# Elaboration with First-Class Implicit Function Types

ANONYMOUS AUTHOR(S)

Implicit functions are dependently typed functions, such that arguments are provided (by default) by inference machinery instead of programmers of the surface language. Implicit functions in Agda are an archetypal example. In the Haskell language as implemented by the Glasgow Haskell Compiler (GHC), polymorphic types are another example. Implicit function types are *first-class* if they are treated as any other type in the surface language. This holds in Agda and partially holds in GHC. Inference and elaboration in the presence of first-class implicit functions poses a challenge; in the context of Haskell and ML-like languages, this has been dubbed "impredicative instantiation" or "impredicative inference". We propose a new solution for elaborating first-class implicit functions, which is applicable to full dependent type theories and compares favorably to prior solutions in terms of power, generality and simplicity. We build atop Norell's bidirectional elaboration algorithm for Agda, and we note that the key issue is incomplete information about insertions of implicit abstractions and applications. We make it possible to track and refine information related to such insertions, by adding a function type to a core Martin-Löf type theory, which supports strict (definitional) currying. This allows us to represent undetermined domain arities of implicit function types, and we can decide at any point during elaboration whether implicit abstractions should be inserted.

Additional Key Words and Phrases: impredicative polymorphism, type theory, elaboration, type inference

## 1 INTRODUCTION

Programmers and users of proof assistants do not like to write out obvious things. Type inference and elaboration serve the purpose of filling in tedious details, translating terse surface-level languages to explicit core languages. Modern systems such as Agda have gotten quite adept at this task. However, in practice, programmers still have to tell the compiler where to try filling in details on its own.

**Implicit function types** are a common mechanism for conveying to the compiler that particular function arguments should be inferred by default. In Agda and Coq, one can use bracketed function domains for this purpose:

$$id : \{A : \mathsf{Set}\} \to A \to A \qquad\qquad \mathsf{Definition}\ id\,\{A : \mathsf{Type}\}(x : A) := x.$$
$$id\,x = x$$

In GHC, one can use forall to define implicit function types[1]

$$id :: \mathsf{forall}\,(a :: *).\, a \to a$$
$$id\,x = x$$

In all of the above cases, if we apply $id$ to an argument, the implicit type argument is provided by elaboration. For example, in Agda, $id\,true$ is elaborated to $id\,\{Bool\}\,true$, and analogously in GHC and Coq. In all three systems, there is also a way to explicitly specify implicit arguments: in Agda we may put arguments in brackets as we have seen, in Coq we can prefix a name with @ to make every implicit argument explicit, as in $@id\,bool\,true$, and in GHC we can enable the language extension TypeApplications [Eisenberg et al. 2016] and write $id\,@Bool\,True$.

Implicit functions are **first-class** if they can be manipulated like any other type. Coq is an example for a system where this is *not* the case. In Coq, the core language does not have an actual

[1] This notation requires language extensions KindSignatures and RankNTypes; one could also write the type $a \to a$ and GHC would silently insert the quantification.

2018. 2475-1421/2018/1-ART1 $15.00
https://doi.org/

implicit function type, instead, implicitness is tied to particular *names*, and while we can write *list* (forall $\{A : \mathsf{Type}\}$, $A \to A$) for a list type with polymorphic elements, the brackets here are simply ignored by Coq. For example, Coq accepts the following definition:

$$\begin{aligned}&\mathsf{Definition}\, poly\ : \mathsf{forall}\,(f\ : \mathsf{forall}\,\{A : \mathsf{Type}\},\ A \to A),\ bool\ *\ nat :=\\ &\quad \mathsf{fun}\, f \Rightarrow (f\, bool\, true,\ f\, nat\, 0)\end{aligned}$$

This is a higher-rank polymorphic function which returns a pair. Note that $f$ is applied to two arguments, because the implicitness in forall $\{A : \mathsf{Type}\}$, $A \to A$ is silently dropped.

In GHC Haskell, forall types are more flexible. We can write the following, with RankNTypes enabled:

$$\begin{aligned}&poly :: (\mathsf{forall}\, a.\, a \to a) \to (Bool,\ Int)\\ &poly\, f = (f\, True,\ f\, 0)\end{aligned}$$

However, polymorphic types are only supported in function domains and as fields of algebraic data constructors. We cannot instantiate an arbitrary type parameter to a forall, as in $[\mathsf{forall}\, a.\, a \to a]$ for a list type with polymorphic elements. While this type is technically allowed by the ImpredicativeTypes language extension, as of GHC 8.8 this extension is deprecated and is not particularly usable in practice.

In Agda, implicit functions are truly a first-class notion, and we may have $List\ (\{A : \mathsf{Set}\} \to A \to A)$ without issue. However, Agda's elaboration still has limitations when it comes to handling implicit functions. Assume that we have $[]$ for the empty list and $-\ ::\ -$ for list extension, and consider the following code:

$$\begin{aligned}&polyList : List\ (\{A : \mathsf{Set}\} \to A \to A)\\ &polyList = (\lambda\, x \to x)\ ::\ []\end{aligned}$$

Agda 2.6.0.1 does not accept this. However, it does accept $polyList = (\lambda\,\{A\}\, x \to x)\ ::\ []$. The issue is the following. Agda first infers a type for $(\lambda\, x \to x)\ ::\ []$, then tries to unify the inferred type with the given $List\ (\{A : \mathsf{Set}\} \to A \to A)$ annotation. However, when Agda elaborates $\lambda\, x \to x$, it does not yet know anything about the element type of the list; it is an undetermined unification variable. Hence, Agda does not know whether it should insert an extra $\lambda\,\{A\}$ or not. If the element type is later found to be an implicit function, then it should, otherwise it should not. To solve this conundrum, Agda simply assumes that any unknown type is *not* an implicit function type, and elects to not insert a lambda. This assumption is often correct, but sometimes — as in the current case — it is not.

There is significant literature on type inference in the presence of first-class polymorphic types, mainly in relation to GHC and ML-like languages; see e.g. [Leijen 2008, 2009; Serrano et al. 2018; Vytiniotis et al. 2006]. The above issue in Agda is a specific instance of the challenges described in the mentioned works. Currently none of the above solutions are supported in production compilers, for reasons of complexity, fragility and interaction with other language features. A recent GHC development [Serrano et al. 2020] offers a solution which is relatively simple, and which is likely to land in an official GHC release. However, none of these works support dependent types, which is a key point in our work.

The solution presented in this paper is to gradually accummulate information about implicit insertions, and to have a setup where insertions can be refined and performed at any time after a particular expression is elaborated. In the current example, our algorithm wraps $\lambda\, x \to x$ in an implicit lambda with unknown arity, whose domain is later refined to be $A : \mathsf{Set}$ when the inferred type is unified with the annotation.

### 1.1 Contributions

- We propose an elaboration algorithm which translates from a small Agda-like surface language to a small Martin-Löf type theory extended with implicit function types, telescopes and *strictly curried function types* with telescope domain. We use these extensions to accumulate information about implicit insertions. Our algorithm is based on Norell's bidirectional elaborator for Agda [Norell 2007, Chapter 3].
- In the System F fragment, the presented elaborator is comparable or superior to previous solutions for impredicative inference. However, it also supports full dependent type theory. Our inference is also global, i.e. it can consider the whole program and not just particular n-ary applications.
- We provide an executable implementation of the elaborator described in this paper.
- Our solution is simple: we implemented elaboration, evaluation and unification in about 670 lines of Haskell, of which 215 lines implement the novel enhancements on the top of Norell's basic elaborator.
- The target theory of elaboration serves as a general platform for elaborating implicit function types: our concrete elaborator is a relatively simple one, and there is room to further develop it.

*1.1.1 Note on terminology.* We prefer to avoid the term "impredicative inference" in order to avoid confusion with impredicativity in type theory. The two notions sometimes coincided historically, but currently they are largely orthogonal. In type theory, impredicativity is a property of a universe, i.e. closure of a universe under arbitrary products. In the type inference literature, impredicativity means the ability to instantiate type variables and metavariables to polymorphic types. In particular, we have that

- Agda has type-theory-predicative universes, but implements type-inference-impredicative elaboration with first-class implicit function types.
- Coq has type-theory-impredicative Prop universe (and optionally also Set), but implements type-inference-predicative elaboration, because of the lack of implicit function types.
- GHC is type-theory-impredicative with RankNTypes enabled and ImpredicativeTypes *disabled*, as we have $(\mathsf{forall}\,(a :: *).\, a \to a) :: *$.

## 2 BIDIRECTIONAL ELABORATION

First, we present a variant of Norell's bidirectional elaborator [Norell 2007, Chapter 3]. Compared to ibid. we make some extensions and simplifications; what we end up with can be viewed as a toy version of the actual Agda elaborator. In this section, we use it to build the backbone of our algorithm and illustrate the key issues. We extend this elaborator in Section 5.

### 2.1 Surface syntax

Figure 1 shows the possible constructs in the surface language. We only have terms, as we have Russell-style universe in the core, and we can conflate types and terms for convenience. The surface syntax does not have semantics or any well-formedness relations attached; its sole purpose is to serve as input to elaboration. Hence, the surface syntax can be also viewed as a small untyped tactic language which is interpreted by the elaborator.

The syntactic constructs are almost the same in the surface language as in the core syntax. The difference is that _ holes only appear in surface syntax. The _ can be used to request a term to be inferred by elaboration, the same way as in Agda. This can be used to give let-definitions without type annotation, as in $\textbf{let}\, x : \_ = \mathsf{U}\, \textbf{in}\, x$.

$$
\begin{array}{rcll}
t, u, v, A, B, C & ::= & x & \text{variable} \\
& | & (x : A) \to B & \text{function type} \\
& | & \{x : A\} \to B & \text{implicit function type} \\
& | & t\, u & \text{application} \\
& | & t\, \{u\} & \text{implicit application} \\
& | & \lambda\, x.\, t & \text{lambda abstraction} \\
& | & \lambda\, \{x\}.\, t & \text{implicit abstraction} \\
& | & \mathsf{U} & \text{universe} \\
& | & \textbf{let}\, x : A = t\, \textbf{in}\, u & \text{let-definition} \\
& | & \_ & \text{hole for inferred term}
\end{array}
$$

Fig. 1. Syntax of the surface language.

### 2.2 Core syntax

Figure 2 lists selected rules of the core language. We avoid a fully formal presentation in this paper. Some notes on what is elided:

- We use nameful notation and implicit weakening, i.e. whenever a term is well-formed in some context, it is assumed to be well-formed (as it is) in extended contexts. We also assume that any specifically mentioned name is fresh, e.g. when we write $\Theta, \alpha : A$, we assume that $\alpha$ is fresh in $\Theta$. Formally, we would use de Bruijn indices for variables, and define variable renaming and parallel substitution by recursion on presyntax, e.g. as in [Schäfer et al. 2015].
- Fixing any $\Theta$ metacontext, parallel substitutions of bound and defined variables form morphisms of a category, where the identity substitution id maps each variable to itself and composition $- \circ -$ is given by pointwise substitution. The action of parallel substitution on terms is functorial, i.e. $t[\sigma][\delta] \equiv t[\sigma \circ \delta]$ and $t[\mathrm{id}] \equiv t$, and typing is stable under substitution.
- Definitional equality is understood to be a congruence and an equivalence relation, which is respected by substitution and typing.
- We elide a number of well-formedness assumptions in rules. For instance, whenever a context appears in a rule, it is assumed to be well-formed. Likewise, whenever we have $\Theta|\Gamma \vdash t : A$, we assume that $\Theta|\Gamma \vdash A : \mathsf{U}$.

From now on, we will only consider well-formed core syntax, and unless otherwise mentioned, presented constructions on the core syntax respect definitional equality.

Alternatively, one could present the syntax as a generalized algebraic theory [Sterling 2019] or a quotient inductive-inductive type [Altenkirch and Kaposi 2016], in which case we would get congruences and quotienting for free, and we would also get a rich model theory for our syntax. However, it seems that there are a number of possible choices for giving an algebraic presentation of metacontexts, and existing works on algebraic presentations of dependent modal contexts (e.g. [Birkedal et al. 2018]) do not precisely cover the current use case. We leave this to future work, along with the investigation of elaboration from an algebraic perspective.

Metacontexts are used to record metavariables which are created during elaboration. In our case, metacontexts are simply a context prefix, and we have variables pointing into it. This corresponds to a particularly simple variant of *crisp type theory* [Licata et al. 2018], where we do not have modal type operators or functions with crisp ("meta") domain. The non-meta typing context additionally

$$\boxed{\Theta \vdash} \quad \textit{metacontext formation}$$
$$\boxed{\Theta|\Gamma \vdash} \quad \textit{context formation}$$
$$\boxed{\Theta|\Gamma \vdash t : A} \quad \textit{typing}$$
$$\boxed{\Theta|\Gamma \vdash t \equiv u : A} \quad \textit{term equality}$$

$$\textsc{metacon/empty}\ \frac{}{\bullet \vdash} \qquad \textsc{metacon/bind}\ \frac{\Theta \vdash \qquad \Theta \vdash A : \mathsf{U}}{\Theta, \alpha : A \vdash} \qquad \textsc{con/empty}\ \frac{\Theta \vdash}{\Theta|\bullet \vdash} \qquad \textsc{con/bind}\ \frac{\Theta|\Gamma \vdash \qquad \Theta|\Gamma \vdash A : \mathsf{U}}{\Theta|\Gamma,\ x : A \vdash}$$

$$\textsc{con/define}\ \frac{\Theta|\Gamma \vdash \qquad \Theta|\Gamma \vdash t : A}{\Theta|\Gamma,\ x : A = t \vdash} \qquad \textsc{metavar}\ \frac{}{\Theta_0,\ \alpha : A,\ \Theta_1|\Gamma \vdash \alpha : A} \qquad \textsc{bound-var}\ \frac{}{\Theta|\Gamma,\ x : A,\ \Delta \vdash x : A}$$

$$\textsc{defined-var}\ \frac{}{\Theta|\Gamma,\ x : A = t,\ \Delta \vdash x : A} \qquad \textsc{universe}\ \frac{}{\Theta|\Gamma \vdash \mathsf{U} : \mathsf{U}} \qquad \textsc{let}\ \frac{\Theta|\Gamma \vdash t : A \qquad \Theta|\Gamma,\ x : A = t \vdash u : B}{\Theta|\Gamma \vdash \mathbf{let}\, x : A = t\, \mathbf{in}\, u : B[x \mapsto t]}$$

$$\textsc{fun}\ \frac{\Theta|\Gamma \vdash A : \mathsf{U} \qquad \Theta|\Gamma,\ x : A \vdash B : \mathsf{U}}{\Theta|\Gamma \vdash (x : A) \to B : \mathsf{U}} \qquad \textsc{implicit-fun}\ \frac{\Theta|\Gamma \vdash A : \mathsf{U} \qquad \Theta|\Gamma,\ x : A \vdash B : \mathsf{U}}{\Theta|\Gamma \vdash \{x : A\} \to B : \mathsf{U}}$$

$$\textsc{app}\ \frac{\Theta|\Gamma \vdash t : (x : A) \to B \qquad \Theta|\Gamma \vdash u : A}{\Theta|\Gamma \vdash t\, u : B[x \mapsto u]} \qquad \textsc{implicit-app}\ \frac{\Theta|\Gamma \vdash t : \{x : A\} \to B \qquad \Theta|\Gamma \vdash u : A}{\Theta|\Gamma \vdash t\, \{u\} : B[x \mapsto u]}$$

$$\textsc{lam}\ \frac{\Theta|\Gamma,\ x : A \vdash t : B}{\Theta|\Gamma \vdash \lambda x. t : (x : A) \to B} \qquad \textsc{implicit-lam}\ \frac{\Theta|\Gamma,\ x : A \vdash t : B}{\Theta|\Gamma \vdash \lambda\, \{x\}. t : \{x : A\} \to B}$$

$$\textsc{fun-}\beta\ \frac{\Theta|\Gamma,\ x : A \vdash t : B \qquad \Theta|\Gamma \vdash u : A}{\Theta|\Gamma \vdash (\lambda x. t)\, u \equiv t[x \mapsto u] : B[x \mapsto u]} \qquad \textsc{implicit-fun-}\beta\ \frac{\Theta|\Gamma,\ x : A \vdash t : B \qquad \Theta|\Gamma \vdash u : A}{\Theta|\Gamma \vdash (\lambda\, \{x\}. t)\, \{u\} \equiv t[x \mapsto u] : B[x \mapsto u]}$$

$$\textsc{fun-}\eta\ \frac{\Theta|\Gamma \vdash t : (x : A) \to B}{\Theta|\Gamma \vdash (\lambda x. t\, x) \equiv t : (x : A) \to B} \qquad \textsc{implicit-fun-}\eta\ \frac{\Theta|\Gamma \vdash t : \{x : A\} \to B}{\Theta|\Gamma \vdash (\lambda\, \{x\}. t\, \{x\}) \equiv t : \{x : A\} \to B}$$

$$\textsc{definition}\ \frac{}{\Theta|\Gamma,\ x : A = t,\ \Delta \vdash x \equiv t : A}$$

Fig. 2. Selected rules of the core language.

supports *defined variables*, which is used in the typing rule for **let**-definitions, and we have that any defined variable is equal to its definition. We support **let** as a convenience feature which is important for defining more involved example programs[2].

The universe U is Russell-style, and we have the type-in-type rule. This causes our core syntax to be non-total, and our elaboration algorithm to be possibly non-terminating. We use type-in-type to simplify presentation, since consistent universe setups are orthogonal to the focus of this work.

Function types only differ from each other in notation: implicit functions have the same rules as "explicit" functions. The primary purpose of implicit function types is to *guide elaboration*: the elaborator will at times compute a type and branch on whether it is an implicit function.

*Notations.* We use Agda-like syntactic sugar both in the surface syntax and in core syntax.

- We use $A \to B$ to refer to non-dependent functions.
- We group domain types together in functions, and omit function arrows, as in $\{A\,B : \mathsf{U}\}(x : A) \to B \to A$.
- We group multiple $\lambda$-s, as in $\lambda\,\{A\}\,\{B\}\,x\,y.\,x$.

*Definition 2.1 (Spines).* We use a spine notation for neutral terms. A spine is a list of terms, noted as $\overline{t}$, where terms may be wrapped in brackets to signal implicit application. For example, if $\overline{u} \equiv (\{A\}, \{B\}, x)$, then $t\,\overline{u}$ denotes $t\,\{A\}\,\{B\}\,x$. In $t\,\overline{u}$, we call $t$ the *head* of the neutral term. In particular, if $t$ is a metavariable, the neutral term is *meta-headed*.

*Example 2.2.* The core syntax is quite expressive as a programming language, thanks to **let**-definitions and the type-in-type rule which allows Church-encodings for a large class of inductive types. For example, the following term computes a list of types by mapping:

$$
\begin{aligned}
&\textbf{let}\,\mathit{List} : \mathsf{U} \to \mathsf{U}\\
&\quad = \lambda\,A.\,(L : \mathsf{U}) \to (A \to L \to L) \to L \to L\ \textbf{in}\\
&\textbf{let}\,\mathit{map} : \{A\,B : \mathsf{U}\} \to (A \to B) \to \mathit{List}\,A \to \mathit{List}\,B\\
&\quad = \lambda\,\{A\}\,\{B\}\,f\,\mathit{as}\,L\,\mathit{cons}\,\mathit{nil}.\,\mathit{as}\,L\,(\lambda\,a.\,\mathit{cons}\,(f\,a))\,\mathit{nil}\ \textbf{in}\\
&\mathit{map}\,\{\mathsf{U}\}\,\{\mathsf{U}\}\,(\lambda\,A.\,A \to A)\,(\lambda\,L\,\mathit{cons}\,\mathit{nil}.\,\mathit{cons}\,\mathsf{U}\,(\mathit{cons}\,\mathsf{U}\,\mathit{nil}))
\end{aligned}
$$

## 2.3 Metasubstitutions

Before we can move on to the description of the elaborator, we need to specify metasubstitutions. These are essentially just parallel substitutions of metacontexts, and their purpose is to keep track of meta-operations (e.g. fresh meta creation or solution of a meta).

- A metasubstitution $\boxed{\theta : \Theta_0 \Rightarrow \Theta_1}$ assigns to each variable in $\Theta_1$ a term in $\Theta_0$ , hence it is represented as a list of terms $(\alpha_1 \mapsto t_1,\ \dots\ \alpha_i \mapsto t_i)$.
- We define the action of a metasubstitution on contexts and terms by recursion; we notate action on contexts as $\Gamma[\theta]$ and action on terms as $t[\theta]$. We remark that there is no abstraction for metavariables in the core syntax, so we do not have to handle variable capture (or index shifting).

[2]In dependent type theories, the **let** rule is not derivable from function application, unlike in simple type theories.

The following are admissible:

$$\frac{\Theta \vdash}{() : \Theta \Rightarrow \bullet}\ \textsc{metasub/empty} \qquad \frac{\theta : \Theta_0 \Rightarrow \Theta_1 \qquad \Theta_0|\bullet \vdash t : A[\theta]}{(\theta,\ \alpha \mapsto t) : \Theta_0 \Rightarrow (\Theta_1,\ \alpha : A)}\ \textsc{metasub/extended} \qquad \frac{\theta : \Theta_0 \Rightarrow \Theta_1 \qquad \Theta_1|\Gamma \vdash}{\Theta_0|\Gamma[\theta] \vdash}\ \textsc{metasub/con-action}$$

$$\frac{\theta : \Theta_0 \Rightarrow \Theta_1 \qquad \Theta_1|\Gamma \vdash t : A}{\Theta_0|\Gamma[\theta] \vdash t[\theta] : A[\theta]}\ \textsc{metasub/tm-action} \qquad \frac{}{\mathsf{id} : \Theta \Rightarrow \Theta}\ \textsc{metasub/identity} \qquad \frac{\theta_0 : \Theta_1 \Rightarrow \Theta_2 \qquad \theta_1 : \Theta_0 \Rightarrow \Theta_1}{\theta_0 \circ \theta_1 : \Theta_0 \Rightarrow \Theta_2}\ \textsc{metasub/composition}$$

$$\frac{}{\mathsf{p} : (\Theta,\ x : A) \Rightarrow \Theta}\ \textsc{metasub/weakening}$$

The identity substitution id maps each variable to itself. Composition is given by pointwise term substitution, id and $- \circ -$ yields a category, and the action of metasubstitution on contexts and terms is functorial. The weakening substitution p (the naming comes from categories-with-families terminology [Dybjer 1995]) can be defined as dropping the last entry from $\mathsf{id} : (\Theta,\ x : A) \Rightarrow (\Theta,\ x : A)$.

### 2.4 Fresh Metavariables

Using *contextual metavariables* is a standard practice in the implementation of dependently typed languages. This means that every "hole" in the surface language is represented as an unknown function which abstracts over all bound variables in the scope of a hole. Unlike [Nanevski et al. 2008] and similarly to [Gundry 2013], we do not have a first-class notion of contextual types, and instead reuse the standard dependent function type to abstract over enclosing contexts.

*Definition 2.3 (Closing type).* For each $\Theta|\Gamma \vdash A : \mathsf{U}$, we define $\Gamma \Rightarrow A$ by recursion on $\Gamma$, such that $\Theta|\bullet \vdash \Gamma \Rightarrow A : \mathsf{U}$.

$$\begin{aligned}
((\Gamma,\ x : A) \Rightarrow B) &:\equiv (\Gamma \Rightarrow ((x : A) \to B)) \\
((\Gamma,\ x : A = t) \Rightarrow B) &:\equiv (\Gamma \Rightarrow B[x \mapsto t]) \\
(\bullet \Rightarrow B) &:\equiv B
\end{aligned}$$

*Definition 2.4 (Contextualization).* For each $\Theta|\bullet \vdash t : \Gamma \Rightarrow A$, we define the spine $\overline{\mathsf{vars}_\Gamma}$ such that that $\Theta|\Gamma \vdash t\,\overline{\mathsf{vars}_\Gamma} : A$, which is $t$ applied to all bound variables in $\Gamma$.

$$\begin{aligned}
(t\,\overline{\mathsf{vars}_{\Gamma,\ x:A}}) &:\equiv (t\,\overline{\mathsf{vars}_\Gamma})\ x \\
(t\,\overline{\mathsf{vars}_{\Gamma,\ x:A=t}}) &:\equiv (t\,\overline{\mathsf{vars}_\Gamma}) \\
(t\,\overline{\mathsf{vars}_\bullet}) &:\equiv t
\end{aligned}$$

*Example 2.5.* If we have $\Gamma \equiv (\bullet,\ A : \mathsf{U},\ B : A \to \mathsf{U})$, then $(\Gamma \Rightarrow \mathsf{U}) \equiv ((A : \mathsf{U})(B : A \to \mathsf{U}) \to \mathsf{U})$ and $t\,\overline{\mathsf{vars}_\Gamma} \equiv t\,A\,B$.

*Definition 2.6 (Fresh meta creation).* We specify $\mathsf{freshMeta}_{\,\Theta|\Gamma}\,A$ as follows:

$$\frac{\Theta|\Gamma \vdash A : \mathsf{U}}{\mathsf{freshMeta}_{\,\Theta|\Gamma}\,A \in \{(\Theta',\ \theta,\ t) \mid (\theta : \Theta' \Rightarrow \Theta)\ \wedge\ (\Theta'|\Gamma[\theta] \vdash t : A[\theta])\}}$$

The definition $\mathsf{freshMeta}_{\,\Theta|\Gamma}\,A :\equiv ((\Theta,\ \alpha : \Gamma \Rightarrow A),\ \mathsf{p},\ \alpha\,\overline{\mathsf{vars}_\Gamma})$, where $\alpha$ is fresh in $\Theta$, satisfies this specification. We extend $\Theta$ with a fresh meta, which has the closing type $\Gamma \Rightarrow A$. The p weakening relates the new metacontext to the old one, by "dropping" the new entry. Lastly, $\alpha\,\overline{\mathsf{vars}_\Gamma}$ is the fresh meta applied to all bound variables.

### 2.5 Unification

We assume that there is a unification procedure, which returns a unifying metasubstitution on success. We only have *homogeneous* unification, i.e. the two terms to be unified must have the same type. The specification is as follows:

$$\frac{\Theta|\Gamma \vdash t : A \qquad \Theta|\Gamma \vdash u : A}{\mathsf{unify}\, t\, u \in \{(\Theta', \theta) \mid (\theta : \Theta' \Rightarrow \Theta) \ \wedge\ (\Theta'|\Gamma[\theta] \vdash t[\theta] \equiv u[\theta] : A[\theta])\} \cup \{\mathsf{fail}\}}$$

For a simple example, assuming $\Theta :\equiv (\bullet, \alpha : \mathsf{U}, \beta : \mathsf{U})$, $\mathsf{unify}\, \alpha\, (\beta \to \beta)$ yields $\Theta' :\equiv (\bullet, \beta : \mathsf{U})$ and the substitution $\theta :\equiv (\alpha \mapsto (\beta \to \beta), \beta \mapsto \beta)$, where $\theta : \Theta' \Rightarrow \Theta$.

Here, we do not require that unification returns most general unifiers, nor do we go into the details of how unification is implemented. Gundry describes unification in detail in [Gundry 2013, Chapter 4] for a similar syntax, with a similar (though more featureful) setup for metacontexts. See also [Abel and Pientka 2011] for a reference on unification. Note that our unification algorithm does not support *constraint postponing*, as we have not talked about constraints at all. In our concrete prototype implementation, unification supports basic pattern unification and metavariable pruning.

### 2.6 Elaboration

Elaboration consists of two (partial) functions, checking and inferring, which are defined by mutual induction on surface syntax. We also have implicit argument insertion as a helper function, which is defined by recursion on core types, and which is used as a post-processing step after inference. First, about the used notations:

- We use a Haskell-like monadic pseudocode notation, where the side effect is failure via fail.
- We use pattern matching notation on core terms; e.g. we may match on whether a type is a function type. This assumes an evaluation/normalization procedure on core terms; but note that we already assume this in unification.
- We abbreviate $\theta_1 \circ \theta_2$ as $\theta_{12}$, $\theta_1 \circ \theta_2 \circ \theta_3$ as $\theta_{123}$ and analogously in other cases. We do this to reduce the visual noise caused by threading composed metasubstitutions everywhere in the elaboration algorithm.

We present the specifications and definitions below, then we describe them in order.

INSERT

$$\frac{\Theta_0|\Gamma \vdash \qquad (\Theta, \theta, t, A) \in \{(\Theta, \theta, t, A) \mid (\theta : \Theta \Rightarrow \Theta_0) \ \wedge\ (\Theta|\Gamma[\theta] \vdash t : A)\} \cup \{\mathsf{fail}\}}{\mathsf{insert}(\Theta, \theta, t, A) \in \{(\Theta, \theta, t, A) \mid (\theta : \Theta \Rightarrow \Theta_0) \ \wedge\ (\Theta|\Gamma[\theta] \vdash t : A)\} \cup \{\mathsf{fail}\}}$$

$$\begin{array}{lll}
\mathsf{insert}\ (\Theta_1, \theta_1, t, \{x : A\} \to B) & :\equiv & \mathbf{let}\ (\Theta_2, \theta_2, u) = \mathsf{freshMeta}_{\Theta_1|\Gamma[\theta_1]}\, A \\
 & & \mathbf{in}\ \mathsf{insert}\ (\Theta_2, \theta_{12}, (t[\theta_2])\, \{u\}, B[\theta_2][x \mapsto u]) \\
\mathsf{insert}\ (\Theta_1, \theta_1, t, A) & :\equiv & \mathbf{return}\ (\Theta_1, \theta_1, t, A) \\
\mathsf{insert\ fail} & :\equiv & \mathsf{fail}
\end{array}$$

$$\textsc{check}\quad \frac{t \text{ is a surface expression} \qquad \Theta|\Gamma \vdash A : \mathsf{U}}{[\![t]\!]\Downarrow_{\Theta|\Gamma} A \in \{(\Theta', \theta, t') \mid (\theta : \Theta' \Rightarrow \Theta) \;\wedge\; (\Theta'|\Gamma[\theta] \vdash t' : A[\theta])\} \cup \{\mathsf{fail}\}}$$

$$\textsc{infer}\quad \frac{t \text{ is a surface expression} \qquad \Theta|\Gamma \vdash}{[\![t]\!]\Uparrow_{\Theta|\Gamma} \;\in \{(\Theta', \theta, t', A) \mid (\theta : \Theta' \Rightarrow \Theta) \;\wedge\; (\Theta'|\Gamma[\theta] \vdash t' : A)\} \cup \{\mathsf{fail}\}}$$

$$
\begin{array}{l}
[\![\lambda\, x.\, t]\!]\Downarrow_{\Theta|\Gamma} ((x : A) \to B) :\equiv \textbf{do} \\
\quad (\Theta', \theta, t') \leftarrow [\![t]\!]\Downarrow_{\Theta|\Gamma,\, x:A} B \\
\quad \textbf{return}\,(\Theta', \theta, \lambda\, x.\, t') \\
[\![\lambda\, \{x\}.\, t]\!]\Downarrow_{\Theta|\Gamma} (\{x : A\} \to B) :\equiv \textbf{do} \\
\quad (\Theta', \theta, t') \leftarrow [\![t]\!]\Downarrow_{\Theta|\Gamma,\, x:A} B \\
\quad \textbf{return}\,(\Theta', \theta, \lambda\, \{x\}.\, t') \\
[\![t]\!]\Downarrow_{\Theta|\Gamma} (\{x : A\} \to B) :\equiv \textbf{do} \\
\quad (\Theta', \theta, t') \leftarrow [\![t]\!]\Downarrow_{\Theta|\Gamma,\, x:A} B \\
\quad (\Theta', \theta, \lambda\, \{x\}.\, t') \\
[\![\textbf{let}\, x : A = t \,\textbf{in}\, u]\!]\Downarrow_{\Theta_0|\Gamma} B :\equiv \textbf{do} \\
\quad (\Theta_1, \theta_1, A') \leftarrow [\![A]\!]\Downarrow_{\Theta_0|\Gamma} \mathsf{U} \\
\quad (\Theta_2, \theta_2, t') \leftarrow [\![t]\!]\Downarrow_{\Theta_1|\Gamma[\theta_1]} A' \\
\quad (\Theta_3, \theta_3, u') \leftarrow [\![u]\!]\Downarrow_{\Theta_2|\Gamma[\theta_{12}]} (B[\theta_{12}]) \\
\quad \textbf{return}\,(\Theta_3, \theta_{123}, \textbf{let}\, x : A'[\theta_{23}] = t'[\theta_3] \,\textbf{in}\, u') \\
[\![\_]\!]\Downarrow_{\Theta|\Gamma} A :\equiv \textbf{do} \\
\quad \textbf{return}\,(\mathsf{freshMeta}_{\Theta|\Gamma}\, A) \\
[\![t]\!]\Downarrow_{\Theta_0|\Gamma} A :\equiv \textbf{do} \\
\quad (\Theta_1, \theta_1, t', B) \leftarrow \mathsf{insert}\,([\![t]\!]\Uparrow_{\Theta_0|\Gamma}) \\
\quad (\Theta_2, \theta_2) \leftarrow \mathsf{unify}\,(A[\theta_1])\, B \\
\quad \textbf{return}\,(\Theta_2, \theta_{12}, t'[\theta_2])
\end{array}
$$

$$
\begin{array}{l}
[\![x]\!]\Uparrow_{\Theta|\Gamma} :\equiv \textbf{do} \\
\quad \textbf{if}\,(\Gamma = (\Gamma_0, x : A, \Gamma_1)) \vee (\Gamma = (\Gamma_0, x : A = t, \Gamma_1)) \\
\quad\quad \textbf{then return}\,(\Theta, \mathsf{id}, x, A) \\
\quad\quad \textbf{else}\, \mathsf{fail} \\
[\![\mathsf{U}]\!]\Uparrow_{\Theta|\Gamma} :\equiv \textbf{do} \\
\quad \textbf{return}\,(\Theta, \mathsf{id}, \mathsf{U}, \mathsf{U}) \\
[\![(x : A) \to B]\!]\Uparrow_{\Theta_0|\Gamma} :\equiv \textbf{do} \\
\quad (\Theta_1, \theta_1, A') \leftarrow [\![A]\!]\Downarrow_{\Theta_1|\Gamma} \mathsf{U} \\
\quad (\Theta_2, \theta_2, B') \leftarrow [\![B]\!]\Downarrow_{\Theta_2|\Gamma[\theta_1],\, x:A'} \mathsf{U} \\
\quad \textbf{return}\,(\Theta_2, \theta_{12}, ((x : A'[\theta_2]) \to B'), \mathsf{U})
\end{array}
$$

$$
\begin{array}{l}
\llbracket \{x : A\} \to B \rrbracket \Uparrow_{\Theta_0 | \Gamma} :\equiv \mathbf{do} \\
\quad (\Theta_1,\, \theta_1,\, A') \leftarrow \llbracket A \rrbracket \Downarrow_{\Theta_1 | \Gamma}\, \mathsf{U} \\
\quad (\Theta_2,\, \theta_2,\, B') \leftarrow \llbracket B \rrbracket \Downarrow_{\Theta_2 | \Gamma[\theta_1],\, x:A'}\, \mathsf{U} \\
\quad \mathbf{return}\ (\Theta_2,\, \theta_{12},\, (\{x : A'[\theta_2]\} \to B'),\, \mathsf{U}) \\
\llbracket \lambda\, x.\, t \rrbracket \Uparrow_{\Theta_0 | \Gamma} :\equiv \mathbf{do} \\
\quad \mathbf{let}\ (\Theta_1,\, \theta_1,\, A) = \mathsf{freshMeta}_{\Theta_0 | \Gamma}\, \mathsf{U} \\
\quad (\Theta_2,\, \theta_2,\, t',\, B) \leftarrow \mathsf{insert}\, (\llbracket t \rrbracket \Uparrow_{\Theta_1 | \Gamma[\theta_1],\, x:A}) \\
\quad \mathbf{return}\ (\Theta_2,\, \theta_{12},\, \lambda\, x.\, t',\, (x : A[\theta_2]) \to B) \\
\llbracket \lambda\, \{x\}.\, t \rrbracket \Uparrow_{\Theta_0 | \Gamma} :\equiv \mathbf{do} \\
\quad \mathbf{let}\ (\Theta_1,\, \theta_1,\, A) = \mathsf{freshMeta}_{\Theta_0 | \Gamma}\, \mathsf{U} \\
\quad (\Theta_2,\, \theta_2,\, t',\, B) \leftarrow \mathsf{insert}\, (\llbracket t \rrbracket \Uparrow_{\Theta_1 | \Gamma[\theta_1],\, x:A}) \\
\quad \mathbf{return}\ (\Theta_2,\, \theta_{12},\, \lambda\, \{x\}.\, t',\, \{x : A[\theta_2]\} \to B) \\
\llbracket t\, u \rrbracket \Uparrow_{\Theta_0 | \Gamma} :\equiv \mathbf{do} \\
\quad (\Theta_1,\, \theta_1,\, t',\, A) \leftarrow \mathsf{insert}\, (\llbracket t \rrbracket \Uparrow_{\Theta_0 | \Gamma}) \\
\quad \mathbf{let}\ (\Theta_2,\, \theta_2, A_0) = \mathsf{freshMeta}_{\Theta_1 | \Gamma[\theta_1]}\, \mathsf{U} \\
\quad \mathbf{let}\ (\Theta_3,\, \theta_3, A_1) = \mathsf{freshMeta}_{\Theta_2 | \Gamma[\theta_{12}],\, x:A_0}\, \mathsf{U} \\
\quad (\Theta_4,\, \theta_4) \leftarrow \mathsf{unify}\, (A[\theta_{23}])\, ((x : A_0[\theta_3]) \to A_1) \\
\quad (\Theta_5,\, \theta_5,\, u') \leftarrow \llbracket u \rrbracket \Downarrow_{\Theta_4 | \Gamma[\theta_{1234}]}\, (A_0[\theta_{34}]) \\
\quad \mathbf{return}\ (\Theta_5,\, \theta_{12345},\, (t'[\theta_{2345}])\, u',\, A_1[\theta_{45}][x \mapsto u']) \\
\llbracket t\, \{u\} \rrbracket \Uparrow_{\Theta_0 | \Gamma} :\equiv \mathbf{do} \\
\quad (\Theta_1,\, \theta_1,\, t',\, A) \leftarrow \llbracket t \rrbracket \Uparrow_{\Theta_0 | \Gamma} \\
\quad \mathbf{let}\ (\Theta_2,\, \theta_2, A_0) = \mathsf{freshMeta}_{\Theta_1 | \Gamma[\theta_1]}\, \mathsf{U} \\
\quad \mathbf{let}\ (\Theta_3,\, \theta_3, A_1) = \mathsf{freshMeta}_{\Theta_2 | \Gamma[\theta_{12}],\, x:A_0}\, \mathsf{U} \\
\quad (\Theta_4,\, \theta_4) \leftarrow \mathsf{unify}\, (A[\theta_{23}])\, (\{x : A_0[\theta_3]\} \to A_1) \\
\quad (\Theta_5,\, \theta_5,\, u') \leftarrow \llbracket u \rrbracket \Downarrow_{\Theta_4 | \Gamma[\theta_{1234}]}\, (A_0[\theta_{34}]) \\
\quad \mathbf{return}\ (\Theta_5,\, \theta_{12345},\, (t'[\theta_{2345}])\, \{u'\},\, A_1[\theta_{45}][x \mapsto u']) \\
\llbracket \mathbf{let}\, x : A = t\, \mathbf{in}\, u \rrbracket \Uparrow_{\Theta_0 | \Gamma} :\equiv \mathbf{do} \\
\quad (\Theta_1,\, \theta_1,\, A') \leftarrow \llbracket A \rrbracket \Downarrow_{\Theta_0 | \Gamma}\, \mathsf{U} \\
\quad (\Theta_2,\, \theta_2,\, t') \leftarrow \llbracket t \rrbracket \Downarrow_{\Theta_1 | \Gamma[\theta_1]}\, A' \\
\quad (\Theta_3,\, \theta_3,\, u',\, B) \leftarrow \llbracket u \rrbracket \Uparrow_{\Theta_2 | \Gamma[\theta_{12}]} \\
\quad \mathbf{return}\ (\Theta_3,\, \theta_{123},\, (\mathbf{let}\, x : A'[\theta_{23}] = t'[\theta_3]\, \mathbf{in}\, u'),\, B) \\
\llbracket \_ \rrbracket \Uparrow_{\Theta | \Gamma} :\equiv \mathbf{do} \\
\quad \mathbf{let}\ (\Theta',\, \theta,\, A) = \mathsf{freshMeta}_{\Theta | \Gamma}\, \mathsf{U} \\
\quad \mathbf{return}\ (\mathsf{freshMeta}_{\Theta' | \Gamma[\theta]}\, A)
\end{array}
$$

#### 2.6.1 *Implicit argument insertion.*

This inserts implicit applications around a core term. For example, if we have a defined name *id* with type $\{A : \mathsf{U}\} \to A \to A$ in a surface program, we usually want to expand $id$ to $id\,\{\alpha\}$, where $\alpha$ is a fresh metavariable. We define insert to take as input the output of $[\![-]\!]\Uparrow$, so that it is more convenient to use as an optional post-processing step after inference.

*2.6.2 Checking.* The first two clauses are checking $\lambda$-s, where the expected type exactly matches the $\lambda$ binders. Hence, we simply check under binders with $[\![t]\!]\Downarrow_{\,\Theta|(\Gamma,\,x:A)} B$, and wrap the resulting term in the appropriate (implicit or explicit) $\lambda$.

The third clause for $[\![t]\!]\Downarrow_{\,\Theta|\Gamma} (\{x : A\} \to B)$ is more interesting. Here, we are checking a surface term which is *not* a $\lambda$ (this follows from our top-down pattern matching notation), with an implicit function expected type. Here, we check $t$ in the extended $\Gamma,\ x : A$ context, and we insert a new implicit $\lambda$ in the elaboration output. This is the only point where implicit $\lambda$-s are introduced by elaboration. Practically, this rule is commonly useful whenever we have a higher-order function where some arguments have implicit function type. For example, in the surface syntax, assume natural numbers, and an induction principle for them:

$$NatInd : (P : Nat \to \mathsf{U}) \to P\,\mathsf{zero} \to (\{n : Nat\} \to P\,n \to P\,(\mathsf{suc}\,n)) \to (n : Nat) \to P\,n$$

Then, define addition using induction:

$$\begin{aligned}&\textbf{let}\,NatPlus : Nat \to Nat \to Nat\\ &\qquad = NatInd\,(\lambda\,n.\,\mathsf{Nat} \to Nat)\,(\lambda\,m.\,m)\,(\lambda\,f\,m.\,\mathsf{suc}\,(f\,m))\ \textbf{in}\ \ldots\end{aligned}$$

When the above is elaborated, the $\lambda\,f\,m.\,\mathsf{suc}\,(f\,m)$ function is checked with the expected type $\{n : Nat\} \to (Nat \to Nat) \to (Nat \to Nat)$, and the elaboration output is $\lambda\,\{n\}\,f\,m.\,\mathsf{suc}\,(f\,m)$. Hence, in this case we do not have to write implicit $\lambda$ in the surface syntax.

For $[\![\textbf{let}\,x : A = t\,\textbf{in}\,u]\!]\Downarrow_{\,\Theta_0|\Gamma} B$, we simply let checking fall through. For $[\![\_]\!]\Downarrow_{\,\Theta|\Gamma} A$, we return a fresh metavariable with the expected type. In any other $[\![t]\!]\Downarrow_{\,\Theta_0|\Gamma} A$ case, we have a *change of direction*: we infer a type for $t$ (with implicit insertions) then unify the expected and inferred types.

*2.6.3 Inferring.* For $[\![x]\!]\Uparrow_{\,\Theta|\Gamma}$, we look up the type of $x$ in $\Gamma$. In the case of U, we always succeed and infer U as type. In the cases for function types, we check that the domains and codomains have type U. For $\lambda$-s, we create a fresh meta for the domain type (since our surface $\lambda$-s are not annotated), and infer types for the bodies.

The $[\![t\,u]\!]\Uparrow_{\,\Theta_0|\Gamma}$ and $[\![t\,\{u\}]\!]\Uparrow_{\,\Theta_0|\Gamma}$ cases are again interesting. Here, we first infer a type for $t$, then refine the type to a function type, and lastly check the argument with the domain type. Note the difference between the explicit and implicit case. In the former case we use insert $([\![t]\!]\Uparrow_{\,\Theta_0|\Gamma})$, which inserts implicit applications. In the latter case we do no insertion. This ensures that implicit applications in the surface syntax behave similarly as in Agda. For example, given $id : \{A : \mathsf{U}\} \to A \to A$ in scope, we elaborate $id\,\mathsf{U}$ as follows:

(1) The expression is an explicit application, so we infer id and insert implicit arguments, returning $id\,\{\alpha\}$, where $\alpha$ is a fresh meta.
(2) We check that U has type $\alpha$. Here we immediately change direction, inferring U as type for U and unifying $\alpha$ with U.
(3) Hence, the resulting output is $id\,\{\mathsf{U}\}\,\mathsf{U}$.

On the other hand, we elaborate $id\,\{\mathsf{U}\}$ as follows:

(1) This is an implicit application, so we infer a type for $id$ without inserting implicit arguments. This yields the inferred type $\{A : \mathsf{U}\} \to A \to A$, and we check that U has type U.
(2) We change direction and infer U as type for U, and successfully unify U with U.

It would be more efficient (and also allow more user-friendly error messages) to not immediately refine the inferred type of $t$ to a function type, but rather match on the inferred type, and only perform refining when the type is meta-headed. We present the unoptimized version here for the sake of brevity.

In the case of **let**, inference again just falls through, and we infer a type for the **let** body. For $[\![\_]\!] \Uparrow_{ins|\Theta} \Gamma$, we create a fresh meta for the type of the hole, and another fresh meta for the hole itself.

*2.6.4 Starting and finishing elaboration.* Given a surface term $t$, we initiate elaboration by computing $[\![t]\!] \Uparrow_{\bullet|\bullet}$. If this succeeds, we get a $(\Theta,\ \theta, t')$ result. Elaboration is successful overall if $\Theta = \bullet$ and $\theta = \mathrm{id}$, i.e. no unsolved metas remain.

*2.6.5 Properties of elaboration.* First, elaboration is *sound* in the sense that it only produces well-formed output.

Theorem 2.7 (Soundness). *The definitions of $[\![-]\!] \Downarrow$ and $[\![-]\!] \Uparrow$ conform to the* check *and* infer *specifications. This follows by induction on surface syntax, while also relying on the properties of substitution, metasubstitution,* insert*,* unify *and* freshMeta*.* □

We remark that this notion of soundness is only a "sanity" or well-typing statement for elaboration. In fact, we could define elaboration as a constantly failing partial function, and it would also conform to the specification. The right way to view this, is that $[\![-]\!] \Downarrow$ and $[\![-]\!] \Uparrow$ together with their specification constitute the semantics of surface syntax. We do not give any other semantics to the surface syntax, nor does it support any other operation.

We do not present any *completeness* result for elaboration in this paper. For an example of what this would entail, in [Dunfield and Krishnaswami 2013] completeness means that whenever there is a way to fill in missing details in the surface syntax, algorithmic typechecking *always* finds it. In ibid. this means figuring out domain types for $\lambda$-s and inserting all implicit applications. However, our elaborator targets a far stronger theory, and it is beyond our reach to succinctly characterize which annotations are inferable in the surface language, and we do not know of any prior work which accomplishes this for a comparably strong elaborator. Our experience in Agda is that it is not tractable in general to figure out which arguments are inferable, by looking at function types, and often we have to run elaboration to see what works.

We can still say something about the behavior of our elaborator. For this, we consider a translation from core terms to surface terms, the evident forgetful translation, which maps core terms to surface counterparts. Now, this is an "evil" construction on core terms, since it does not preserve definitional equality, but we shall only use this evil notion in the following statement.

Theorem 2.8 (Conservativity). *Elaboration is conservative over the surface syntax, in the sense that for any surface term $t$, if checking or inference outputs $t'$, then the forgetful translation of $t'$ differs from $t$ only by*

- *Having all _ holes filled with core expressions.*
- *Having extra implicit $\lambda$-s and implicit applications inserted.*

*This follows by straightforward induction on surface syntax.* □

*Remark.* It is *not* the case that for every term in the core syntax there exists a surface term which elaborates to it. The main reason is that $\lambda$-s in the surface syntax are not annotated, and it is easy to find core $\lambda$ expressions with uninferable domain types. We skip optional surface $\lambda$ domain type annotations for the sake of brevity.

*2.6.6 Omitted features.*

- *Let-generalization.* This is an open research topic in settings with dependent types, and we make no attempt at covering it. See [Eisenberg 2016] for a treatment in a proposed dependent version of Haskell.
- *Polymorphic subtyping.* In some prior works, e.g. in [Dunfield and Krishnaswami 2013; Vytiniotis et al. 2008], there is a subtyping relation arising along instantiations of polymorphic types. In GHC 8 polymorphic subtyping is implemented for function types only. Polymorphic subtyping complicates type inference, and to our knowledge it has not been implemented in any dependently typed setting. We also believe that it is undesirable in dependent settings, because elaboration of subtyping must insert coercions which significantly change the intensional character of programs. For example, if we have covariant list types, then coercing $t : \mathit{List}\,(\{A : \mathsf{U}\} \to A \to A)$ to $t : \mathit{List}\,(\mathit{Bool} \to \mathit{Bool})$ requires mapping over $t$ and inserting implicit applications to $\mathit{Bool}$ for each list element. In System F, all such coercions are erasible, since types are computationally irrelevant, but in our core syntax we have implicit functions with arbitrary (relevant) domains. In GHC, subtyping coercions for functions change operational semantics; this is a reason for abandoning subtyping in recent developments of impredicative inference for GHC [Serrano et al. 2020].

## 3 ISSUES WITH FIRST-CLASS IMPLICIT FUNCTIONS

We revisit now the *polyList* example from Section 1. We assume the following:

$$\mathit{List} : \mathsf{U} \to \mathsf{U} \qquad \mathit{nil} : \{A : \mathsf{U}\} \to \mathit{List}\,A \qquad \mathit{cons} : \{A : \mathsf{U}\} \to A \to \mathit{List}\,A \to \mathit{List}\,A$$

In the following, we present a trace of checking $\mathit{cons}\,(\lambda\,x.\,x)\,\mathit{nil}$ at type $\mathit{List}\,(\{A : \mathsf{U}\} \to A \to A)$. We omit context and metacontext parameters everywhere, and notate recursive calls by indentation. We also omit some checking, inference, implicit insertion and unification calls which are not essential for illustration.

$$
\begin{array}{l}
[\![\mathit{cons}\,(\lambda\,x.\,x)\,\mathit{nil}]\!] \Downarrow\ (\mathit{List}\,(\{A : \mathsf{U}\} \to A \to A)) \\
\quad [\![\mathit{cons}\,(\lambda\,x.\,x)\,\mathit{nil}]\!] \Uparrow \\
\quad\quad [\![\mathit{cons}\,(\lambda\,x.\,x)]\!] \Uparrow \\
\quad\quad\quad [\![\mathit{cons}]\!] \Uparrow \\
\quad\quad\quad = \mathit{cons}\,\{\alpha_0\} \ :\ \alpha_0 \to \mathit{List}\,\alpha_0 \to \mathit{List}\,\alpha_0 \\
\quad\quad\quad [\![\lambda\,x.\,x]\!] \Downarrow\ \alpha_0 \\
\quad\quad\quad = \lambda\,x.\,x \\
\quad\quad = \mathit{cons}\,\{\alpha_1 \to \alpha_1\}\,(\lambda\,x.\,x) : \mathit{List}\,(\alpha_1 \to \alpha_1) \to \mathit{List}\,(\alpha_1 \to \alpha_1) \\
\quad\quad [\![\mathit{nil}]\!] \Downarrow\ (\mathit{List}\,(\alpha_1 \to \alpha_1)) \\
\quad\quad = \mathit{nil}\,\{\alpha_1 \to \alpha_1\} \\
\quad = \mathit{cons}\,\{\alpha_1 \to \alpha_1\}\,(\lambda\,x.\,x)\,(\mathit{nil}\,\{\alpha_1 \to \alpha_1\}) : \mathit{List}\,(\alpha_1 \to \alpha_1) \\
\quad \mathsf{unify}\,(\mathit{List}\,(\{A : \mathsf{U}\} \to A \to A))\,(\mathit{List}\,(\alpha_1 \to \alpha_1)) \\
\quad\quad \mathsf{unify}\,(\{A : \mathsf{U}\} \to A \to A)\,(\alpha_1 \to \alpha_1) \\
\quad\quad = \mathsf{fail}
\end{array}
$$

Above, we first infer $\mathit{cons}\,(\lambda\,x.\,x)\,\mathit{nil}$, which inserts implicit applications to fresh metas in $\mathit{cons}$ and $\mathit{nil}$, and returns $\mathit{cons}\,\{\alpha_1 \to \alpha_1\}\,(\lambda\,x.\,x)\,(\mathit{nil}\,\{\alpha_1 \to \alpha_1\}) : \mathit{List}\,(\alpha_1 \to \alpha_1)$. Here, the $\alpha_0$ meta is refined to $\alpha_1 \to \alpha_1$ when we check $\lambda\,x.\,x$. In the end, we need to unify the expected and inferred types, which fails, since we have an implicit function type on one side and an explicit function on the other side.

Why does this fail? The culprit is line 5, where we call $[\![\lambda\, x.\, x]\!] \Downarrow\ \alpha_0$. At this point, the checking type is not an implicit function type (it is a meta), so we do not insert an implicit $\lambda$. At the heart of the issue is that elaboration makes insertion choices based on core types.

(1) $[\![t]\!] \Downarrow\ _{\Theta|\Gamma}\ A$ can insert a $\lambda$ only if $A$ is an implicit function type.
(2) insert $(\Theta,\ \theta,\ t,\ A)$ inserts an application only if $A$ is an implicit function type.

In both of these cases, if $A$ is of the form $\alpha\,\overline{u}$ (i.e. meta-headed), then it is possible that $\alpha$ is later refined to an implicit function, but at that point we have already missed our shot at implicit insertion.

There is a potential naive solution: just *postpone* checking a term until the shape of the checking type is known for sure. This was included as part of a proposed solution for smarter $\lambda$-insertions in [Johansson and Lloyd 2015]. This means that checking with a meta-headed type returns a "guarded constant" [Norell 2007, Chapter 3], an opaque stand-in which only computes to an actual core term when the checking type becomes known. In practice, this solution has a painful drawback: *we get no information at all from checked terms before the guard is unblocked.*

For an example for unexpected behavior with this solution, let us assume $\mathit{Bool} : \mathsf{U}$ and $\mathit{true} : \mathit{Bool}$, and try to infer type for the surface term $\mathbf{let}\, x : \_ = \mathit{true}\ \mathbf{in}\ x$. We first insert a fresh meta $\alpha$ for the hole, and then check $\mathit{true}$ with $\alpha$. We postpone this checking, returning a guarded constant, and then infer a type for $x$, which is $\alpha$. Hence, this small example yields an unsolved meta and a guarded constant in the output.

Now, this particular example can be repaired by special-casing the elaboration of a **let**-definition without an explicit type annotation. However, our experience from playing with an implementation of this solution, is that we are missing too much information by postponing, and this cascades in an unfortunate way: postponing yields more unsolved metas, which cause more postponing.

## 4 TELESCOPES AND STRICTLY CURRIED FUNCTIONS

As part of the proposed solution, we extend the core theory with telescopes and strictly curried functions. Figure 3 lists the typing rules and definitional equalities.

### 4.1 Telescopes

Telescopes can be viewed as a generic implementation of record types. We have $\mathsf{Tel}$ as the type of telescopes. Elements of $\mathsf{Tel}$ are right-nested telescopes of types, with $\epsilon$ denoting the empty telescope, and $- \triangleright -$ telescope extension. For example, we can define the signature of natural number algebras as follows:

$$\mathbf{let}\ \mathit{NatAlgSig} : \mathsf{Tel} = (N : \mathsf{U}) \triangleright (\mathit{zero} : N) \triangleright (\mathit{suc} : N \to N) \triangleright \epsilon\ \mathbf{in} \ldots$$

We interpret an $A : \mathsf{Tel}$ as a record type as $\mathsf{Rec}\, A$, which behaves as the evident iterated $\Sigma$-type corresponding to the telescope. Hence, $\mathsf{Rec}\,\epsilon$ is isomorphic to the unit type, with inhabitant $[]$, and $\mathsf{Rec}\,((x : A) \triangleright B)$ behaves as a $\Sigma$-type, with pairing constructor $- :: -$ and projections $\pi_1$ and $\pi_2$. We also have the $\beta$ and $\eta$ rules for record constructors and projections in Figure 3. We present definitional equalities in a compact form, but note that they still stand for $\boxed{\Theta|\Gamma \vdash t \equiv u : A}$ judgments. Hence, the sides of the equations must have the same types, and in particular the left side of the $[]$-$\eta$ rule has type $\mathsf{Rec}\,\epsilon$.

Telescopes and records are derivable from natural numbers, the unit type and $\Sigma$-types. We use Agda-like pattern matching notation in the following. First, we define length-indexed telescopes.

$$\begin{aligned}&\mathsf{Tel}' : \mathsf{Nat} \to \mathsf{U}\\&\mathsf{Tel}'\,\mathsf{zero} \quad\;\; :\equiv \top\\&\mathsf{Tel}'\,(\mathsf{suc}\,n) :\equiv \Sigma(A : \mathsf{U}).\,(A \to \mathsf{Tel}'\,n)\end{aligned}$$

Then, we have $\mathsf{Tel} :\equiv \Sigma(n : \mathsf{Nat}).\,(\mathsf{Tel}'\,n)$, and define records:

$$\begin{aligned}&\mathsf{Rec} : \mathsf{Tel} \to \mathsf{U}\\&\mathsf{Rec}\,(\mathsf{zero},\,\_) \qquad\quad :\equiv \top\\&\mathsf{Rec}\,(\mathsf{suc}\,n,\,(A,\,B)) :\equiv \Sigma(a : A).\,(\mathsf{Rec}\,(n,\,B\,a))\end{aligned}$$

From the above, $\epsilon$, $- \triangleright -$, $- :: -$ and $[]$ are evident, and all expected equalities hold definitionally. Derivability is good news because we inherit nice properties of the type theory which only contains the base type formers. For instance, if we have a consistent universe setup[3], we inherit consistency, canonicity and decidability of conversion. We currently use native telescopes instead of Nat, $\top$ and $\Sigma$ because in unification and elaboration it is convenient that we are able to restrict some types to records types.

### 4.2 Strictly Curried Functions

These are function types whose domains are telescopes, and they are immediately computed to iterated implicit function types when the domain telescope is canonical. See FUN-$\epsilon$ and FUN-$\triangleright$: a curried function with empty domain computes to simply the codomain, while a function with a non-empty domain computes to an implicit function type. We explicitly notate telescopes in both $\lambda$-abstractions and applications for strictly curried functions, since they are relevant in the computation rules.

Curried function types tend to be computed away, but they can persist if the domain telescope is neutral, and in particular when it is meta-headed. For example, assuming a meta $\alpha$ : Tel, the type $\{x : \overline{\alpha}\} \to B$ cannot be computed further. During elaboration, we will use strictly curried function types to represent unknown insertions, but these types are eventually computed away if a surface expression can be successfully elaborated. Since the surface language remains unchanged, telescopes and curried functions are merely an internal implementation detail from the perspective of programmers.

Curried functions are *mostly* derivable from Nat, $\top$ and $\Sigma$. The type former is defined as follows:

$$\begin{aligned}&\Pi^C : (A : \mathsf{Tel}) \to (\mathsf{Rec}\,A \to \mathsf{U}) \to \mathsf{U}\\&\Pi^C(\mathsf{zero},\,\_)\,B \qquad\;\; :\equiv B\,\mathsf{tt}\\&\Pi^C(\mathsf{suc}\,n,\,(A,\,B))\,C :\equiv \{a : A\} \to \Pi^C\,(n,\,B\,a)(\lambda\,b.\,C\,(a,\,b))\end{aligned}$$

With this, we can also define $\mathsf{app} : \Pi^C\,A\,B \to (a : \mathsf{Rec}\,A) \to B\,a$ and $\mathsf{lam} : ((a : \mathsf{Rec}\,A) \to B\,a) \to \Pi^C\,A\,B$, and all equations in Figure 3 hold definitionally, except CURRIED-$\beta$ and CURRIED-$\eta$. These do not hold strictly, because $\Pi^C$, app and lam are all defined by recursion on the $A$ telescope, but the $\beta\eta$ rules are specified generically for arbitrary (possibly neutral) telescopes. CURRIED-$\beta$ is still provable as a propositional equality, and assuming function extensionality CURRIED-$\eta$ is provable as well. For details, see our Agda formalization of these definitions, which is included alongside the prototype implementation.

[3]Recall that we currently assume type-in-type, which causes consistency and normalization to fail.

$$\textsc{Tel}\quad \frac{}{\Theta|\Gamma \vdash \mathsf{Tel} : \mathsf{U}} \qquad \textsc{Empty-Tel}\quad \frac{}{\Theta|\Gamma \vdash \epsilon : \mathsf{Tel}} \qquad \textsc{Nonempty-Tel}\quad \frac{\Theta|\Gamma \vdash A : \mathsf{U} \qquad \Theta|\Gamma,\, x : A \vdash B : \mathsf{Tel}}{\Theta|\Gamma \vdash (x : A) \triangleright B : \mathsf{Tel}}$$

$$\textsc{Record-Type}\quad \frac{\Theta|\Gamma \vdash A : \mathsf{Tel}}{\Theta|\Gamma \vdash \mathsf{Rec}\, A : \mathsf{U}} \qquad \textsc{Empty-Record}\quad \frac{}{\Theta|\Gamma \vdash [] : \mathsf{Rec}\, \epsilon} \qquad \textsc{Nonempty-Record}\quad \frac{\Theta|\Gamma \vdash t : A \qquad \Theta|\Gamma \vdash u : \mathsf{Rec}\, (B[x \mapsto t])}{\Theta|\Gamma \vdash t :: u : \mathsf{Rec}\, ((x : A) \triangleright B)}$$

$$\textsc{Record-Projection}\quad \frac{\Theta|\Gamma \vdash t : \mathsf{Rec}\, ((x : A) \triangleright B)}{\Theta|\Gamma \vdash \pi_1\, t : A \qquad \Theta|\Gamma \vdash \pi_2\, t : \mathsf{Rec}\, (B[x \mapsto \pi_1\, t])} \qquad \textsc{Curried-Fun}\quad \frac{\Theta|\Gamma \vdash A : \mathsf{Tel} \qquad \Theta|\Gamma,\, x : \mathsf{Rec}\, A \vdash B : \mathsf{U}}{\Theta|\Gamma \vdash \{x : \overline{A}\} \to B : \mathsf{U}}$$

$$\textsc{Curried-Lam}\quad \frac{\Theta|\Gamma,\, x : \mathsf{Rec}\, A \vdash t : B}{\Theta|\Gamma \vdash \lambda\, \{x : \overline{A}\}.\, t : \{x : \overline{A}\} \to B} \qquad \textsc{Curried-App}\quad \frac{\Theta|\Gamma \vdash t : \{x : \overline{A}\} \to B \qquad \Theta|\Gamma \vdash u : \mathsf{Rec}\, A}{\Theta|\Gamma \vdash t\, \{u : \overline{A}\} : B[x \mapsto u]}$$

$$\begin{array}{lll}
\pi_1\text{-}\beta & \pi_1\, (t :: u) & \equiv t \\
\pi_2\text{-}\beta & \pi_2\, (t :: u) & \equiv u \\
::\text{-}\eta & (\pi_1\, t :: \pi_2\, t) & \equiv t \\
[]\text{-}\eta & t & \equiv [] \\
\textsc{fun-}\epsilon & \{x : \overline{\epsilon}\} \to B & \equiv B[x \mapsto []] \\
\textsc{fun-}\triangleright & \{x : \overline{(y : A) \triangleright B}\} \to C & \equiv \{y : A\} \to (\{b : \overline{B}\} \to C[x \mapsto (y :: b)]) \\
\textsc{lam-}\epsilon & \lambda\, \{x : \overline{\epsilon}\}.\, t & \equiv t[x \mapsto []] \\
\textsc{lam-}\triangleright & \lambda\, \{x : \overline{(y : A) \triangleright B}\}.\, t & \equiv \lambda\{y\}.\, \lambda\, \{b : \overline{B}\}.\, t[x \mapsto (y :: b)] \\
\textsc{app-}\epsilon & t\, \{u : \overline{\epsilon}\} & \equiv t \\
\textsc{app-}\triangleright & t\, \{u : \overline{(x : A) \triangleright B}\} & \equiv t\, \{\pi_1\, u\}\, \{\pi_2\, u : \overline{B[x \mapsto \pi_1\, u]}\} \\
\textsc{curried-}\beta & \lambda\, (\{x : \overline{A}\}.\, t)\, \{u : \overline{A}\} & \equiv t[x \mapsto u] \\
\textsc{curried-}\eta & \lambda\, \{x : \overline{A}\}.\, t\, \{x : \overline{A}\} & \equiv t
\end{array}$$

Fig. 3. Rules for telescopes and strictly curried functions.

Hence, we can derive a somewhat weaker version of curried functions, with propositional $\beta$ and $\eta$. From this, we still get consistency and canonicity very cheaply. This is because models proving consistency or canonicity for the base theory with Nat, $\top$ and $\Sigma$ usually support equality reflection. For consistency, standard set-theoretical models and models in extensional type theory have this property, for canonicity, glued models e.g. as in [Kaposi et al. 2019; Sterling 2019] also have this property.

In contrast, showing normalization and decidability of conversion would require some extra work. We leave this to future work, but we expect that it is not difficult to extend previous proofs to cover strict $\beta$ and $\eta$ for curried functions.

## 5 EXTENDING ELABORATION

We shall utilize the extended core theory to implement smarter elaboration. Recall from Section 3 that the old elaborator makes two kinds of unforced insertion choices:

(1) $[\![t]\!] \Downarrow_{\Theta|\Gamma} A$ does not insert an implicit $\lambda$ when $A$ is meta-headed.
(2) insert $(\Theta,\ \theta,\ t,\ A)$ does not insert an implicit application when $A$ is meta-headed.

In the following, we shall only enhance $\lambda$-insertions. This allows a simple implementation which only requires minimal changes to unification, and which is already remarkably powerful. It seems that enhancing implicit application insertions requires extending unification; we discuss this in Section 7.3. First, we modify closing types and contextualization to take advantage of telescopes.

*Definition 5.1 (Closing types).* We use curried function types to close over record types in the scope. If a bound variable does not have a record type, then we do as before[4]. We prepend the following clause to Definition 2.3:

$$((\Gamma,\ x : \mathsf{Rec}\, A) \Rightarrow B) :\equiv (\Gamma \Rightarrow (\{x : \overline{A}\} \to B))$$

*Definition 5.2 (Contextualization).* We extend spine notation to applications of curried functions. For example, we may have a spine $\overline{t} \equiv (\{x : \overline{A}\},\ \{y : \overline{B}\})$. We accordingly revise Definition 2.4 for $\overline{\mathsf{vars}_\Gamma}$ so that we use curried function application for each record type in $\Gamma$.

### 5.1 Handling Superfluous Implicit Functions

Before we can move on to unification and elaboration, we have to address a curious issue. Assuming $Bool : \mathsf{U}$, $true : Bool$ and $false : Bool$, consider the following surface expression:

$$\mathbf{let}\, x : \_ = true\ \mathbf{in}\ x$$

What should this expression elaborate to? We would expect the result to be simply

$$\mathbf{let}\, x : Bool = true\ \mathbf{in}\ x$$

However, there are infinitely many core terms which are conservative over the surface expression in the sense of Theorem 2.8. That is, we can wrap definitions with any number of implicit $\lambda$-s, and add implicit applications accordingly to usage sites of the defined name. For example, we could have

$$\mathbf{let}\, x : \{y : Bool\} \to Bool = \lambda\, \{y\}.\, true\ \mathbf{in}\ x\, \{true\}$$

This is clearly undesirable. With the type $\{y : Bool\} \to Bool$, the implicit argument $y$ is never inferable, because the codomain type does not depend on the domain, and the argument is never constrained. Hence, with the above definition, we always have to write $x\,\{true\}$ or $x\,\{false\}$ when we want to use $x$. In order to avoid such nonsense, we adopt the following principle: *elaboration should never invent non-dependent implicit function types.*

This was a non-issue in the old elaborator, because it was not able to invent implicit function types; it was only utilizing the type annotations present in the surface input. In the case of $\mathbf{let}\, x : \_ = true$, the old elaborator checks $true$ with a fresh meta, and just assumes that the meta does not stand for an implicit function type.

[4] This implies that we close over meta-headed types using plain functions. In theory, this causes a higher-order version of the basic implicit insertion problem: we are uncertain about whether we should be uncertain about implicit insertions. So far, this higher-order insertion problem seems to be irrelevant in practice, in the prototype implementation.

$$\frac{\Theta|\Gamma,\, x : \mathsf{Rec}\,A \vdash B : \mathsf{U}}{\Theta,\, \mathsf{constancy}_{\Gamma,\, x:\mathsf{Rec}\,A}\, B \vdash}\ \textsc{metacon/constancy}$$

$$\frac{x \notin \mathsf{FreeVars}(B)}{\Theta_0,\, \mathsf{constancy}_{\Gamma_0,\, x:\mathsf{Rec}\,A}\, B,\, \Theta_1|\Gamma_1 \vdash A \equiv \epsilon : \mathsf{Tel}}\ \textsc{constancy-}\equiv$$

$$\frac{\theta : \Theta_0 \Rightarrow \Theta_1 \qquad x \notin \mathsf{FreeVars}(B[\theta]) \text{ implies } \Theta_0|\Gamma[\theta] \vdash A[\theta] \equiv \epsilon : \mathsf{Tel}}{(\theta,\, \mathsf{solve}_{\Gamma,\, x:\mathsf{Rec}\,A}\, B) : \Theta_0 \Rightarrow (\Theta_1,\, \mathsf{constancy}_{\Gamma,\, x:\mathsf{Rec}\,A}\, B)}\ \textsc{metasub/constancy}$$

$$\frac{}{\mathsf{p} : (\Theta,\, \mathsf{constancy}_{\Gamma,\, x:\mathsf{Rec}\,A}\, B) \Rightarrow \Theta}\ \textsc{metasub/weaken-constancy}$$

Fig. 4. Rules for constancy constraints

*5.1.1 Constancy constraints.* We use these constraints to get rid of curried function types as soon as we learn that they are non-dependent. They are constraints in the usual sense in unification algorithms (e.g. as in [Abel and Pientka 2011] or [Vytiniotis et al. 2011]). We formalize them in a compact way, by adding a new kind of context extension for metacontexts. The rules are given in Figure 4.

In the rule METACON/CONSTANCY we specify extension of a metacontext with a constraint. The CONSTANCY-≡ rule expresses that, assuming we have a constancy constraint for $A$ and $B$ in context, if $B$ does not depend on the $x : \mathsf{Rec}\,A$ domain variable, then $A$ is equal to the empty telescope $\epsilon$.

The METASUB/CONSTANCY rule defines metasubstitutions whose codomains are extended with constraints. Intuitively, while the METASUB/EXTENDED rule from Section 2.3 can be used to solve a metavariable (by mapping it to a term), METASUB/CONSTANCY solves a constraint. We can only extend $\theta : \Theta_0 \Rightarrow \Theta_1$ to map into an additional constraint if $\theta$ forces the constraint to hold. In METASUB/WEAKEN-CONSTANCY, we overload p for the weakening substitution which drops a constraint.

*Definition 5.3 (Creating a new constraint).* We do this similarly to Definition 2.6, by simply returning a weakening substitution.

$$\mathsf{newConstancy}_{\Theta|\Gamma,\, x:\mathsf{Rec}\,A}\, B :\equiv ((\Theta,\, \mathsf{constancy}_{\Gamma,\, x:\mathsf{Rec}\,A}\, B),\, \mathsf{p})$$

*5.1.2 Algorithmic implementation of constraint solving.* The above specification for constancy constraints is compact but not particularly algorithmic: we just magically get new definitional equalities whenever we have constraints in contexts. In our prototype implementation, we implement eager removal of solvable constraints.

After solving a meta $\alpha$ during unification, which yields a unifying $\theta$ substitution, we review all $(\mathsf{constancy}_{\Gamma,\, x:\mathsf{Rec}\,A}\, B)$ constraints in the context, such that $x$ occurs in $B$ inside a $\bar{t}$ spine of some $\alpha\,\bar{t}$ term. In other words, we review constraints where the new meta solution might make a difference.

(1) If we have $x \in \mathsf{FreeVars}(B[\theta])$, where $x$ occurs rigidly in $B[\theta]$, i.e. the occurrence is not in a spine of a meta, then no metasubstitution can possibly remove this occurrence. In this case the constraint holds vacuously, so we can use the METASUB/CONSTANCY rule to return a $\theta'$ substitution which also solves the constraint.
(2) If we have $x \notin \mathsf{FreeVars}(B[\theta])$, we recursively unify $A[\theta]$ with $\epsilon$. If that succeeds, we get a $\theta'$ which unifies $A[\theta]$ and $\epsilon$ and thus forces the constraint to hold, so we can again use METASUB/CONSTANCY to solve the constraint.
(3) In any other case we simply return $\theta$ and keep the constraint around.

Also, when we create a new constancy constraint, we immediately review it as described above.

*Remark.* In the case with $x \in \mathsf{FreeVars}(B[\theta])$, it would be also sound to solve the constraint when the occurrence is not rigid. However, this way we could lose potential non-$\epsilon$ solutions of $A[\theta]$.

### 5.2 Unification For Strictly Curried Functions

Although we omit most details of unification, we shall discuss it for curried functions, as it is essential in the extended elaboration algorithm. The most interesting case is when we unify a curried function type with an implicit function type. In this case, we learn that the domain of the curried function is non-empty, so we refine the $A$ domain to an extended $(x_0 : A_0) \triangleright A_1$ telescope. Since we invent a fresh $A_1$ domain for a curried function type, we need to add a constancy constraint for it as well.

$$
\begin{aligned}
&\mathsf{unify}_{\Theta_0\,|\,\Gamma}(\{x : \overline{A}\} \to B)\,(\{x_0 : A_0\} \to B') :\equiv \mathbf{do}\\
&\quad \mathbf{let}\,(\Theta_1,\,\theta_1,\,A_1) = \mathsf{freshMeta}_{\,\Theta_0\,|\,\Gamma,\,x_0:A_0}\,\mathsf{Tel}\\
&\quad (\Theta_2,\,\theta_2) \leftarrow \mathsf{unify}_{\Theta_1\,|\,\Gamma[\theta_1]}\,(A[\theta_1])\,((x : A_0[\theta_1]) \triangleright A_1)\\
&\quad \mathbf{let}\,(\Theta_3,\,\theta_3) = \mathsf{newConstancy}_{\Theta_2\,|\,\Gamma[\theta_{12}],\,x_0:A_0[\theta_{12}],\,x_1:\mathsf{Rec}\,(A_1[\theta_2])}\,(B[\theta_{12}][x \mapsto (x_0 :: x_1)])\\
&\quad \mathsf{unify}_{\Theta_3\,|\,\Gamma[\theta_{123}],\,x_0:A_0[\theta_{123}]}\,(\{x_1 : A_1[\theta_{23}]\} \to B[\theta_{123}][x \mapsto (x_0 :: x_1)])\,B'
\end{aligned}
$$

We have the symmetric $\mathsf{unify}_{\Theta_0\,|\,\Gamma}(\{x_0 : A_0\} \to B')\,(\{x : \overline{A}\} \to B)$ case the same way as above.

Now, let us assume that $B'$ is not an implicit function type, curried function type or meta-headed. Then, we have the following case, where we solve a telescope domain to be empty.

$$
\begin{aligned}
&\mathsf{unify}_{\Theta_0\,|\,\Gamma}(\{x : \overline{A}\} \to B)\,B' :\equiv \mathbf{do}\\
&\quad (\Theta_1,\,\theta_1) \leftarrow \mathsf{unify}_{\Theta_0\,|\,\Gamma}\,A\,\epsilon\\
&\quad \mathsf{unify}_{\Theta_1\,|\,\Gamma[\theta_1]}\,(B[\theta_1][x \mapsto []])\,(B'[\theta_1])
\end{aligned}
$$

Again, we also have the symmetric case. For $\lambda\,\{x : \overline{A}\}.\,t$ and $t\,\{u : \overline{A}\}$, unification is structural, and other cases remain the same as in the basic elaborator of Section 2.

### 5.3 Elaboration

In the definition of checking, we insert a new clause after $[\![t]\!] \Downarrow_{\,\Theta\,|\,\Gamma}(\{x : A\} \to B)$:

$$
\begin{aligned}
&[\![t]\!] \Downarrow_{\,\Theta_0\,|\,\Gamma}(\alpha\,\overline{u}) :\equiv \mathbf{do}\\
&\quad \mathbf{let}\,(\Theta_1,\,\theta_1,\,A) = \mathsf{freshMeta}_{\,\Theta_0\,|\,\Gamma}\,\mathsf{Tel}\\
&\quad (\Theta_2,\,\theta_2,\,t',\,B) \leftarrow [\![t]\!] \Uparrow_{\,\Theta_1\,|\,\Gamma[\theta_1],\,x:\mathsf{Rec}\,A}\\
&\quad \mathbf{let}\,(\Theta_3,\,\theta_3) = \mathsf{newConstancy}_{\Theta_2\,|\,\Gamma[\theta_{12}],\,x:\mathsf{Rec}\,(A[\theta_2])}\,B\\
&\quad (\Theta_4,\,\theta_4) \leftarrow \mathsf{unify}\,((\alpha\,\overline{u})[\theta_{123}])\,(\{x : \overline{A[\theta_{23}]}\} \to B[\theta_3])\\
&\quad \mathbf{return}\,(\Theta_4,\,\theta_{1234},\,(\lambda\,\{x : \overline{A[\theta_{234}]}\}.\,t'[\theta_{34}]))
\end{aligned}
$$

Hence, when checking a term with a meta-headed type, we create a fresh meta with Tel type, infer a type for the term, and wrap the result in a strictly curried $\lambda$. We again need to create new constancy constraint. This way, if $B$ does not depend on $x : \mathsf{Rec}\,A$, the strictly curried $\lambda$ in the output immediately computes away, since $A$ is solved to $\epsilon$.

This concludes the definition of the extended elaborator. The new algorithm is sound with respect to the extended core syntax, and it also has the conservativity property from Theorem 2.8 if we additionally allow elaboration to insert $\lambda$-s for strictly curried functions. We present some examples of the algorithm in action.

*Example 5.4.* We return to the *polyList* example from Section 3. We trace elaboration using the extended algorithm.

$$
\begin{array}{ll}
0 & \llbracket \mathit{cons}\,(\lambda\, x.\, x)\; \mathit{nil} \rrbracket \Downarrow\; (\mathit{List}\,(\{A : \mathsf{U}\} \to A \to A)) \\
1 & \quad \llbracket \mathit{cons}\,(\lambda\, x.\, x)\; \mathit{nil} \rrbracket \Uparrow \\
2 & \quad\quad \llbracket \mathit{cons}\,(\lambda\, x.\, x) \rrbracket \Uparrow \\
3 & \quad\quad\quad \llbracket \mathit{cons} \rrbracket \Uparrow \\
4 & \quad\quad\quad = \mathit{cons}\,\{\alpha_0\} \;:\; \alpha_0 \to \mathit{List}\,\alpha_0 \to \mathit{List}\,\alpha_0 \\
5 & \quad\quad\quad \llbracket \lambda\, x.\, x \rrbracket \Downarrow\; \alpha_0 \\
6 & \quad\quad\quad = \lambda\,\{y : \overline{\alpha_1}\}.\, \lambda\, x.\, x \\
7 & \quad\quad = \mathit{cons}\,\{\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\}\}\,(\lambda\,\{y : \overline{\alpha_1}\}.\, \lambda\, x.\, x) \\
8 & \quad\quad \llbracket \mathit{nil} \rrbracket \Downarrow\; (\mathit{List}\,(\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\})) \\
9 & \quad\quad = \mathit{nil}\,\{\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\}\} \\
10 & \quad = \mathit{cons}\,\{\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\}\}\,(\lambda\,\{y : \overline{\alpha_1}\}.\, \lambda\, x.\, x) \\
 & \quad\quad\quad\quad (\mathit{nil}\,\{\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\}\}) \\
 & \quad\quad\quad : \mathit{List}\,(\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\}) \\
11 & \quad \mathsf{unify}\,(\mathit{List}\,(\{A : \mathsf{U}\} \to A \to A))\,(\mathit{List}\,(\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\})) \\
12 & \quad\quad \mathsf{unify}\,(\{A : \mathsf{U}\} \to A \to A)\,(\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\}) \\
13 & \quad\quad\quad \mathsf{unify}\,\alpha_1\,((A : \mathsf{U}) \triangleright (\alpha_3\,A)) \\
14 & \quad\quad\quad \mathsf{unify}\,(A \to A)\,(\{z : \overline{\alpha_3\,A}\} \to \alpha_2\,\{A\}\,\{z : \overline{\alpha_3\,A}\} \to \alpha_2\,\{A\}\,\{z : \overline{\alpha_3\,A}\}) \\
15 & \quad\quad\quad\quad \mathsf{unify}\,(\alpha_3\,A)\,\epsilon \\
16 & \quad\quad\quad\quad \mathsf{unify}\,(A \to A)\,(\alpha_2\,\{A\} \to \alpha_2\,\{A\}) \\
17 & \quad\quad\quad\quad\quad \mathsf{unify}\,A\,(\alpha_2\,\{A\}) \\
18 & \quad\quad\quad\quad\quad \mathsf{unify}\,A\,A \\
19 & = \mathit{cons}\,\{\{A : \mathsf{U}\} \to A \to A\}\,(\lambda\,\{A\}\,x.\, x)\,(\mathit{nil}\,\{\{A : \mathsf{U}\} \to A \to A\})
\end{array}
$$

We diverge from the previous attempt at line 5. Here, we check $\lambda\, x.\, x$ with the meta $\alpha_0$, so we create a fresh $\alpha_1$ : Tel meta and wrap the result as $\lambda\,\{y : \overline{\alpha_1}\}.\, \lambda\, x.\, x$. This result has the inferred type $\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\} \to \alpha_2\,\{y : \overline{\alpha_1}\}$, and we promptly unify $\alpha_0$ (which stands for the list element type) with it. On line 10, we return the elaborated *cons* expression, where the list element type is made explicit. It only remains to unify the inferred and expected types. On line 12 we have the case where a curried function type is matched with an implicit function type, so on line 13 we refine $\alpha_1$ to a non-empty telescope, and proceed unifying the domains. Note that on line 14 we have $\alpha_2\,\{A\}\,\{z : \overline{\alpha_3\,A}\}$, which results from applying the substitution $y \mapsto (A :: z)$, and computing $\alpha_2\,\{(A :: z) : \overline{(A : \mathsf{U}) \triangleright (\alpha_3\,A)}\}$ further using the App-$\triangleright$ rule from Figure 3.

On line 14 we have the case when a curried function type is matched with a type which is not a curried function, implicit function or a meta-headed type. We accordingly unify domain telescope $\alpha_3\,A$ with $\epsilon$, which causes $\alpha_3$ to be solved as $\lambda\, A.\, \epsilon$ according to standard pattern unification. Now, $\alpha_2\,\{A\}\,\{z : \overline{\alpha_3\,A}\}$ computes to $\alpha_2\,\{A\}\,\{z : \epsilon\}$, which further computes to $\alpha_2\,\{A\}$ by App-$\epsilon$. From here, unification finishes with success. On the last line, the expected output is returned: since the $\alpha_1$ telescope meta is now solved, curried function types and abstractions are computed away.

*Example 5.5.* We illustrate now the action of constancy constraints, using the example from Section 5.1. Again, assume $Bool : \mathsf{U}$ and $true : Bool$.

$$\begin{array}{ll}
0 & [\![\mathbf{let}\, x : _ = true\, \mathbf{in}\, x]\!] \Uparrow \\
1 & \quad [\![_]\!] \Downarrow \mathsf{U} = \alpha_0 \\
3 & \quad [\![true]\!] \Downarrow \alpha_0 \\
4 & \qquad \mathsf{freshMeta\ Tel} = \alpha_1 \\
6 & \qquad [\![true]\!] \Uparrow = true : Bool \\
8 & \qquad \mathsf{newConstancy}_{\bullet,\, x:\mathsf{Rec}\,\alpha_1}\, Bool \\
9 & \qquad\quad \mathsf{unify}\, \alpha_1\, \epsilon \\
10 & \qquad \mathsf{unify}\, \alpha_0\, Bool \\
11 & \quad\; = true \\
12 & \quad [\![x]\!] \Uparrow \mathsf{true} = true : Bool \\
14 & = (\mathbf{let}\, x : Bool = true\, \mathbf{in}\, x) : Bool
\end{array}$$

First we create a fresh meta $\alpha_0$ for the hole on line 2, then we check true with it. Since we are checking with a meta-headed type, we create a fresh telescope meta $\alpha_1$, then infer *Bool* for *true*. On line 8 we create a constancy constraint and immediately try to solve it, as described in Section 5.1.2. Since *Bool* does not depend on the domain, we solve the constraint and thereby solve $\alpha_1$ as $\epsilon$. Hence, we simply return *true* on line 11, since $\lambda\{x : \overline{\epsilon}\}.\, true$ computes to that, and elaboration succeeds with the expected output.

## 6 IMPLEMENTATION

We provide an implementation of the elaborator from Section 5. It is a standalone Haskell program which reads a surface expression from standard input, and outputs the result of elaboration, or optionally the type or the normal form of the result. It is implemented in 1056 lines of Haskell. Of this, 669 lines constitute the core syntax, evaluation, unification and elaboration. Of these 669 lines, 454 lines implement the basic elaborator of Section 2 and 215 lines implement the extended elaborator.

Elaboration is implemented in the style of Coquand's algorithm [Coquand 1996], where elaboration is interleaved with normalization-by-evaluation. Hence, we do not perform any substitution operations on core syntax, instead we evaluate core terms into the semantic domain, and perform unification, strengthening and occurrence checking on semantic values. There is also a quoting (or readback) operation which yields normalized core terms from values, and which is used most prominently when we have to generate solutions for metavariables.

Metacontexts are a key point in the implementation. We avoid the tedious (and inefficient) threading of composed metasubstitutions, instead we have a mutable reference which stores the current metasubstitution. We have a "forcing" operation which computes a semantic value to a head normal form with respect to the current metasubstitution. Hence, instead of constantly updating every term and type by performing metasubstitution, we only force them whenever we need to pattern match on the shape of types, for example when we are inserting implicit arguments based on types. In this paper we still stick with the "threaded metasubstitution" presentation because it is more conventional, and also more straightforward to formalize.

We use normalization-by-evaluation in Abel's style (see e.g. [Abel 2013, Chapter 3]), where semantic values use de Bruijn levels, and the core syntax uses de Bruijn indices. This is practically

| | | | |
|---|---|---|---|
| $IdTy$ | $:\equiv \{A : \mathsf{U}\} \to A \to A$ | $inc$ | $: Int \to Int$ |
| $single$ | $: \{A : \mathsf{U}\} \to A \to List\, A$ | $auto$ | $: IdTy \to IdTy$ |
| $id$ | $: IdTy$ | $auto'$ | $: \{B : \mathsf{U}\} \to IdTy \to B \to B$ |
| $ids$ | $: List\, IdTy$ | $choose$ | $: \{A : \mathsf{U}\} \to A \to A \to A$ |
| $nil$ | $: \{A : \mathsf{U}\} \to List\, A$ | $app$ | $: \{A\, B : \mathsf{U}\} \to (A \to B) \to A \to B$ |
| $cons$ | $: \{A : \mathsf{U}\} \to A \to List\, A \to List\, A$ | $revapp$ | $: \{A\, B : \mathsf{U}\} \to A \to (A \to B) \to B$ |
| $head$ | $: \{A : \mathsf{U}\} \to List\, A \to A$ | $runST$ | $: \{A : \mathsf{U}\} \to (\{S : \mathsf{U}\} \to ST\, S\, A) \to A$ |
| $tail$ | $: \{A : \mathsf{U}\} \to List\, A \to List\, A$ | $argST$ | $: \{S : \mathsf{U}\} \to ST\, S\, Int$ |
| $map$ | $: \{A\, B : \mathsf{U}\} \to (A \to B) \to List\, A \to List\, B$ | $poly$ | $: IdTy \to Pair\, Int\, Bool$ |

Fig. 5. Types used in Figure 6.

very favorable, because the evaluator never has to perform weakening on values. The implementation of curried functions presents a bit of a complication, because the computation rules are type-directed, so we have to annotate applications and abstractions with telescopes (just as in our notation for the core syntax). We implement constraints in a fairly optimized way: we keep track of relevant "blocking" metas for each constraint and upon solving a meta we only review constraints which were blocked on the meta.

## 7 RELATED WORKS AND EVALUATION

In this section we examine how the elaborator fares in practice, its limitations, and how it compares to related works. We refer to our elaborator as FCIF when comparing it to others.

### 7.1 Related Works

MLF [Le Botlan and Rémy 2014] extends System F with polymorphic subtyping bounds, and supports strong inference for first class polymorphism. HML [Leijen 2009]) is a simplified variant of MLF. These systems rely critically on subtyping and thus diverge markedly from (fragments of) Martin-Löf type theory. HMF [Leijen 2008] is a relatively simple system with first-class polymorphism; however, it is also weaker than others in terms of inferable annotations, and its more powerful variant (extended with n-ary application handling) is more complex.

There have been several attempts in the context of GHC. Boxy types [Vytiniotis et al. 2006] and FPH [Vytiniotis et al. 2008] were two early iterations which suffered from complex specification, complex implementation or fragility. More recent works are guarded impredicativity (GI) [Serrano et al. 2018] and quick look impredicativity (QL) [Serrano et al. 2020]. They both work by examining n-ary applications to find metavariable instantiations which arise from occurrences guarded by rigid type constructors. GI's use of *generalization constraints* anticipates our use of telescopes, but overall GI is also burdened by a great deal of complexity. QL streamlines GI by eschewing constraints in favor of an eager preprocessing pass on neutral expressions which finds polymorphic instantiations. QL also takes advantage of bidirectional type propagation. Although QL seems to be practically the most favorable so far, it is far from being elegant: it rechecks expressions multiple times, and the restriction of preprocessing to (nested) neutral applications is rather ad-hoc.

### 7.2 Evaluation

We borrow a very useful collection of inference benchmarks from the GI [Serrano et al. 2018] and QL [Serrano et al. 2020] papers: see Figure 5 and Figure 6. We also include a source file in

| | | |
|---|---|---|
| A | POLYMORPHIC INSTANTIATION | |
| A1 | $\lambda\, x\, y.\, y$<br>we infer a type which may be solved to $\{A\, B : \mathsf{U}\} \to A \to B \to A$<br>but not to $\{A : \mathsf{U}\} \to A \to \{B : \mathsf{U}\} \to B \to A$ | Yes* |
| A2 | *choose id* | Yes* |
| A3 | *choose nil ids* | Yes |
| A4 | $\lambda\, (x : \{A : \mathsf{U}\} \to A \to A).\, x\, x$ | Yes* |
| A5 | *id auto* | Yes |
| A6 | *id auto′* | Yes* |
| A7 | *choose id auto* | Yes |
| A8 | *choose id auto′* | No |
| A9 | $\lambda\, (f : \{A : \mathsf{U}\} \to (A \to A) \to \mathit{List}\, A \to A).\, f\, (\mathit{choose}\, \mathit{id})\, \mathit{ids}$ | Yes* |
| A10 | *poly id* | Yes |
| A11 | $\mathit{poly}\, (\lambda\, x.\, x)$ | Yes |
| B | INFERENCE OF POLYMORPHIC ARGUMENTS | |
| B1 | $\lambda\, f.\, \mathit{pair}\, (f\, \mathit{zero})\, (f\, \mathit{true})$ | No |
| B2 | $\lambda\, xs.\, \mathit{poly}\, (\mathit{head}\, xs)$ | No |
| C | FUNCTIONS ON POLYMORPHIC LISTS | |
| C1 | *length ids* | Yes |
| C2 | *tail ids* | Yes |
| C3 | *head ids* | Yes* |
| C4 | *single id* | Yes* |
| C5 | *cons id ids* | Yes |
| C6 | $\mathit{cons}\, (\lambda\, x.\, x)\, \mathit{ids}$ | Yes |
| C7 | $\mathit{append}\, (\mathit{single}\, \mathit{inc})\, (\mathit{single}\, \mathit{id})$ | Yes |
| C8 | $\lambda\, (g : \{A : \mathsf{U}\} \to \mathit{List}\, A \to \mathit{List}\, A \to A).\, g\, (\mathit{single}\, \mathit{id})\, \mathit{ids}$ | Yes* |
| C9 | $\mathit{map}\, \mathit{poly}\, (\mathit{single}\, \mathit{id})$ | Yes |
| C10 | $\mathit{map}\, \mathit{head}\, (\mathit{single}\, \mathit{ids})$ | Yes |
| D | APPLICATION FUNCTIONS | |
| D1 | *app poly id* | Yes |
| D2 | *revapp id poly* | Yes |
| D3 | *runST argST* | Yes |
| D4 | *app runST argST* | Yes |
| D5 | *revapp argST runST* | Yes |
| E | $\eta$-EXPANSION | |
| | assuming $k : \{A : \mathsf{U}\} \to A \to \mathit{List}\, A \to A$,<br>$h : \mathit{Int} \to \{A : \mathsf{U}\} \to A \to A$<br>and $\mathit{lst} : \mathit{List}\, (\{A : \mathsf{U}\} \to \mathit{Int} \to A \to A)$ | |
| E1 | *k h lst* | No |
| E2 | $k\, (\lambda\, x.\, h\, x)\, \mathit{lst}$ | Yes* |
| E3 | $\lambda\, (r : (\{A : \mathsf{U}\} \to A \to \{B : \mathsf{U}\} \to B \to B) \to \mathit{Int}).\, r\, (\lambda\, x\, y.\, y)$ | Yes |

"Yes*" means that our system can infer a type for the expression, but the lack of let-generalization yields unsolved metas in the type.

Fig. 6. Elaboration benchmark from [Serrano et al. 2018].

the implementation which reproduces these results. In ibid. a comparison is presented between multiple systems, but here we only include FCIF. The relative performance of FCIF here is easy to remember: it handles exactly the same cases as QL (which is slightly more than what GI covers). We mark some cases as "Yes*", where FCIF successfully infers a type, but since FCIF does no let-generalization, the inferred types are not fully constrained without extra contextual information.

In the $A1$ case, FCIF inserts only a single curried $\lambda$ on the outside, which is why the inferred type is not unifiable with $\{A : \mathsf{U}\} \to A \to \{B : \mathsf{U}\} \to B \to A$. However, this could be easily remedied by adding an extra curried $\lambda$ insertion in the definition of $[\![\lambda\, x.\, t]\!] \Uparrow_{\,\Theta|\Gamma}$.

In $A8$, the failure is intentional: the types of *id* and *auto′* are not unifiable because of the mismatched order of implicit and explicit arguments. We do not float out or reorder implicit arguments in any way, because we want to support arbitrary mixing of implicit/explicit arguments in the surface language, and such reordering would be problematic anyway in the presence of dependent types. The same situation arises in $E1$, where we cannot unify $\{A : \mathsf{U}\} \to Int \to A \to A$ with $Int \to \{A : \mathsf{U}\} \to A \to A$.

Some general comments on the comparison of FCIF to prior works. First, FCIF supports dependent types and prior solutions do not. It is also unclear whether prior solutions can scale to dependent types. MLF and HML rely on subtyping, and solutions which work one neutral spine at a time (HMF, GI, QL) also face issues with dependently typed spines. With such spines, we cannot simply process later arguments when previous ones are not yet elaborated, as return types depend on argument values, and skipping over arguments may clog up type computation in an unacceptable way.

Secondly, FCIF supports global inference. For an example for global inference, $\mathbf{let}\, x : \_ = single\ id\ \mathbf{in}\ \ cons\ \{IdTy\}\ x\ nil$ works in FCIF, MLF and HML but does not work in GI, QL and HMF. MLF and HML work here by immediately giving a principal type to *single id* which involves generalization with a subtyping bound. In contrast, FCIF makes no promises about principal typing — which is not feasible with dependent types — and does no generalization, but it can still infer a type for *id* which can be later constrained to $IdTy$.

### 7.3 Inferring Polymorphic Types for Function Arguments

It is apparent from the $B1$ and $B2$ cases on Figure 6 that FCIF cannot infer implicit function types for function arguments. MLF is the only system which can do this, in cases where the polymorphic argument is used only once. Consider the following:

$$\mathbf{let}\, f : IdTy \to Bool = \lambda\, \_.\, true\ \mathbf{in}\ \lambda\, x.\, f\ x$$

FCIF will attempt to elaborate $f\ x$ to $f\ (\lambda\{A\}.\, x)$, but fails to infer a type for $x$, because $A$ is not in the scope of $x$'s type. We sketch an extension of FCIF with *curried application insertion*, which could possibly handle this example. We extend insertion with the following case. We omit metacontexts and metasubstitutions for brevity.

$$\begin{aligned}
&\mathsf{insert}\,(t,\ \alpha\,\overline{u}) :\equiv \mathbf{do}\\
&\quad A \leftarrow \mathsf{freshMeta}_{\Gamma}\ \mathsf{Tel}\\
&\quad B \leftarrow \mathsf{freshMeta}_{\Gamma,\ x:\mathsf{Rec}\,A}\ \mathsf{U}\\
&\quad \mathsf{unify}\ (\alpha\,\overline{u})\ (\{x : \overline{A}\} \to B)\\
&\quad \mathsf{newConstancy}_{\Gamma, x:\mathsf{Rec}\,A}\ B\\
&\quad u \leftarrow \mathsf{freshMeta}_{\Gamma}\ (\mathsf{Rec}\,A)\\
&\quad \mathbf{return}\ (t\ \{u : \overline{A}\},\ B[x \mapsto u])
\end{aligned}$$

In short, we insert a new curried application whenever we insert with meta-headed type. Now, elaboration returns $f\ (\lambda\,\{A\}.\, x\ \{\alpha_2\, x\, A : \overline{\alpha_0}\})$, and right before we try to unify expected and inferred types, we have $x : \{y : \overline{\alpha_0}\} \to \alpha_1\ \{y : \overline{\alpha_0}\}$, and we have to compute unify $(\alpha_1\ \{\alpha_2\, x\, A : \overline{\alpha_0}\})\ (A \to A)$. This problem does not fall in any usual pattern fragment, and does not have an obvious most general solution. However, it is reasonable to make the following assumption: metas which return in unknown record types, like $\alpha_2$, are only solved with *order-preserving embeddings*, so for $\alpha_2$ the four potential solutions are $(\lambda\, x\, A.\, [])$, $(\lambda\, x\, A.\, (A :: []))$, $(\lambda\, x\, A.\, (x :: []))$ and $(\lambda\, x\, A.\, (x :: A :: []))$[5]. The reason for only allowing such solutions is that these can possibly yield solvable pattern unification problems, while solutions containing arbitrary terms or non-linear variable occurrences cannot. Out of these, $(\lambda\, x\, A.\, (x :: A :: []))$ and $(\lambda\, x\, A.\, (x :: []))$ are ruled out because they would require a cyclic solution for $\alpha_0$, and $(\lambda\, x\, A.\, [])$ is ruled out because it would yield the unsolvable unify $\alpha_1$ $(A \to A)$ problem. Thus, $(\lambda\, x\, A.\, (A :: []))$ is the unique order-preserving embedding solution in this case, which does yield the expected elaboration output. However, this idea is far from being fully fleshed out, and it could be subject of future research.

## 8 CONCLUSION AND FUTURE WORK

In type theory, it is a common endeavor to search for theories and features which confer the greatest amount of expressive power for the least amount of formal and conceptual complexity. In relation to elaboration and inference algorithms, we would similarly like to use tools and concepts with high power-to-weight ratio. Sometimes core theories need to be extended to allow more powerful elaboration. Certainly, type inference would be cumbersome without the notion of metavariables. Likewise, contextual metavariables are essential for elaboration in the presence of type dependencies.

In this paper we propose the concept of strictly curried functions, which supports elaboration of first-class implicit functions. It seems that trying to solve this problem by fiddling with postponing and heuristics, without extending the core theory, does not really cut it. In a dependently typed spine $t\, u\, v$, when elaborating $v$ we already want to have an elaborated version of $u$ with computational behavior. While contextual metavariables yield a nice modal type theory which computes in the presence of *unknown terms*, our curried functions extend this to a system which also computes in the presence of unknown *implicit insertions*. Since implicit insertions are in a way also just unknown terms, we can reuse most of the infrastructure of contextual/crisp type theory, and only add modest extensions.

Often it is the case that constructions on dependent type theory are forced to be more principled and structured than constructions on less powerful theories, because the many interactions and intricacies leave less room for whims. In our case, we believe that focusing on dependent type theory helped us hone in on the essential parts of the problem, and it is likely that restrictions of our solution would be also favorably simple and powerful in simpler settings such as System F.

In future work, we would like to

- Investigate extending elaboration with curried application insertions, along the lines of Section 7.3.
- Formalize elaboration and unification from an algebraic/categorical perspective, in particular give algebraic definitions for the core theories.
- Investigate whether features of our elaborator could be implemented in a reasonably practical way in production systems such as Agda. Also, investigate simplifying our algorithm to non-dependent (e.g. System F) settings.
- Investigate completeness and properties related to inferability.

## REFERENCES

Andreas Abel. 2013. Normalization by evaluation: dependent types and impredicativity. Habilitation thesis, Ludwig-Maximilians-Universität München.

[5]Technically, we could also allow permutations such as $\lambda\, x\, A.\, (A :: x :: [])$, but these are superfluous since nothing in our system depends on the ordering of permutable function arguments.

Andreas Abel and Brigitte Pientka. 2011. Higher-order dynamic pattern unification for dependent types and records. In *International Conference on Typed Lambda Calculi and Applications*. Springer, 10–26.

Thorsten Altenkirch and Ambrus Kaposi. 2016. Type theory in type theory using quotient inductive types. In *Proceedings of the 43rd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2016, St. Petersburg, FL, USA, January 20 - 22, 2016*, Rastislav Bodik and Rupak Majumdar (Eds.). ACM, 18–29. https://doi.org/10.1145/2837614.2837638

Lars Birkedal, Ranald Clouston, Bassel Mannaa, Rasmus Ejlers Møgelberg, Andrew M Pitts, and Bas Spitters. 2018. Modal dependent type theory and dependent right adjoints. *arXiv preprint arXiv:1804.05236* (2018).

Thierry Coquand. 1996. An algorithm for type-checking dependent types. *Science of Computer Programming* 26, 1-3 (1996), 167–177.

Joshua Dunfield and Neelakantan R Krishnaswami. 2013. Complete and easy bidirectional typechecking for higher-rank polymorphism. *ACM SIGPLAN Notices* 48, 9 (2013), 429–442.

Peter Dybjer. 1995. Internal type theory. In *International Workshop on Types for Proofs and Programs*. Springer, 120–134.

Richard A Eisenberg. 2016. *Dependent Types in Haskell: Theory and Practice*. University of Pennsylvania.

Richard A Eisenberg, Stephanie Weirich, and Hamidhasan G Ahmed. 2016. Visible type application. In *European Symposium on Programming*. Springer, 229–254.

Adam Michael Gundry. 2013. *Type inference, Haskell and dependent types*. Ph.D. Dissertation. University of Strathclyde.

Marcus Johansson and Jesper Lloyd. 2015. *Eliminating the problems of hidden-lambda insertion-Restricting implicit arguments for increased predictability of type checking in a functional programming language with depending types*. Master's thesis.

Ambrus Kaposi, Simon Huber, and Christian Sattler. 2019. Gluing for type theory. In *4th International Conference on Formal Structures for Computation and Deduction (FSCD 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Didier Le Botlan and Didier Rémy. 2014. MLF: raising ML to the power of System F. *ACM SIGPLAN Notices* 49, 4S (2014), 52–63.

Daan Leijen. 2008. HMF: Simple type inference for first-class polymorphism. In *Proceedings of the 13th ACM SIGPLAN international conference on Functional programming*. 283–294.

Daan Leijen. 2009. Flexible types: robust type inference for first-class polymorphism. In *Proceedings of the 36th annual ACM SIGPLAN-SIGACT symposium on Principles of programming languages*. 66–77.

Daniel R Licata, Ian Orton, Andrew M Pitts, and Bas Spitters. 2018. Internal universes in models of homotopy type theory. *arXiv preprint arXiv:1801.07664* (2018).

Aleksandar Nanevski, Frank Pfenning, and Brigitte Pientka. 2008. Contextual modal type theory. *ACM Transactions on Computational Logic (TOCL)* 9, 3 (2008), 1–49.

Ulf Norell. 2007. *Towards a practical programming language based on dependent type theory*. Ph.D. Dissertation. Chalmers University of Technology.

Steven Schäfer, Tobias Tebbi, and Gert Smolka. 2015. Autosubst: Reasoning with de Bruijn terms and parallel substitutions. In *International Conference on Interactive Theorem Proving*. Springer, 359–374.

Alejandro Serrano, Jurriaan Hage, Simon Peyton Jones, and Dimitrios Vytiniotis. 2020. A quick look at impredicativity. (January 2020). https://www.microsoft.com/en-us/research/publication/a-quick-look-at-impredicativity/ In submission.

Alejandro Serrano, Jurriaan Hage, Dimitrios Vytiniotis, and Simon Peyton Jones. 2018. Guarded impredicative polymorphism. In *Proceedings of the 39th ACM SIGPLAN Conference on Programming Language Design and Implementation*. 783–796.

Jonathan Sterling. 2019. Algebraic type theory and universe hierarchies. *arXiv preprint arXiv:1902.08848* (2019).

Dimitrios Vytiniotis, Simon Peyton Jones, Tom Schrijvers, and Martin Sulzmann. 2011. OutsideIn (X) Modular type inference with local assumptions. *Journal of functional programming* 21, 4-5 (2011), 333–412.

Dimitrios Vytiniotis, Stephanie Weirich, and Simon Peyton Jones. 2006. Boxy types: inference for higher-rank types and impredicativity. In *Proceedings of the eleventh ACM SIGPLAN international conference on Functional programming*. 251–262.

Dimitrios Vytiniotis, Stephanie Weirich, and Simon Peyton Jones. 2008. FPH: First-class polymorphism for Haskell. *ACM Sigplan Notices* 43, 9 (2008), 295–306.