\xi] \mid \xi \triangleright \tau \mid \{\tau_1, \dots, \tau_n\} \mid \Pi \rho \mid \Sigma \rho$						
Terms	$M, N ::= x \mid \lambda x : \tau. M \mid M N \mid \Lambda \alpha : \kappa. M \mid M[\tau]$ $\mid \ell \mid M \triangleright M \mid M/M \mid \text{prj}_d M \mid M \# M \mid \text{inj}_d M \mid M \nabla M$ $\mid \text{syn}_\phi M \mid \text{ana}_\phi M \mid \text{fold } M M M M$						
Environments	$\Gamma ::= \varepsilon \mid \Gamma, \alpha : \kappa \mid \Gamma, x : \tau \mid \Gamma, \pi$						

Fig. 8. Syntax

We write $R\omega(\mathcal{T})$ to indicate $R\omega$ instantiated with theory \mathcal{T} . Our description of $R\omega$ syntax (§4.1), types (§4.2), and terms (§4.3) are all given generically over an arbitrary row theory \mathcal{T} . We then provide three concrete row theories. The *minimal* row theory (§4.4) captures labeled rows, but makes no commitment to when (non-singleton) rows are well-formed. The examples in the previous section are all well-typed given only the minimal row theory. We then describe the *simple* row theory (§4.5), which captures commutative Rémy-style rows, and the *scoped* row theory (§4.6), which captures non-commutative Leijen-style rows. We develop the expected metatheory in the following section, when we discuss our denotational interpretation of $R\omega$ in Agda.

4.1 Syntax

The syntax of $R\omega(\mathcal{T})$ is given in Figure 8.

Kinds include types \star , labels L , rows R^κ of kind κ , and type constructors $\kappa \rightarrow \kappa$. Not all possible kinds are currently used in $R\omega$. For example: while nothing prevents describing a type of kind R^L (i.e., a row of labels), we have no primitives that operate on such a type, and indeed suspect that such a type would be very difficult to use (§7).

Predicates include containment $\rho \lesssim_d \rho$ and combination $\rho \odot \rho \sim \rho$. To account for non-commutative row theories, we include directed variants of the containment predicate; intuitively, if $\rho_1 \odot \rho_2 \sim \rho_3$, then $\rho_1 \lesssim_L \rho_3$ and $\rho_2 \lesssim_R \rho_3$. Given a commutative row theory, these predicates are equivalent. In a practical language based on $R\omega$, we anticipate that the predicate language would be extended with other forms of predicates, such as type classes [Wadler and Blott 1989], linearity constraints [Gan et al. 2014; Morris 2016], or general equality constraints [Peyton Jones et al. 2006].

We let ϕ, τ, v, ρ and ξ range over types; when possible, we use ϕ where we expect a type constructor, ρ where we expect a row type, and ξ where we expect a label. Standard type constructs include variables α , constants (here only the function arrow), quantifiers, abstractions, and applications. Predicates appear in qualified types $\pi \Rightarrow \tau$. To incorporate labeling, we include labels (ℓ) themselves, singletons $[\xi]$, and labeled types $\xi \triangleright \tau$. Following ROSE, we treat labeled types and row types independently. Finally, we include rows $\{\tau_1, \dots, \tau_n\}$ (including the empty row), records, and variants. Well-formedness of concrete rows is delegated to the row theory \mathcal{T} .

We let M, N range over terms. Standard terms include variables, type and term abstractions, and applications. Introduction and elimination of qualifiers is implicit. To support labeling terms, we include label (singleton) constants ℓ and terms to label ($M \triangleright M$) and unlabel (M/M). As the singleton record and variant types are isomorphic to their underlying single field or constructor type, we do not provide separate syntax to construct singleton records and variants. Finally, we include the (directed) variant and record operators of ROSE, and the label-generic operators new to $R\omega(\mathcal{T})$.

$$\boxed{\vdash \Gamma}$$

$$\begin{array}{c}
\text{(C-EMP)} \frac{}{\vdash \varepsilon} \quad \text{(C-TVAR)} \frac{\vdash \Gamma}{\vdash \Gamma, \alpha : \kappa} \quad \text{(C-VAR)} \frac{\vdash \Gamma \quad \Gamma \vdash \tau : \star}{\vdash \Gamma, x : \tau} \quad \text{(C-PRED)} \frac{\vdash \Gamma \quad \Gamma \vdash \pi}{\vdash \Gamma, \pi} \\
\boxed{\Gamma \vdash \tau : \kappa} \quad \boxed{\Gamma \vdash \pi} \\
\text{(K-VAR)} \frac{\vdash \Gamma \quad \alpha : \kappa \in \Gamma}{\Gamma \vdash \alpha : \kappa} \quad \text{(K-}\rightarrow\text{)} \frac{\vdash \Gamma}{\Gamma \vdash (\rightarrow) : \star \rightarrow \star \rightarrow \star} \quad \text{(K-}\Rightarrow\text{)} \frac{\Gamma \vdash \pi \quad \Gamma, \pi \vdash \tau : \star}{\Gamma \vdash \pi \Rightarrow \tau : \star} \\
\text{(K-}\forall\text{)} \frac{\Gamma, \alpha : \kappa \vdash \tau : \star}{\Gamma \vdash \forall \alpha : \kappa. \tau : \star} \quad \text{(K-}\rightarrow I\text{)} \frac{\Gamma, \alpha : \kappa_1 \vdash \tau : \kappa_2}{\Gamma \vdash \lambda \alpha : \kappa_1. \tau : \kappa_2} \quad \text{(K-}\rightarrow E\text{)} \frac{\Gamma \vdash \tau_1 : \kappa_1 \rightarrow \kappa_2 \quad \Gamma \vdash \tau_2 : \kappa_1}{\Gamma \vdash \tau_1 \tau_2 : \kappa_2} \\
\text{(K-LAB)} \frac{\vdash \Gamma}{\Gamma \vdash \ell : L} \quad \text{(K-SING)} \frac{\Gamma \vdash \xi : L}{\Gamma \vdash [\xi] : \star} \quad \text{(K-LTY)} \frac{\Gamma \vdash \xi : L \quad \Gamma \vdash \tau : \kappa}{\Gamma \vdash \xi \triangleright \tau : \kappa} \quad \text{(K-ROW)} \frac{\Gamma \vdash \tau \{ \overline{\xi \triangleright \tau} \} : R^\kappa}{\Gamma \vdash \{ \overline{\xi \triangleright \tau} \} : R^\kappa} \\
\text{(K-}\Pi\text{)} \frac{\Gamma \vdash \rho : R^\kappa}{\Gamma \vdash \Pi \rho : \kappa} \quad \text{(K-}\Sigma\text{)} \frac{\Gamma \vdash \rho : R^\kappa}{\Gamma \vdash \Sigma \rho : \kappa} \quad \text{(K-LIFT}_1\text{)} \frac{\Gamma \vdash \rho : R^{\kappa_1 \rightarrow \kappa_2} \quad \Gamma \vdash \tau : \kappa_1}{\Gamma \vdash \rho \tau : R^{\kappa_2}} \\
\text{(K-LIFT}_2\text{)} \frac{\Gamma \vdash \phi : \kappa_1 \rightarrow \kappa_2 \quad \Gamma \vdash \rho : R^{\kappa_1}}{\Gamma \vdash \phi \rho : R^{\kappa_2}} \quad \text{(K-}\lesssim_d\text{)} \frac{\Gamma \vdash \rho_i : R^\kappa}{\Gamma \vdash \rho_1 \lesssim_d \rho_2} \quad \text{(K-}\odot\text{)} \frac{\Gamma \vdash \rho_i : R^\kappa}{\Gamma \vdash \rho_1 \odot \rho_2 \sim \rho_3}
\end{array}$$

Fig. 9. Contexts and kinding.

Environments track three kinds of assumptions: kindings of type variables $\alpha : \kappa$, typings of term variables $x : \tau$, and predicates π (as qualified type elimination is implicit, we do not need to name predicate assumptions). We combine these assumptions into a single context Γ simply to avoid a superfluity of (mostly unchanging) metavariables.

4.2 Types and Kinds

Figure 9 gives rules for context formation ($\vdash \Gamma$), kinding ($\Gamma \vdash \tau : \kappa$), and predicate formation ($\Gamma \vdash \pi$), parameterized by row theory \mathcal{T} .

The kinding rules are mostly standard. Rule (K-ROW) delegates well-formedness of rows to the row theory \mathcal{T} . Rules (K- Π) and (K- Σ) capture the formation of record and variant types, lifted to arbitrary kinds κ . In the functor example (Figure 5, §3.3), we had a row of type constructors $z : R^{\star \rightarrow \star}$. Applying (K- Σ), we can conclude that $\Sigma z : \star \rightarrow \star$, and so that $(\Sigma z) t : \star$.

Rules (K-LIFT₁) and (K-LIFT₂) license the lifting that has played a prominent role in our examples. Rule (K-LIFT₂) says that a type constructor $\kappa_1 \rightarrow \kappa_2$, applied to a row of κ_1 s, generates a row of κ_2 s. Consider a type like $z \rightarrow t$, or more pedantically $(\rightarrow) z t$, where $z : R^{\star}$ and $t : \star$. We begin by applying (K-LIFT₂) to apply (\rightarrow) to z , concluding $(\rightarrow) z : R^{\star \rightarrow \star}$. Then, we apply (K-LIFT₁) to apply $(\rightarrow) z$ to t , concluding that $(\rightarrow) z t : R^{\star}$.

One might argue that these are simply syntactic abbreviations, and complicate the reading of types. Instead, we should follow the lead of Featherweight Ur [Chlipala 2010], and use an explicit map operation to lift types over rows. However, in developing our examples, we found that the extra weight introduced by a more explicit approach obscured the meaning of the terms. For example,

$$\begin{array}{c}
\boxed{\tau \equiv \tau} \quad \boxed{\pi \equiv \pi} \\
\\
(\text{E-REFL}) \frac{}{\tau \equiv \tau} \quad (\text{E-SYM}) \frac{\tau_1 \equiv \tau_2}{\tau_2 \equiv \tau_1} \quad (\text{E-TRANS}) \frac{\tau_1 \equiv \tau_2 \quad \tau_2 \equiv \tau_3}{\tau_1 \equiv \tau_3} \quad (\text{E-}\beta) \frac{}{(\lambda \alpha : \kappa. \tau) v \equiv \tau[v/\alpha]} \\
\\
(\text{E-}\xi_{\Rightarrow}) \frac{\pi_1 \equiv \pi_2 \quad \tau_1 \equiv \tau_2}{\pi_1 \Rightarrow \tau_1 \equiv \pi_2 \Rightarrow \tau_2} \quad (\text{E-}\xi_{\forall}) \frac{\tau[Y/\alpha] \equiv v[Y/\beta]}{\forall \alpha : \kappa. \tau \equiv \forall \beta : \kappa. v} (Y \notin \text{fv}(\tau, v)) \quad (\text{E-}\xi_{\text{APP}}) \frac{\tau_i \equiv v_i}{\tau_1 \tau_2 \equiv v_1 v_2} \\
\\
(\text{E-}\xi_{\triangleright}) \frac{\xi_1 \equiv \xi_2 \quad \tau_1 \equiv \tau_2}{\xi_1 \triangleright \tau_1 \equiv \xi_2 \triangleright \tau_2} \quad (\text{E-ROW}) \frac{\{\overline{\xi_i \triangleright \tau_i}\} \equiv_{\mathcal{T}} \{\overline{\xi'_j \triangleright \tau'_j}\}}{\{\xi_i \triangleright \tau_i\} \equiv \{\xi'_j \triangleright \tau'_j\}} \quad (\text{E-}\xi_{[\cdot]}) \frac{\xi_1 \equiv \xi_2}{[\xi_1] \equiv [\xi_2]} \\
\\
(\text{E-LIFT}_1) \frac{}{\{\xi \triangleright \phi\} \tau \equiv \{\xi \triangleright \phi \tau\}} \quad (\text{E-LIFT}_2) \frac{}{\phi \{\xi \triangleright \tau\} \equiv \{\xi \triangleright \phi \tau\}} \\
\\
(\text{E-}\xi_{\Pi\Sigma}) \frac{\rho_1 \equiv \rho_2}{K\rho_1 \equiv K\rho_2} \quad (\text{E-LIFT}_3) \frac{}{(K\rho) \tau \equiv K(\rho \tau)} \quad (\text{E-SING}) \frac{}{K\{\xi \triangleright \tau\} \equiv \xi \triangleright \tau} \quad (K \in \{\Pi, \Sigma\}) \\
\\
(\text{E-}\xi_{\lesssim_d}) \frac{\tau_i \equiv v_i}{\tau_1 \lesssim_d \tau_2 \equiv v_1 \lesssim_d v_2} \quad (\text{E-}\xi_{\odot}) \frac{\tau_i \equiv v_i}{\tau_1 \odot \tau_2 \sim \tau_3 \equiv v_1 \odot v_2 \sim v_3}
\end{array}$$

Fig. 10. Type and predicate equivalence

contrast our types for *reify* and *reflect* (Figure 3, §3.2) with the more explicit

$$\begin{aligned}
\text{reify} &: \forall z : R^{\star}, t : \star. (\Sigma z \rightarrow t) \rightarrow \Pi(\text{map } (\lambda(s : \star). s \rightarrow t) z) \\
\text{reflect} &: \forall z : R^{\star}, t : \star. \Pi(\text{map } (\lambda(s : \star). s \rightarrow t) z) \rightarrow (\Sigma z \rightarrow t)
\end{aligned}$$

Or, similarly, contrast our type for fmap_{Σ} (Figure 5, §3.3) with the more explicit

$$\text{fmap}_{\Sigma} : \forall z : R^{\star \rightarrow \star}. \Pi(\text{map Functor } z) \rightarrow \text{Functor } (\lambda(t : \star). \Sigma(\text{map } (\lambda f : \star \rightarrow \star. f t) z))$$

But in the end, this is a matter of taste; restricting Σ and Π to arguments of kind R^{\star} and making row mapping explicit would not fundamentally restrict the expressiveness of $R\omega$.

The type equivalence rules are shown in Figure 10. The first three lines are standard. The rules (E-LIFT₁) and (E-LIFT₂) realize the promise made in (κ-LIFT₁) and (κ-LIFT₂), lifting single type operators or type arguments to rows. Rule (E-ROW) delegates equivalence of row types to the row theory \mathcal{T} . Rule (E-LIFT₃) gives Π and Σ their intended meaning at higher kinds. Finally, (E-SING) captures the isomorphism between singleton records, singleton variants, and their underlying field (or constructor) type. Again, this latter rule is not integral to $R\omega$; a more explicit version, with separate terms to introduce and eliminate singleton records and variants, would be just as expressive.

4.3 Terms

Figure 11 gives the typing rules for $R\omega$. We have already developed its novelties in the previous section, but will briefly highlight the remaining features of the type system. Lines 1–3 contain a standard treatment of functions, qualified types, and quantified types. Rule (T-SING) is used to introduce label singleton constants, which can then be used to label (T-▷I) or unlabel (T-▷E) terms. Rule (T-≡) can be used (among other things) to move between labeled terms and singleton records or variants. The rules for projection, concatenation, injection, and branching are identical to the

$$\boxed{\Gamma \vdash M : \tau}$$

$$\begin{array}{c}
\text{(T-VAR)} \frac{\vdash \Gamma \quad x : \tau \in \Gamma}{\Gamma \vdash x : \tau} \quad \text{(T-}\rightarrow\text{I)} \frac{\Gamma \vdash \tau_1 : \star \quad \Gamma, x : \tau_1 \vdash M : \tau_2}{\Gamma \vdash \lambda x : \tau_1. M : \tau_1 \rightarrow \tau_2} \quad \text{(T-}\rightarrow\text{E)} \frac{\Gamma \vdash M_1 : \tau_1 \rightarrow \tau_2 \quad \Gamma \vdash M_2 : \tau_1}{\Gamma \vdash M_1 M_2 : \tau_2} \\
\\
\text{(T-}\equiv\text{)} \frac{\Gamma \vdash M : \tau \quad \tau \equiv v}{\Gamma \vdash M : v} \quad \text{(T-}\Rightarrow\text{I)} \frac{\Gamma \vdash \pi \quad \Gamma, \pi \vdash M : \tau}{\Gamma \vdash M : \pi \Rightarrow \tau} \quad \text{(T-}\Rightarrow\text{E)} \frac{\Gamma \vdash M : \pi \Rightarrow \tau \quad \Gamma \Vdash_{\mathcal{T}} \pi}{\Gamma \vdash M : \tau} \\
\\
\text{(T-}\forall\text{I)} \frac{\Gamma, \alpha : \kappa \vdash M : \tau}{\Gamma \vdash \Lambda \alpha : \kappa. M : \forall \alpha : \kappa. \tau} \quad \text{(T-}\forall\text{E)} \frac{\Gamma \vdash M : \forall \alpha : \kappa. \tau \quad \Gamma \vdash v : \kappa}{\Gamma \vdash M[v] : \tau[v/\alpha]} \\
\\
\text{(T-SING)} \frac{\vdash \Gamma}{\Gamma \vdash \ell : [\ell]} \quad \text{(T-}\triangleright\text{I)} \frac{\Gamma \vdash M_1 : [\ell] \quad \Gamma \vdash M_2 : \tau}{\Gamma \vdash M_1 \triangleright M_2 : \ell \triangleright \tau} \quad \text{(T-}\triangleright\text{E)} \frac{\Gamma \vdash M_1 : \ell \triangleright \tau \quad \Gamma \vdash M_2 : [\ell]}{\Gamma \vdash M_1 / M_2 : \tau} \\
\\
\text{(T-II E)} \frac{\Gamma \vdash M : \Pi \rho_1 \quad \Gamma \Vdash_{\mathcal{T}} \rho_2 \lesssim_d \rho_1}{\Gamma \vdash \text{prj}_d M : \Pi \rho_2} \quad \text{(T-III I)} \frac{\Gamma \vdash M_1 : \Pi \rho_1 \quad \Gamma \vdash M_2 : \Pi \rho_2 \quad \Gamma \Vdash_{\mathcal{T}} \rho_1 \odot \rho_2 \sim \rho_3}{\Gamma \vdash M_1 \# M_2 : \Pi \rho_3} \\
\\
\text{(T-}\Sigma\text{I)} \frac{\Gamma \vdash M : \Sigma \rho_1 \quad \Gamma \Vdash_{\mathcal{T}} \rho_1 \lesssim_d \rho_2}{\Gamma \vdash \text{inj}_d M : \Sigma \rho_2} \quad \text{(T-}\Sigma\text{E)} \frac{\Gamma \vdash M_1 : \Sigma \rho_1 \rightarrow \tau \quad \Gamma \vdash M_2 : \Sigma \rho_2 \rightarrow \tau \quad \Gamma \Vdash_{\mathcal{T}} \rho_1 \odot \rho_2 \sim \rho_3}{\Gamma \vdash M_1 \nabla M_2 : \Sigma \rho_3 \rightarrow \tau} \\
\\
\text{(T-ana)} \frac{\Gamma \vdash \rho : R^K \quad \Gamma \vdash \phi : \kappa \rightarrow \star \quad \Gamma \vdash M : \forall l : L, u : \kappa, y_1, z, y_2 : R^K. (y_1 \odot \{l \triangleright u\} \sim z, z \odot y_2 \sim \rho) \Rightarrow [l] \rightarrow \phi u \rightarrow \tau}{\Gamma \vdash \text{ana}_{\phi} M : \Sigma(\phi \rho) \rightarrow \tau} \\
\\
\text{(T-syn)} \frac{\Gamma \vdash \rho : R^K \quad \Gamma \vdash \phi : \kappa \rightarrow \star \quad \Gamma \vdash M : \forall l : L, u : \kappa, y_1, z, y_2 : R^K. (y_1 \odot \{l \triangleright u\} \sim z, z \odot y_2 \sim \rho) \Rightarrow [l] \rightarrow \phi u}{\Gamma \vdash \text{syn}_{\phi} M : \Pi(\phi \rho)} \\
\\
\text{(T-fold)} \frac{M_1 : \forall l : L, t : \star, y_1, z, y_2 : R^K. (y_1 \odot \{l \triangleright u\} \sim z, z \odot y_2 \sim \rho) \Rightarrow [l] \rightarrow t \rightarrow v \quad \Gamma \vdash M_2 : v \rightarrow v \rightarrow v \quad \Gamma \vdash M_3 : v \quad \Gamma \vdash N : \Pi \rho}{\Gamma \vdash \text{fold } M_1 M_2 M_3 N : v}
\end{array}$$

Fig. 11. Typing

corresponding rules for ROSE. Finally, the rules for analyzing variants and synthesizing and folding rows are discussed in the previous section.

4.4 Minimal Rows

Figure 12 gives the minimal row theory \mathcal{M} .

The minimal row theory only includes singleton rows, and so $R\omega(\mathcal{M})$ can express very few practical uses of extensible data types. However, the minimal row theory captures the fundamental properties that all (labeled) row theories share. Our motivating examples (§3) all type in $R\omega(\mathcal{M})$.

The interesting content of the minimal row theory is its entailment relation. Rules (N-REFL) and (N-TRANS) make containment a preorder. Rules (N- \lesssim LIFT₁) and (N- \lesssim LIFT₂) capture that containment is preserved by lifted application. Rules (N- \odot LIFT₁) and (N- \odot LIFT₂) similarly capture that combination is preserved by lifted application. Finally, rules (N- \odot \lesssim L) and (N- \odot \lesssim R) capture the relationship between containment and combination.

$$\begin{array}{c}
\boxed{\Gamma \vdash_m \rho : \kappa} \quad \boxed{\rho \equiv_m \rho} \\
\\
(\text{K-MROW}) \frac{\Gamma \vdash \xi : L \quad \Gamma \vdash \tau : \kappa}{\Gamma \vdash_m \{\xi \triangleright \tau\} : R^\kappa} \quad (\text{E-MROW}) \frac{\xi \equiv \xi' \quad \tau \equiv \tau'}{\{\xi \triangleright \tau\} \equiv_m \{\xi' \triangleright \tau'\}} \\
\boxed{\Gamma \Vdash_m \pi} \\
\\
(\text{N-AX}) \frac{\pi \in \Gamma}{\Gamma \Vdash_m \pi} \quad (\text{N-REFL}) \frac{}{\Gamma \Vdash_m \rho \lesssim_d \rho} \quad (\text{N-TRANS}) \frac{\Gamma \Vdash_m \rho_1 \lesssim_d \rho_2 \quad \Gamma \Vdash_m \rho_2 \lesssim_d \rho_3}{\Gamma \Vdash_m \rho_1 \lesssim_d \rho_3} \\
(\text{N-}\equiv) \frac{\Gamma \Vdash_m \pi_1 \quad \pi_1 \equiv \pi_2}{\Gamma \Vdash_m \pi_2} \quad (\text{N-}\lesssim\text{LIFT}_1) \frac{\Gamma \Vdash_m \rho_1 \lesssim_d \rho_2}{\Gamma \Vdash_m \phi \rho_1 \lesssim_d \phi \rho_2} \quad (\text{N-}\lesssim\text{LIFT}_2) \frac{\Gamma \Vdash_m \rho_1 \lesssim_d \rho_2}{\Gamma \Vdash_m \rho_1 \tau \lesssim_d \rho_2 \tau} \\
(\text{N-}\odot\text{LIFT}_1) \frac{\Gamma \Vdash_m \rho_1 \odot \rho_2 \sim \rho_3}{\Gamma \Vdash_m \rho_1 \tau \odot \rho_2 \tau \sim \rho_3 \tau} \quad (\text{N-}\odot\text{LIFT}_2) \frac{\Gamma \Vdash_m \rho_1 \odot \rho_2 \sim \rho_3}{\Gamma \Vdash_m \phi \rho_1 \odot \phi \rho_2 \sim \phi \rho_3} \\
(\text{N-}\odot\lesssim_L) \frac{\Gamma \Vdash_m \rho_1 \odot \rho_2 \sim \rho_3}{\Gamma \Vdash_m \rho_1 \lesssim_L \rho_3} \quad (\text{N-}\odot\lesssim_R) \frac{\Gamma \Vdash_m \rho_1 \odot \rho_2 \sim \rho_3}{\Gamma \Vdash_m \rho_2 \lesssim_R \rho_3}
\end{array}$$

Fig. 12. Minimal row theory $\mathcal{M} = \langle \vdash_m, \equiv_m, \Vdash_m \rangle$

$$\begin{array}{c}
\boxed{\Gamma \vdash_s \rho : R^\kappa} \quad \boxed{\rho \equiv_s \rho} \\
\\
(\text{K-SROW}) \frac{\Gamma \vdash \xi_i : L \quad \Gamma \vdash \tau_i : \kappa \quad \forall i, j \neq i. \xi_i \# \xi_j}{\Gamma \vdash_s \left\{ \overline{\xi_i \triangleright \tau_i} \right\}^{i \in 1 \dots n} : R^\kappa} \quad (\text{E-SROW}) \frac{\xi_i \equiv \xi'_{p(i)} \quad \tau_i \equiv \tau'_{p(i)}}{\left\{ \overline{\xi_i \triangleright \tau_i} \right\}^{i \in 1 \dots n} \equiv_s \left\{ \overline{\xi'_j \triangleright \tau'_j} \right\}^{j \in 1 \dots n}} \\
\text{where } \xi \# \xi' \text{ iff } \xi = \ell, \xi' = \ell', \ell \neq \ell' \quad \text{where } p \text{ permutes } 1 \dots n \\
\boxed{\Gamma \Vdash_s \pi} \\
\\
(\text{the rules of } \Vdash_m) \quad (\text{N}_s\text{-}\lesssim_d) \frac{\xi_i \equiv \xi'_{p(i)} \quad \tau_i \equiv \tau'_{p(i)}}{\Gamma \Vdash_s \left\{ \overline{\xi_i \triangleright \tau_i} \right\}^{i \in 1 \dots m} \lesssim_d \left\{ \overline{\xi'_j \triangleright \tau'_j} \right\}^{j \in 1 \dots n}} \\
\text{where } p \text{ injects } 1 \dots m \text{ into } 1 \dots n \\
(\text{N}_s\text{-}\odot) \frac{\xi_i \equiv \xi''_{p(i)} \quad \tau_i \equiv \tau''_{p(i)} \quad \xi'_j \equiv \xi''_{r(j)} \quad \tau'_j \equiv \tau''_{r(j)}}{\Gamma \Vdash_s \left\{ \overline{\xi_i \triangleright \tau_i} \right\}^{i \in 1 \dots m} \odot \left\{ \overline{\xi'_j \triangleright \tau'_j} \right\}^{j \in 1 \dots n} \sim \left\{ \overline{\xi''_k \triangleright \tau''_k} \right\}^{k \in 1 \dots m+n}} \\
\text{where } p \text{ injects } 1 \dots m \text{ into } 1 \dots m+n, r \text{ injects } 1 \dots n \text{ into } 1 \dots m+n, \text{ and for all } i, j, p(i) \neq r(j)
\end{array}$$

Fig. 13. The simple row theory $\mathcal{S} = \langle \vdash_s, \equiv_s, \Vdash_s \rangle$; the entailment relation extends \Vdash_m .

4.5 Simple Rows

Figure 13 gives the simple row theory \mathcal{S} .

The simple theory is a commutative theory, in which labels may appear at most once in any row; it captures the most common approach to row types, originally introduced by Rémy [1989]. The challenge to expressing the simple row theory in $R\omega$ arises from first-class labels. As noted by

Leijen [2004], among others, first-class labels can introduce surprising corner cases. Consider a type like $\Pi\{\xi_1 \triangleright \text{Int}, \xi_2 \triangleright \text{Int}\}$, where ξ_1 and ξ_2 are types of kind L. This type only makes sense if ξ_1 and ξ_2 are guaranteed to be different labels. This restriction is captured in (κ -srow): each pair of labels in a row must be different concrete labels. Of course, this condition is satisfied trivially for the empty and singleton rows. Nor does this requirement limit the use of first-class labels, as longer rows may always be expressed as concatenations of singleton rows—indeed, such an elaboration could be done automatically, treating rows as partial type constructors [Ingle et al. 2022; Jones and Diatchki 2008; Jones et al. 2020].

The entailment relation extends that of the minimal row theory with rules for concrete rows. In each case, the essential evidence is a mapping between rows in the predicate; as we will see in the next section, these mappings are exactly the information needed to implement the record and variant operations. There are more generic entailment rules that could be useful in a practical realization of $R\omega(s)$. For example, combination gives a least upper bound for the containment relation:

$$\frac{\Gamma \Vdash_m \rho_1 \lesssim_d \rho \quad \Gamma \Vdash_m \rho_2 \lesssim_d \rho \quad \Gamma \Vdash_m \rho_1 \odot \rho_2 \sim \rho_3}{\Gamma \Vdash_m \rho_3 \lesssim_d \rho}$$

Nevertheless, the rules we give here capture the essential properties of the simple row theory; we regard further extension of the entailment relation as an orthogonal concern.

4.6 Scoped Rows

Figure 14 gives the scoped row theory C .

The scoped row theory is a non-commutative theory, in which the left-most instance of a given label is preferred; it was introduced by Berthomieu and le Monières de Sagazan [1995] and independently by Leijen [2005]. Because labels can be repeated, there is no difficulty in the kinding rule (κ -crow). However, more care must be taken in the entailment relation: we want to allow $\{y \triangleright \text{Int}\} \lesssim_L \{x \triangleright \text{Int}, y \triangleright \text{Int}\}$, as there is no harm in permuting distinct labels, while excluding $\{x \triangleright \text{Bool}\} \lesssim_L \{x \triangleright \text{Int}, x \triangleright \text{Bool}\}$, as this permutes identical labels. This is captured by the side condition on the permutations in each of the entailment rules, which requires that swapped labels be provably distinct.

5 INTERPRETING (STRATIFIED) $R\omega$ IN AGDA

We have two goals in defining semantics for $R\omega$. Primarily, of course, is to demonstrate the soundness of $R\omega$'s type system. Secondly is to show that $R\omega$ need not introduce runtime dependence on or manipulation of labels compared to extensible data types without label-generic operators.

To accomplish both goals, we embedded $R\omega(\mathcal{M})$ typings in the Agda type theory, and then defined a denotational interpretation of those typings in Agda itself, interpreting the $R\omega$ function space as Agda functions, $R\omega$ records and variants as dependent products and sums with finite natural indices, evidence for containment and combination as maps between finite naturals, and so forth. In particular, labels are interpreted as the unit type, and the indexing of products and sums does not depend on the identities of labels in the source derivations.

While our mechanization of the entailment relation is limited to the minimal row theory, our denotations are not correspondingly limited to singleton rows, records, and variants. To the contrary, because our denotations do not depend on labels directly, they are sufficient for all the row theories discussed in this paper. Concretely: while the minimal theory provides no row z that satisfies the constraint $x \triangleright \text{Int} \odot y \triangleright \text{Int} \sim z$, our Agda denotation includes both suitable instantiations for z and the evidence that they satisfy the constraint.

$$\begin{array}{c}
\boxed{\Gamma \vdash_c \rho : R^K} \quad \boxed{\rho \equiv_c \rho} \\
\\
(\text{K-CROW}) \frac{\Gamma \vdash \xi_i : L \quad \Gamma \vdash \tau_i : \kappa}{\Gamma \vdash_c \left\{ \overline{\xi_i \triangleright \tau_i} \right\}^{i \in 1 \dots n} : R^K} \quad (\text{E-CROW}) \frac{\xi_i \equiv \xi'_{p(i)} \quad \tau_i \equiv \tau'_{p(i)}}{\left\{ \overline{\xi_i \triangleright \tau_i} \right\}^{i \in 1 \dots n} \equiv_c \left\{ \overline{\xi'_j \triangleright \tau'_j} \right\}^{j \in 1 \dots n}} \\
\text{where } p \text{ permutes } 1 \dots n, \text{ if } i < j, p(i) > p(j), \text{ then } \xi_i \# \xi_j \\
\boxed{\Gamma \Vdash_c \pi} \\
\\
(\text{the rules of } \Vdash_m) \quad (\text{Nc-}\lesssim_L) \frac{\xi_i \equiv \xi'_j \quad \tau_i \equiv \tau'_j \quad \text{for } i \in 1 \dots m, p(j) = i}{\Gamma \Vdash_c \left\{ \overline{\xi_i \triangleright \tau_i} \right\}^{i \in 1 \dots m} \lesssim_L \left\{ \overline{\xi'_j \triangleright \tau'_j} \right\}^{j \in 1 \dots n}} \\
\text{where } p \text{ permutes } 1 \dots n, \text{ if } i < j \text{ and } p(i) > p(j), \text{ then } \xi'_i \# \xi'_j \\
\\
(\text{Nc-}\lesssim_R) \frac{\xi_i \equiv \xi'_j \quad \tau_i \equiv \tau'_j \quad \text{for } i \in 1 \dots m, p(j) = n - m + i}{\Gamma \Vdash_c \left\{ \overline{\xi_i \triangleright \tau_i} \right\}^{i \in 1 \dots m} \lesssim_R \left\{ \overline{\xi'_j \triangleright \tau'_j} \right\}^{j \in 1 \dots n}} \\
\text{where } p \text{ permutes } 1 \dots n, \text{ if } i < j \text{ and } p(i) > p(j), \text{ then } \xi'_i \# \xi'_j \\
\\
(\text{Nc-}\odot) \frac{\xi_i \equiv \xi''_k \quad \tau_i \equiv \tau''_k \quad \text{for } i \in 1 \dots m, p(k) = i \quad \xi'_j \equiv \xi''_k \quad \tau'_j \equiv \tau''_k \quad \text{for } j \in 1 \dots n, p(k) = m + j}{\Gamma \Vdash_c \left\{ \overline{\xi_i \triangleright \tau_i} \right\}^{i \in 1 \dots m} \odot \left\{ \overline{\xi'_j \triangleright \tau'_j} \right\}^{j \in 1 \dots n} \sim \left\{ \overline{\xi''_k \triangleright \tau''_k} \right\}^{k \in 1 \dots m+n}} \\
\text{where } p \text{ permutes } 1 \dots m + n, \text{ if } i < j \text{ and } p(i) > p(j) \text{ then } \xi''_i \# \xi''_j.
\end{array}$$

Fig. 14. Scoped rows: kinding, entailment, and equivalence

Our claim of type soundness is semantic in nature and relies on the totality of Agda as a type theory: we show that the denotations of well-kinded types are in the denotations of their kinds, that the denotations of well-typed terms are in the denotations of their types, and so forth. Because our denotations are in a typed theory, we do not have a wrong value (as in [Milner \[1978\]](#)); instead, we extend the guarantees provided by Agda's type system to $R\omega$.

This section gives a high-level overview of our Agda development; interested readers are referred to the full development [[Hubers and Morris 2023](#)]. There are two significant threads. First: our specification of $R\omega$ so far is impredicative, while Agda is a predicative type theory. We address this by stratifying $R\omega$, preserving its practical expressiveness while being suitable for embedding in Agda. Second: we need Agda definitions of the $R\omega$ primitives. With these out of the way, the remainder of the development was pleasingly straightforward, and demonstrates soundness of kinding, typing, and equivalence.

5.1 Stratifying $R\omega$

Our first challenge is developing a predicative version of $R\omega$. Following [Dunfield and Krishnaswami \[2013\]](#), we could identify the monotypes of $R\omega$ (those types without quantifiers), and limit quantifier instantiation to monotypes. However, this approach would unacceptably compromise the

$$\begin{array}{c}
\boxed{\Gamma \vdash_S \tau : \kappa} \\
\\
(\kappa_S\text{-VAR}) \frac{\vdash \Gamma \quad \alpha : \kappa \in \Gamma}{\Gamma \vdash_S \alpha : \kappa} \quad (\kappa_S\text{-}\leq) \frac{\Gamma \vdash_S \tau : \star_i \quad i \leq j}{\Gamma \vdash_S \tau : \star_j} \quad (\kappa_S\text{-}(\rightarrow)) \frac{\vdash \Gamma}{\Gamma \vdash_S (\rightarrow) : \star_i \rightarrow \star_i \rightarrow \star_i} \\
\\
(\kappa_S\text{-}\rightarrow I) \frac{\Gamma, \alpha : \kappa_1 \vdash_S \tau : \kappa_2}{\Gamma \vdash_S \lambda \alpha : \kappa_1. \tau : \kappa_1 \rightarrow \kappa_2} \quad (\kappa_S\text{-}\rightarrow E) \frac{\Gamma \vdash_S \tau_1 : \kappa_1 \rightarrow \kappa_2 \quad \Gamma \vdash_S \tau_2 : \kappa_1}{\Gamma \vdash_S \tau_1 \tau_2 : \kappa_2} \quad (\kappa_S\text{-LAB}) \frac{\vdash \Gamma}{\Gamma \vdash_S \ell : L} \\
\\
(\kappa_S\text{-}\Rightarrow) \frac{\Gamma \vdash_S \pi : i \quad \Gamma, \pi \vdash_S \tau : \star_j}{\Gamma \vdash_S \pi \Rightarrow \tau : \star_{(i+1) \sqcup j}} \quad (\kappa_S\text{-}\forall) \frac{\Gamma, \alpha : \kappa^{(i)} \vdash_S \tau : \star_j}{\Gamma \vdash_S \forall \alpha : \kappa^{(i)}. \tau : \star_{(i+1) \sqcup j}} \quad (\kappa_S\text{-SING}) \frac{\Gamma \vdash_S \xi : L}{\Gamma \vdash_S [\xi] : \star_0} \\
\\
(\kappa_S\text{-LTY}) \frac{\Gamma \vdash_S \xi : L \quad \Gamma \vdash_S \tau : \kappa}{\Gamma \vdash_S \xi \triangleright \tau : \kappa} \quad (\kappa_S\text{-ROW}) \frac{\Gamma \vdash_S \xi : L \quad \Gamma \vdash_S \tau : \kappa}{\Gamma \vdash_S \{\xi \triangleright \tau\} : R^\kappa} \quad (\kappa_S\text{-}\Pi) \frac{\Gamma \vdash_S \rho : R^\kappa}{\Gamma \vdash_S \Pi \rho : \kappa} \\
\\
(\kappa_S\text{-}\Sigma) \frac{\Gamma \vdash_S \rho : R^\kappa}{\Gamma \vdash_S \Sigma \rho : \kappa} \quad (\kappa_S\text{-LIFT1}) \frac{\Gamma \vdash_S \rho : R^{\kappa_1 \rightarrow \kappa_2} \quad \Gamma \vdash_S \tau : \kappa_1}{\Gamma \vdash_S \rho \tau : R^{\kappa_2}} \quad (\kappa_S\text{-LIFT2}) \frac{\Gamma \vdash_S \phi : \kappa_1 \rightarrow \kappa_2 \quad \Gamma \vdash_S \rho : R^{\kappa_1}}{\Gamma \vdash_S \phi \rho : R^{\kappa_2}} \\
\\
\boxed{\Gamma \vdash_S \pi : i} \\
\\
\frac{\Gamma \vdash_S \rho_n : R^{\kappa^{(i)}}}{\Gamma \vdash_S \rho_1 \lesssim \rho_2 : i} \quad \frac{\Gamma \vdash_S \rho_n : R^{\kappa^{(i)}}}{\Gamma \vdash_S \rho_1 \odot \rho_2 \sim \rho_3 : i}
\end{array}$$

Fig. 15. Stratified kinding and predicate formation

expressiveness of $R\omega$. The following type captures a dictionary for Haskell's Monad type class:

$\text{Monad} : \star \rightarrow \star$

$\text{Monad} = \lambda m. \Pi \{ \text{return} \triangleright \forall t : \star. t \rightarrow m t, \text{bind} \triangleright \forall t, u : \star. m t \rightarrow (t \rightarrow m u) \rightarrow m u \}$

with selector functions such as:

$\text{return} : \forall m : \star \rightarrow \star, t : \star. \text{Monad } m \rightarrow t \rightarrow m t$

$\text{return} = \lambda d x. \text{sel } d \text{ return } x$

However, to type *return*, we have to instantiate *sel* with the type of the *return* field, $\forall t : \star, t \rightarrow m t$, which is not a monotype.

Instead, we follow the approach of System SF₂ [Leivant 1991], ensuring predicativity by stratifying the $R\omega$ type system. Each type in stratified $R\omega$ is associated with a level. We write $\kappa^{(i)}$ for the kinds of types at level i :

$$\text{Kinds} \quad \kappa^{(i)} = \star_i \mid L \mid R^{\kappa^{(j)}} \mid \kappa^{(j)} \rightarrow \kappa^{(k)} \quad (\text{where } i = j \sqcup k)$$

The base kind \star is now annotated with a level. Labels are types at any level, and the types of rows and type constructors are determined by their component types. We write κ for the union $\bigcup_{i \in \mathbb{N}} \kappa^{(i)}$.

The stratified kinding relation is shown in Figure 15. Overall, stratification has a relatively minor impact. Rule $(\kappa_S\text{-}\leq)$ includes earlier levels in later levels; our mechanization incorporates this rule into the other rules. Rules $(\kappa_S\text{-}\Rightarrow)$ and $(\kappa_S\text{-}\forall)$ ensure that the result type is at least one level higher than the level of the quantified type or predicate. The remaining rules are unchanged. However, note that we do now require that quantification and type abstraction explicitly mention the level of the quantified or argument type. In our mechanization, in turn, we can abstract derivations over

the base level. Figure 15 also includes a stratified version of the predicate formation rule, tracking the level of types that appear in the predicate.

In mechanizing $R\omega$ kinds and types, we have separated the environment Γ into three: a kinding environment Δ , a predicate environment Φ , and a typing environment Γ . We use an intrinsically-kinded representation of types:

$$\begin{aligned} \text{Kind} &: \text{Level} \rightarrow \text{Set} \\ \text{KEnv} &: \text{Level} \rightarrow \text{Set} \\ \text{Ty} &: \forall \{ij : \text{Level}\} \rightarrow \text{KEnv } i \rightarrow \text{Kind } j \rightarrow \text{Set} \end{aligned}$$

We define interpretation functions for kinds, kinding environments, and types:

$$\begin{aligned} \llbracket _ \rrbracket_k &: \forall \{i : \text{Level}\} \rightarrow \text{Kind } i \rightarrow \text{Set} (\text{lsuc } i) \\ \llbracket _ \rrbracket_{ke} &: \forall \{i : \text{Level}\} \rightarrow \text{KEnv } i \rightarrow \text{Set} (\text{lsuc } i) \\ \llbracket _ \rrbracket_t &: \forall \{ij : \text{Level}\} \{\Delta : \text{KEnv } i\} \{\kappa : \text{Kind } j\} \rightarrow \text{Ty } \Delta \kappa \rightarrow \llbracket \Delta \rrbracket_{ke} \rightarrow \llbracket \kappa \rrbracket_k \end{aligned}$$

These definitions are unsurprising. For example: the kind \star_i is interpreted as $\text{Set } i$; kinding environments are interpreted as tuples of types; the type $\forall \alpha : \kappa. \tau$ in kinding environment H is interpreted as a dependent function $(X : \llbracket \kappa \rrbracket_k) \rightarrow \llbracket \tau \rrbracket_t(H, X)$. Label singleton types are all interpreted as \top (the unit type), buttressing our claim that $R\omega$ can be implemented without runtime manipulation or comparison of labels. The interpretation of row types, records, and variants is discussed next.

The interpretation of types gives a constructive proof of the following claim:

THEOREM. *The kind system of $R\omega$ is sound.*

Of course, this is only convincing if the interpretations themselves are non-trivial. Here we rely on the underlying type theory: for example, as we interpret the kind of type constructors $\kappa_1 \rightarrow \kappa_2$ as Agda functions $\llbracket \kappa_1 \rrbracket_k \rightarrow \llbracket \kappa_2 \rrbracket_k$, we can be confident that our interpretations of types of that kind are meaningful. For the full details, please see the Agda development [Hubers and Morris 2023].

5.2 Rows and Indices

We intend our interpretation of records and variants to be both type-safe, and to align with the intuition of those types. That is, a record should be a sequence of its field values, and a variant should be a single tagged value.

We begin with rows themselves. Intuitively, a row is a sequence of types. Our encoding in Agda is almost that direct:

$$\begin{aligned} \text{Row} &: \forall \{i : \text{Level}\} \rightarrow \text{Set } i \rightarrow \text{Set } i \\ \text{Row } A &= \Sigma [n \in \mathbb{N}] (\text{Fin } n \rightarrow A) \end{aligned}$$

That is to say: a row at level i is a dependent pair of its length n and a map from finite indices less than n to types at level i . We can define record and variant constructors (at type \star_i) as dependent functions on rows:

$$\begin{aligned} \Pi &: \forall \{i : \text{Level}\} \rightarrow \text{Row } (\text{Set } i) \rightarrow \text{Set } i \\ \Pi (n, P) &= (i : \text{Fin } n) \rightarrow P \, i \\ \Sigma &: \forall \{i : \text{Level}\} \rightarrow \text{Row } (\text{Set } i) \rightarrow \text{Set } i \\ \Sigma (n, P) &= i : \Sigma [i \in \text{Fin } n] (P \, i) \end{aligned}$$

(We will rely on some overloading to avoid tedious qualified names: Σ followed by a variable binding is the dependent sum constructor; followed by a row, it is the variant constructor.) In each case, we pattern match on the input row, obtaining its length n and a mapping from indices to types P . A variant is the expected tagged value, pairing a tag less than n with a value of the type indexed by the tag. A record is another dependent function: given an index into the record, it returns a value of the type at that index.

We have made one simplification relative to $R\omega$: we implement records and variants only at the base kind, and express the type constructor variants using type functions. This does not reflect a fundamental limitation in our embedding, but simply a choice made for expediency in development.

These definitions emulate our intuition of records and variants. For variants, we are quite close: erasing the types leaves a pair of a tag and a value, just as you might expect to represent a value of a traditional variant type. For records, we are further away: while we emulate accessing fields of a record by offset, the practical construction of records is not emulated by our encoding. Nevertheless, we hope that these encodings demonstrate the potential of a real implementation, even if they do not claim to address all the problems that such an implementation would encounter.

Note that these types have none of the properties we have assumed for the corresponding types in $R\omega$: there are no traces of labels to be found, and order is very much significant in determining the meaning of rows, records, and variants. The mapping between rows in the source language and rows in Agda will be found in the concrete evidence for the row predicates, discussed next.

5.3 Containment and Combination

The next piece of our encoding is the evidence for the containment and combination predicates. The stratification of the entailment relation $\Gamma \Vdash \pi$ is entirely unsurprising. As usual in qualified types, evidence for predicates plays a central role in interpreting the overloaded operators. Our goal is to combine the intuition of a practical realization of $R\omega$ with dependent types to ensure type safety.

Intuitively, containment maps indices in the smaller row to indices in the larger row.

$$\begin{aligned} _ \lesssim _ & : \forall \{i : \text{Level}\} \{A : \text{Set } i\} \rightarrow \text{Row } A \rightarrow \text{Row } A \rightarrow \text{Set } i \\ (n, P) \lesssim (m, Q) & = (i \in \text{Fin } n) \rightarrow \Sigma[j \in \text{Fin } m](P \equiv Q j) \end{aligned}$$

(We have omitted some straightforward but tedious bookkeeping to do with levels.) The evidence for containment is a dependent function over indices in the smaller row, associating each with both an index in the larger row and a proof that the associated types are the same. Implementing record projection and variant injection in terms of this evidence is simple: the former simply precomposes with the evidence function while the latter replaces the existing tag with its image in the evidence function.

Similarly, combination maps indices in the resulting row to indices in one of the two starting rows:

$$\begin{aligned} _ \odot _ & : \forall \{i : \text{Level}\} \{A : \text{Set } i\} \rightarrow \text{Row } A \rightarrow \text{Row } A \rightarrow \text{Row } A \rightarrow \text{Set } i \\ (l, P) \odot (m, Q) \sim (n, R) & = (i \in \text{Fin } n) \rightarrow (\Sigma[j \in \text{Fin } l](P j \equiv R i) \text{ or } (\Sigma[j \in \text{Fin } m](Q j \equiv R i))) \end{aligned}$$

We pair the intuitive mapping on indices with evidence that types agree. As for containment, the implementation of the branching and concatenation operators in terms of this evidence is immediate. Unfortunately, however, this is not sufficient to implement all of the entailment rules of $R\omega$. Our intuition is not just that this be *any* map between the indices, but a *surjective* map: every index in one of the original rows should appear somewhere in the combined row. This intuition justifies the entailment rules $(N \odot \lesssim L)$ and $(N \odot \lesssim R)$, which conclude containment from combination. However, this intuition is not captured in our evidence. We have taken a brute force approach to doing so, by storing the evidence for the two containments in the evidence for combination:

$$\begin{aligned} (l, P) \odot (m, Q) \sim (n, R) & = (i \in \text{Fin } n) \rightarrow (\Sigma[j \in \text{Fin } l](P j \equiv R i) \text{ or } (\Sigma[j \in \text{Fin } m](Q j \equiv R i))) \\ & \times (l, P) \lesssim (n, R) \times (m, Q) \lesssim (n, R) \end{aligned}$$

This definition allows us to realize all of $R\omega$'s entailment rules.

We define an intrinsically-kinded representation of predicates, interpreted as evidence:

$$\begin{aligned} \text{Pred} & : \forall \{ij : \text{Level}\} \rightarrow \text{KEnv } i \rightarrow \text{Kind } j \rightarrow \text{Set} \\ \llbracket _ \rrbracket_p & : \forall \{ij : \text{Level}\} \{\Delta : \text{KEnv } i\} \{\kappa : \text{Kind } j\} \rightarrow \text{Pred } \Delta \kappa \rightarrow \llbracket \Delta \rrbracket_{ke} \rightarrow \text{Set } (\text{lsuc } j) \end{aligned}$$

We define a corresponding intrinsically well-formed definition of predicate environments and entailment:

$$\begin{aligned} \text{PEnv} &: \forall \{i : \text{Level}\} \rightarrow \text{KEnv } i \rightarrow \text{Level} \rightarrow \text{Set} \\ \text{Ent} &: \forall \{i_1 \ i_2 \ i_3 : \text{Level}\} \{\kappa : \text{Kind } i_3\} \rightarrow (\Delta : \text{KEnv } i_1) \rightarrow \text{PEnv } \Delta \ i_2 \rightarrow \text{Pred } \Delta \ \kappa \rightarrow \text{Set} \end{aligned}$$

Finally, we define the meaning of an entailment judgment in terms of the meaning of the predicate it entails:

$$\begin{aligned} \llbracket _ \rrbracket_{pe} &: \forall \{i \ j : \text{Level}\} \{\Delta : \text{KEnv } i\} \rightarrow \text{PEnv } \Delta \ j \rightarrow \llbracket \Delta \rrbracket_{ke} \rightarrow \text{Set } j \\ \llbracket _ \rrbracket_n &: \forall \{i_1 \ i_2 \ i_3 : \text{Level}\} \{\kappa : \text{Kind } i_3\} \{\Delta : \text{KEnv } i_1\} \{\Phi : \text{PEnv } \Delta \ i_2\} \{\pi : \text{Pred } \Delta \ i_3\} \rightarrow \\ &\quad \text{Ent } \Delta \ \Phi \ \pi \rightarrow (H : \llbracket \Delta \rrbracket_{ke}) \rightarrow \llbracket \Phi \rrbracket_{pe} H \rightarrow \llbracket \pi \rrbracket_p H \end{aligned}$$

The latter provides a constructive proof of the following.

THEOREM. *The entailment relation of $R\omega$ is sound.*

5.4 Label-Generic Operations

The label-generic operators *ana*, *syn*, and *fold* work by invoking a suitably parametric function on entries in their source rows. To implement this, we must be able to work backwards from the index used in a variant or record to the corresponding evidence that its type is in the original row. We capture this in Agda as follows.

We begin by introducing an abbreviation for indices over a given row.

$$\begin{aligned} \text{Ix} &: \forall \{i : \text{Level}\} \{A : \text{Set } i\} \rightarrow \text{Row } A \rightarrow \text{Set} \\ \text{Ix } (n, _) &= \text{Fin } n \end{aligned}$$

The *pick* operator selects from a row the singleton row at a particular index, and we can construct evidence that each singleton row is contained within the original row.

$$\begin{aligned} _ \text{pick} _ &: \forall \{i : \text{Level}\} \{A : \text{Set } i\} \rightarrow (\rho : \text{Row } A) \rightarrow \text{Ix } \rho \rightarrow \text{Row } A \\ \text{pickedIn} &: \forall \{i : \text{Level}\} \{A : \text{Set } i\} \{\rho : \text{Row } A\} \{n : \text{Ix } \rho\} \rightarrow \rho \text{ pick } n \leq \rho \end{aligned}$$

Similarly, the *delete* operator returns the row containing everything but the given index. We can also construct evidence that this row is contained within the original.

$$\begin{aligned} _ \text{delete} _ &: \forall \{i : \text{Level}\} \{A : \text{Set } i\} \rightarrow (\rho : \text{Row } A) \rightarrow \text{Ix } \rho \rightarrow \text{Row } A \\ \text{deletedIn} &: \forall \{i : \text{Level}\} \{A : \text{Set } i\} \{\rho : \text{Row } A\} \{n : \text{Ix } \rho\} \rightarrow \rho \text{ delete } n \sim \rho \end{aligned}$$

Finally, for a given index into a row, we can produce the evidence needed to invoke the body of a label-generic operator: that combining the singleton row and that index and the remainder of the row gives the original row.

$$\text{recombine} : \forall \{i : \text{Level}\} \{A : \text{Set } i\} \rightarrow (\rho : \text{Row } A) \rightarrow (n : \text{Ix } \rho) \rightarrow \rho \text{ pick } n \odot \rho \text{ delete } n \sim \rho$$

The implementations of the label-generic operators follow easily.

5.5 Terms and Equivalences

Finally, we come to the representations of terms, and of type equivalences. We use intrinsically kinded representations of type and predicate equivalence:

$$\begin{aligned} _ \equiv_{p_} &: \forall \{\Delta : \text{KEnv}\} \{\kappa : \text{Kind}\} \rightarrow \text{Pred } \Delta \ \kappa \rightarrow \text{Pred } \Delta \ \kappa \rightarrow \text{Set} \\ _ \equiv_{t_} &: \forall \{\Delta : \text{KEnv}\} \{\kappa : \text{Kind}\} \rightarrow \text{Ty } \Delta \ \kappa \rightarrow \text{Ty } \Delta \ \kappa \rightarrow \text{Set} \end{aligned}$$

(We will omit the level bookkeeping for the remainder of this section, as it is entirely routine.)

We have made one important simplification in mechanizing the type equivalence relation. If we restrict type equivalence to kinds \star_i , then we have shown that the interpretation of equivalence derivations is an isomorphism in Agda. That is to say, if we have a derivation that $\tau_1 \equiv \tau_2$, then (for a suitable type environment H) we can show not only functions *to* : $\llbracket \tau_1 \rrbracket H \rightarrow \llbracket \tau_2 \rrbracket H$ and *from* : $\llbracket \tau_2 \rrbracket H \rightarrow \llbracket \tau_1 \rrbracket H$, but also that their compositions are the identity function. In particular, we

validate rule (E-SING), that singleton record and variant types are isomorphic to their underlying field type.

However, this definition of isomorphism is not applicable at higher kinds: type constructors have no elements, so it makes little sense to talk about mappings between them. Moreover, if we remove rule (E-SING), we are able to show stronger results, which generalize to all kinds:

$$\begin{aligned} \llbracket _ \rrbracket_{ep} &: \forall \{\Delta : \text{KEnv}\} \{\kappa : \text{Kind}\} \{\pi_1 \pi_2 : \text{Pred } \Delta \kappa\} \rightarrow \pi_1 \equiv_p \pi_2 \rightarrow (H : \llbracket \Delta \rrbracket_{ke}) \rightarrow \llbracket \pi_1 \rrbracket_p H \equiv \llbracket \pi_2 \rrbracket_p H \\ \llbracket _ \rrbracket_{et} &: \forall \{\Delta : \text{KEnv}\} \{\kappa : \text{Kind}\} \{\tau_1 \tau_2 : \text{Ty } \Delta \kappa\} \rightarrow \tau_1 \equiv_t \tau_2 \rightarrow (H : \llbracket \Delta \rrbracket_{ke}) \rightarrow \llbracket \tau_1 \rrbracket_t H \equiv \llbracket \tau_2 \rrbracket_t H \end{aligned}$$

That is to say: we show that when equivalence is derivable between two predicates or two types (at any kind), their interpretations are propositionally equal in Agda. Given our limitations, these provide constructive proofs of the following claim.

THEOREM. *The type and predicate equivalence relations of $R\omega$ are sound.*

To account for the loss of (E-SING), our term language is extended with terms to construct and deconstruct singleton records and variants. We define intrinsically-typed representations of terms, and their interpretation:

$$\begin{aligned} \text{Env} &: \text{KEnv} \rightarrow \text{Set} \\ \text{Tm} &: (\Delta : \text{KEnv}) \rightarrow \text{PEnv } \Delta \rightarrow \text{Env } \Delta \rightarrow \text{Ty } \Delta \star \rightarrow \text{Set} \\ \llbracket _ \rrbracket_e &: \forall \{\Delta : \text{KEnv}\} \rightarrow \text{Env } \Delta \rightarrow \llbracket \Delta \rrbracket_{ke} \text{Set} \\ \llbracket _ \rrbracket_t &: \forall \{\Delta : \text{KEnv}\} \{\Phi : \text{PEnv}\} \{\Gamma : \text{Env}\} \{\tau : \text{Ty } \Delta \star\} \rightarrow \\ &\quad \text{Tm } \Delta \Phi \Gamma \tau \rightarrow (H : \llbracket \Delta \rrbracket_{ke}) \rightarrow \llbracket \Phi \rrbracket_{pe} H \rightarrow \llbracket \Gamma \rrbracket_e H \rightarrow \llbracket \tau \rrbracket_t H \end{aligned}$$

The latter provides a constructive proof of our final claim.

THEOREM. *The type system of $R\omega$ is sound.*

6 RELATED WORK

There is a significant and growing literature on row types and their applications, and a larger literature on extensible data types in general. We highlight that work that is most relevant to $R\omega$.

Featherweight Ur. The most immediately relevant languages are Featherweight Ur and its practical realization in Ur/Web [Chlipala 2010, 2015a,b]. As in ROSE, Ur supports row and record concatenation with first-class labels, enabled by first-class label inequality proofs. As for $R\omega$, Ur is based on System F ω , and supports mapping type-level operations over rows. Ur has practical evaluation as a framework for database-connected web applications. We view $R\omega$ and Ur as complementary explorations of the design space of extensible data types.

There are several differences in focus between $R\omega$ and Ur. Ur does not include extensible variants. Consequently, the duality of records and variants does not appear in Ur, and examples like our *reify* and *reflect* functions do not apply to Ur. We view extensible variants as an important application of row typing, useful for examples like the expression problem and encoding extensible effects; however, we do not think there is any fundamental reason that Ur's approach to extensible records could not be equally applicable to extensible variants. Ur also does not attempt to generalize over different row theories, but assumes that row disjointness is sufficient to capture extensibility.

The more significant difference between $R\omega$ and Ur is in our approach to generic programming with records. Ur provides a family of folding functions for concrete records types. Instead of our view, in which folds should respect the identities of the underlying row theory, Ur uses the type of its folder to capture the particular order in which the programmer intends to visit fields in the records. We believe that $R\omega$'s *synthesis* operator provides a novel, alternative view of generic programming with records. In particular, we are able to define many of our operations to apply to records regardless of their structure; while we believe that Ur's folder could capture the same operation for any concrete record type, it is less clear that Ur captures them in the general case.

Other row type systems. Row types were originally proposed by Wand [1987] as a mechanism for typing records and variants; he defined rows by extension, one field at a time, and allowed subsequent extensions to overwrite fields already in a record. Rémy [1989] generalized Wand’s approach in several significant ways. He restricts row extension to fields not already present in the row, enforced using kinds. His rows record both present and absent fields, with explicit operations to “forget” entries in rows. Finally, he introduces polymorphism over field presence, allowing his calculus to capture patterns like a single operation for both record extension and record update. Rémy’s approach has been used as the foundation for numerous other row type systems. Blume et al. [2006] extends Rémy’s approach to incorporate first-class blocks over extensible variants. Their implementation relies on the duality between records and variants, translating case blocks into records. However, this duality is not exposed to the programmer; unlike $R\omega$, they rely on having a specific type for case blocks distinct from the normal function type. Other application of Rémy-style row type systems include: Makhholm and Wells’s 2005 system for mixin modules; Lindley and Cheney’s 2012 type system for effect polymorphism; Hillerström and Lindley’s 2016 system for extensible effects and handlers; and, Lindley and Morris’s 2017 account of extensible session types. Gaster and Jones [1996] implement a system with operations similar to Rémy’s, but using qualified types instead of kinds to assure that row extension is well-defined. Lindley et al. [2017] start from Rémy-style rows, but consider several extensions including generic support for renaming entries in rows. Berthomieu and le Monières de Sagazan [1995] and Leijen [2005] independently proposed scoped rows, in which row extension preserves both the original and new fields.

Wand [1991] identified the problems that can arise in typing record concatenation, and proposed an approach based on intersection types. Harper and Pierce [1991] support record concatenation using a new form of quantification, in which quantification is over types disjoint from a given row. Their system cannot express Wand’s problem: while it can require that two rows be disjoint, it cannot require that a single field appear in their concatenation without requiring that it appear in a particular input row.

There have been numerous encodings of row types in other type system features, most notably Haskell’s type classes and type families [Bahr 2014; Kiselyov et al. 2004; Morris 2015; Oliveira et al. 2015; Swierstra 2008]. While impressive, these encodings inevitably rely on encoding rows as particular sequences as types, and so struggle to capture the flexibility that row typing is intended to provide. Extensible data types can also be expressed directly using intersection types and the merge operator [Dunfield 2012; Rioux et al. 2023].

$R\omega$ is differentiated from other row type theories by its focus on label-generic operations. It also inherits the expressiveness of ROSE, and its adaptability to multiple different row theories.

Shallow embeddings in Agda. Our approach to mechanizing the metatheory of $R\omega$ is unusual; far more typical would have been to define an operational semantics of $R\omega$ directly, and then mechanize the expected properties of that operational semantics. We chose to embed the semantics of $R\omega$ directly in Agda for two reasons: we wanted an account that clearly did not rely on labels themselves, and we needed to rely on dependent typing to guarantee that record and variant operations were well-typed. This made it natural to embed our semantics in a dependent type theory, and Agda provides flexible dependently-typed programming and a rich standard library.

Embedding simply-typed λ -calculi in rich type theories is well-traveled ground. There is recent work on shallow or mixed deep and shallow embeddings of rich type theories in rich type theories [Kaposi et al. 2019; McBride 2010]. Our embedding is less impressive than theirs: while we demonstrate that our notions of type equality and predicate entailment are sound, we still require explicit equality and entailment proofs in our derivations.

7 CONCLUSION

We have presented a novel approach to programming with extensible data types, based on label-generic operators for variant destruction and record construction and destruction. We conclude by identifying several directions of future work.

Relating row components. Lindley et al. [2017] proposes a renaming operator for rows, as a tool for simulating scoped rows with simple rows. We might hope to capture such an idea in $R\omega$; indeed, our kind system even includes rows of labels, which seem like a promising start. However, while we could attempt to describe a function that relabeled the fields of a row or constructors of a variant, we have no way to guarantee that the renamed fields are unique! That is, we have nothing that accepts the row $\{a \triangleright b, b \triangleright c\}$ while rejecting the row $\{a \triangleright z, b \triangleright z\}$. More generally, we have no way to impose conditions on the relationship between an entry in a row and the remainder, other than that provided by the row combination predicate.

Realizing $R\omega$. $R\omega$'s goal is to demonstrate the expressiveness of its core features. We identify two challenges in making $R\omega$ more practical. The first is exposing its features in a programmer-friendly surface language, such as a variant of Haskell. Doing so would allow us to use $R\omega$ to capture practical examples from algebraic effects and handlers to extensible compiler passes. While adapting $R\omega$ to a type system without type-level functions would certainly make type reconstruction more likely, it may also introduce limitations in $R\omega$'s expressiveness. The second is an efficient implementation of extensible records and variants, in particular, an account of record construction that does not require copying record values or leave records fragmented.

ACKNOWLEDGMENTS

We thank: James McKinna, for providing initial direction to our mechanization of $R\omega$ as well as general discussion of ROSE; Christa Jenkins, for guidance in developing the mechanization; and, Fabian Ruch for extensive feedback on the final mechanization. This work was supported by the National Science Foundation under Grant No. CCF-2044815.

DATA AVAILABILITY STATEMENT

Our Agda mechanization of $R\omega$ is available online [Hubers and Morris 2023].

REFERENCES

- Patrick Bahr. 2014. Composing and decomposing data types: a closed type families implementation of data types à la carte. In *Proceedings of the 10th ACM SIGPLAN workshop on Generic programming, WGP 2014, Gothenburg, Sweden, August 31, 2014*, José Pedro Magalhães and Tiark Rumpf (Eds.). ACM, 71–82.
- Bernard Berthomieu and Camille le Monières de Sagazan. 1995. A Calculus of Tagged Types, with applications to process languages. In *Workshop on types for program analysis*. Aarhus.
- Matthias Blume, Umut A. Acar, and Wonseok Chae. 2006. Extensible programming with first-class cases. In *Proceedings of the 11th ACM SIGPLAN International Conference on Functional Programming, ICFP 2006, Portland, Oregon, USA, September 16-21, 2006*, John H. Reppy and Julia L. Lawall (Eds.). ACM, 239–250.
- Adam Chlipala. 2010. Ur: statically-typed metaprogramming with type-level record computation. In *Proceedings of the 2010 ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI 2010, Toronto, Ontario, Canada, June 5-10, 2010*, Benjamin G. Zorn and Alexander Aiken (Eds.). ACM, 122–133.
- Adam Chlipala. 2015a. An optimizing compiler for a purely functional web-application language. In *Proceedings of the 20th ACM SIGPLAN International Conference on Functional Programming, ICFP 2015, Vancouver, BC, Canada, September 1-3, 2015*, Kathleen Fisher and John H. Reppy (Eds.). ACM, 10–21. <https://doi.org/10.1145/2784731.2784741>
- Adam Chlipala. 2015b. Ur/Web: A Simple Model for Programming the Web. In *Proceedings of the 42nd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2015, Mumbai, India, January 15-17, 2015*, Sriram K. Rajamani and David Walker (Eds.). ACM, 153–165. <https://doi.org/10.1145/2676726.2677004>

- Joshua Dunfield. 2012. Elaborating intersection and union types. In *ACM SIGPLAN International Conference on Functional Programming, ICFP'12, Copenhagen, Denmark, September 9-15, 2012*, Peter Thiemann and Robby Bruce Findler (Eds.). ACM, 17–28.
- Jana Dunfield and Neelakantan R. Krishnaswami. 2013. Complete and easy bidirectional typechecking for higher-rank polymorphism. In *ACM SIGPLAN International Conference on Functional Programming, ICFP'13, Boston, MA, USA - September 25 - 27, 2013*, Greg Morrisett and Tarmo Uustalu (Eds.). ACM, 429–442. <https://doi.org/10.1145/2500365.2500582>
- Edward Gan, Jesse A. Tov, and Greg Morrisett. 2014. Type Classes for Lightweight Substructural Types. In *Proceedings Third International Workshop on Linearity, LINEARITY 2014, Vienna, Austria, 13th July, 2014 (EPTCS, Vol. 176)*, Sandra Alves and Iliano Cervesato (Eds.). 34–48. <https://doi.org/10.4204/EPTCS.176.4>
- Jacques Garrigue. 1998. Programming with Polymorphic Variants. In *ML Workshop*. ACM.
- Benedict R. Gaster and Mark P. Jones. 1996. *A Polymorphic Type System for Extensible Records and Variants*. Technical Report NOTTCS-TR-96-3. University of Nottingham.
- Robert Harper and Benjamin Pierce. 1991. A Record Calculus Based on Symmetric Concatenation. In *Proceedings of the 18th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL '91)*. ACM, 131–142.
- Daniel Hillerström and Sam Lindley. 2016. Liberating effects with rows and handlers. In *TyDe@ICFP*. ACM, 15–27.
- Alex Hubers and J. Garrett Morris. 2023. Generic Programming with Extensible Data Types; Or, Making Ad Hoc Extensible Data Types Less Ad Hoc—Artifact. <https://doi.org/10.5281/zenodo.8116889>.
- Apoorv Ingle, Alex Hubers, and J. Garrett Morris. 2022. Partial type constructors in practice. In *Haskell '22: 15th ACM SIGPLAN International Haskell Symposium, Ljubljana, Slovenia, September 15 - 16, 2022*, Nadia Polikarpova (Ed.). ACM, 95–107. <https://doi.org/10.1145/3546189.3549923>
- Mark P. Jones. 1994. *Qualified Types: Theory and Practice*. Cambridge University Press.
- Mark P. Jones and Iavor S. Diatchki. 2008. Language and program design for functional dependencies. In *Proceedings of the first ACM SIGPLAN symposium on Haskell (Haskell '08)*. ACM, Victoria, BC, Canada, 87–98.
- Mark P. Jones, J. Garrett Morris, and Richard A. Eisenberg. 2020. Partial type constructors: or, making ad hoc datatypes less ad hoc. *Proc. ACM Program. Lang.* 4, POPL (2020), 40:1–40:28.
- Ambrus Kaposi, András Kovács, and Nicolai Kraus. 2019. Shallow Embedding of Type Theory is Morally Correct. In *Mathematics of Program Construction - 13th International Conference, MPC 2019, Porto, Portugal, October 7-9, 2019, Proceedings (Lecture Notes in Computer Science, Vol. 11825)*, Graham Hutton (Ed.). Springer, 329–365. https://doi.org/10.1007/978-3-030-33636-3_12
- Andrew W. Keep and R. Kent Dybvig. 2013. A nanopass framework for commercial compiler development. In *ACM SIGPLAN International Conference on Functional Programming, ICFP'13, Boston, MA, USA - September 25 - 27, 2013*, Greg Morrisett and Tarmo Uustalu (Eds.). ACM, 343–350. <https://doi.org/10.1145/2500365.2500618>
- Oleg Kiselyov, Ralf Lämmel, and Kean Schupke. 2004. Strongly typed heterogeneous collections. In *Proceedings of the ACM SIGPLAN Workshop on Haskell, Haskell 2004, Snowbird, UT, USA, September 22-22, 2004*, Henrik Nilsson (Ed.). ACM, 96–107.
- Daan Leijen. 2004. *First-class labels for extensible rows* (technical report uu-cs-2004-51 ed.). Technical Report UU-CS-2004-51. Dept. of Computer Science, Universiteit Utrecht. <https://www.microsoft.com/en-us/research/publication/first-class-labels-for-extensible-rows/>
- Daan Leijen. 2005. Extensible records with scoped labels. In *Revised Selected Papers from the Sixth Symposium on Trends in Functional Programming, TFP 2005, Tallinn, Estonia, 23-24 September 2005*. 179–194.
- Daan Leijen. 2014. Koka: Programming with Row Polymorphic Effect Types. In *Proceedings 5th Workshop on Mathematically Structured Functional Programming, MSFP@ETAPS 2014, Grenoble, France, 12 April 2014 (EPTCS, Vol. 153)*, Paul Levy and Neel Krishnaswami (Eds.). 100–126.
- Daan Leijen. 2017. Type directed compilation of row-typed algebraic effects. In *Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages, POPL 2017, Paris, France, January 18-20, 2017*, Giuseppe Castagna and Andrew D. Gordon (Eds.). ACM, 486–499.
- Daniel Leivant. 1991. Finitely Stratified Polymorphism. *Inf. Comput.* 93, 1 (1991), 93–113. [https://doi.org/10.1016/0890-5401\(91\)90053-5](https://doi.org/10.1016/0890-5401(91)90053-5)
- Sam Lindley and James Cheney. 2012. Row-based effect types for database integration. In *Proceedings of TLDI 2012: The Seventh ACM SIGPLAN Workshop on Types in Languages Design and Implementation, Philadelphia, PA, USA, Saturday, January 28, 2012*, Benjamin C. Pierce (Ed.). ACM, 91–102.
- Sam Lindley, Conor McBride, and Craig McLaughlin. 2017. Do be do be do. In *Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages, POPL 2017, Paris, France, January 18-20, 2017*, Giuseppe Castagna and Andrew D. Gordon (Eds.). ACM, 500–514.
- Sam Lindley and J. Garrett Morris. 2017. Lightweight functional session types. In *Behavioural Types: from Theory to Tools*, Simon Gay and António Ravara (Eds.). River Publishers.

- Henning Makholm and J. B. Wells. 2005. Type inference, principal typings, and let-polymorphism for first-class mixin modules. In *Proceedings of the 10th ACM SIGPLAN International Conference on Functional Programming, ICFP 2005, Tallinn, Estonia, September 26-28, 2005*, Olivier Danvy and Benjamin C. Pierce (Eds.). ACM, 156–167.
- Conor McBride. 2010. Outrageous but meaningful coincidences: dependent type-safe syntax and evaluation. In *Proceedings of the ACM SIGPLAN Workshop on Generic Programming, WGP 2010, Baltimore, MD, USA, September 27-29, 2010*, Bruno C. d. S. Oliveira and Marcin Zalewski (Eds.). ACM, 1–12. <https://doi.org/10.1145/1863495.1863497>
- Robin Milner. 1978. A theory of type polymorphism in programming. *J. Comput. System Sci.* (1978), 348–375.
- J. Garrett Morris. 2015. Variations on variants. In *Proceedings of the 8th ACM SIGPLAN Symposium on Haskell (Haskell '15)*, Ben Lippmeier (Ed.). ACM, Vancouver, BC, 71–81.
- J. Garrett Morris. 2016. The best of both worlds: linear functional programming without compromise. In *Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming, ICFP 2016, Nara, Japan, September 18-22, 2016*, Jacques Garrigue, Gabriele Keller, and Eijiro Sumii (Eds.). ACM, 448–461. <https://doi.org/10.1145/2951913.2951925>
- J. Garrett Morris and James McKinna. 2019. Abstracting extensible data types: or, rows by any other name. *Proc. ACM Program. Lang.* 3, POPL (2019), 12:1–12:28. <https://doi.org/10.1145/3290325>
- Bruno Oliveira, Shin-Cheng Mu, and Shu-Hung You. 2015. Modular reifiable matching: a list-of-functors approach to two-level types. In *Haskell*. ACM, 82–93.
- Simon L. Peyton Jones, Dimitrios Vytiniotis, Stephanie Weirich, and Geoffrey Washburn. 2006. Simple unification-based type inference for GADTs. In *Proceedings of the 11th ACM SIGPLAN International Conference on Functional Programming, ICFP 2006, Portland, Oregon, USA, September 16-21, 2006*, John H. Reppy and Julia Lawall (Eds.). ACM, 50–61. <https://doi.org/10.1145/1159803.1159811>
- Gordon D. Plotkin and John Power. 2003. Algebraic Operations and Generic Effects. *Applied Categorical Structures* 11, 1 (2003), 69–94. <https://doi.org/10.1023/A:1023064908962>
- Gordon D. Plotkin and Matija Pretnar. 2009. Handlers of Algebraic Effects. In *Programming Languages and Systems, 18th European Symposium on Programming, ESOP 2009, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2009, York, UK, March 22-29, 2009. Proceedings (Lecture Notes in Computer Science, Vol. 5502)*, Giuseppe Castagna (Ed.). Springer, 80–94. https://doi.org/10.1007/978-3-642-00590-9_7
- François Pottier and Didier Rémy. 2005. The essence of ML type inference. In *Advanced Topics in Types and Programming Languages*, Benjamin C. Pierce (Ed.). The MIT Press.
- Didier Rémy. 1989. Typechecking Records and Variants in a Natural Extension of ML. In *Conference Record of the Sixteenth Annual ACM Symposium on Principles of Programming Languages, Austin, Texas, USA, January 11-13, 1989*. ACM Press, 77–88.
- Didier Rémy. 1992. Typing Record Concatenation for Free. In *POPL '92*. ACM, Albuquerque, New Mexico, 166–176.
- Nick Rioux, Xuejing Huang, Bruno C. d. S. Oliveira, and Steve Zdancewic. 2023. A Bowtie for a Beast: Overloading, Eta Expansion, and Extensible Data Types in F_⋆. *Proc. ACM Program. Lang.* 7, POPL (2023), 515–543. <https://doi.org/10.1145/3571211>
- Dipankar Sarkar, Oscar Waddell, and R. Kent Dybvig. 2004. A nanopass infrastructure for compiler education. In *Proceedings of the Ninth ACM SIGPLAN International Conference on Functional Programming, ICFP 2004, Snow Bird, UT, USA, September 19-21, 2004*, Chris Okasaki and Kathleen Fisher (Eds.). ACM, 201–212. <https://doi.org/10.1145/1016850.1016878>
- Martin Sulzmann. 1997. *Designing Record Systems*. Technical Report YALEU/DCS/RR-1128. Yale University.
- Wouter Swierstra. 2008. Data types à la carte. *J. Funct. Program.* 18, 04 (2008), 423–436.
- Philip Wadler. 1998. The Expression Problem. <http://homepages.inf.ed.ac.uk/wadler/papers/expression/expression.txt>.
- Philip Wadler and Stephen Blott. 1989. How to Make ad-hoc Polymorphism Less ad-hoc. In *Conference Record of the Sixteenth Annual ACM Symposium on Principles of Programming Languages, Austin, Texas, USA, January 11-13, 1989*. ACM Press, 60–76. <https://doi.org/10.1145/75277.75283>
- Mitchell Wand. 1987. Complete Type Inference for Simple Objects. In *Proceedings of the Symposium on Logic in Computer Science (LICS '87), Ithaca, New York, USA, June 22-25, 1987*. IEEE Computer Society, 37–44.
- Mitchell Wand. 1989. Type Inference for Record Concatenation and Multiple Inheritance. In *Proceedings of the Fourth Annual Symposium on Logic in Computer Science (LICS '89), Pacific Grove, California, USA, June 5-8, 1989*. IEEE Computer Society, 92–97.
- Mitchell Wand. 1991. Type Inference for Record Concatenation and Multiple Inheritance. *Inf. Comput.* 93, 1 (1991), 1–15.

Received 2023-03-01; accepted 2023-06-27