Equational Reasoning about Programs with General Recursion and Call-by-value Semantics

Garrin Kimmell  
University of Iowa  
garrin-kimmell@uiowa.edu

Aaron Stump  
University of Iowa  
astump@acm.org

Harley D. Eades III  
University of Iowa  
harley-eades@uiowa.edu

Peng Fu  
University of Iowa  
peng-fu@uiowa.edu

Tim Sheard  
Portland State University  
sheard@cis.pdx.edu

Stephanie Weirich  
University of Pennsylvania  
sweirich@cis.upenn.edu

Chris Casinghino  
University of Pennsylvania  
ccasin@cis.upenn.edu

Vilhelm Sjöberg  
University of Pennsylvania  
vilhelm@cis.upenn.edu

Nathan Collins  
Portland State University  
nathan.collins@gmail.com

Ki Yung Ahn  
Portland State University  
kya@cs.pdx.edu

ABSTRACT
Dependently typed programming languages provide a mechanism for integrating verification and programming by encoding invariants as types. Traditionally, dependently typed languages have been based on constructive type theories, where the connection between proofs and programs is based on the Curry-Howard correspondence. This connection comes at a price, however, as it is necessary for the languages to be normalizing to preserve logical soundness. Trellys is a call-by-value dependently typed programming language currently in development that is designed to integrate a type theory with unsound programming features, such as general recursion, Type:Type, and arbitrary data types. In this paper we outline one core language design for Trellys, and demonstrate the use of the key language constructs to facilitate sound reasoning about potentially diverging programs.

KEYWORDS
1 Introduction

When writing verified programs, there are two traditional approaches. Theorem provers like ACL2 and Isabelle are used to perform verification externally [13, 24]. Programs are defined independently of the desired invariants, and then those invariants are verified after the fact. In addition to external verification, languages based on constructive type theory, such as Coq [34] and Agda [3] also support encoding program invariants internally, using dependent types. Specifications are tightly connected to code, and the burden of external proof can be reduced.

For both approaches, general recursion (or the definition of partial functions) poses challenges. Constructive type theories require functions to terminate on all inputs to preserve soundness of the logic under the Curry-Howard isomorphism. Sophisticated techniques, such as encoding possibly-diverging computations as co-inductive data, are required to define truly partial functions [5]. Alternatively, one can formulate a domain of definition for which the functions are, in fact, total, using an accessibility predicate [2]. This basic idea has also been used for higher-order logics [15]. Relatively few theories have been proposed for direct reasoning about general recursion; examples are LCF, LTC and VeriFun [20, 4, 35].

The Trellys project is a collaborative research initiative to design a dependently typed programming language with direct support for general recursion and other features such as Type:Type, which, like general recursion, are unsound under the Curry-Howard isomorphism. The goal of Trellys is to bridge the gap between dependently typed languages and program logics, allowing a programmer to utilize both internal and external verification techniques in the presence of these logically unsound features.

Trellys is also intended as a practical programming language. By removing the constraint that all programs terminate, we are forced to consider details such as evaluation strategy, since the termination behavior of a term in a language with general recursion can vary if the reduction strategy is changed. For Trellys, we have chosen call-by-value reduction because of the simplicity of the cost model it provides. This choice has far reaching consequences on the design of the logic used for verification in Trellys. In this paper, we address and propose solutions to a number of the issues encountered at this particular point in the design space: a call-by-value dependently typed language with general recursion.

We identify several problems encountered when trying to integrate dependent types with general recursion:

- How do we exploit the fact that inside programs, variables bound in those programs range over values, while allowing proofs abstractions to range over all programs\(^1\), including ones that diverge?

- The theory of call-by-value \(\beta\)-equality is fairly weak, given the restriction that the argument to \(\beta\)-reduction must be a value. How can we strengthen this equality to include all arguments that provably have values, but are not syntactically values?

- The natural way to prove theorems like associativity of append is by induction on the structure of a value. How can we strengthen such theorems to apply to all programs, including possibly diverging ones?

In solving these problems, we develop the main contributions of this paper:

- We define a judgmental notion of value that distinguishes variables contextually. This is achieved by marking variables introduced during quantification by whether or not they should be treated as values. The syntactic notion of value is then changed to be a judgmental notion, which classifies those variables as ranging over values or expressions, depending on how they are marked in the context.

- We integrate a notion of termination cast from our previous work (see Section 8) with call-by-value reduction, so that programs that are proved to be terminating can be considered to be values, for purposes of \(\beta\)-reduction.

- We introduce a non-computational termination case form, which allows us to case-split during reasoning on whether a program terminates or diverges. Using this, many theorems can be generalized to hold for all programs, not

\(^1\)We write “program” for terms in the non-logical fragment of the language, and write “proof” for terms in the logical fragment.
just terminating ones. This generalization is quite useful in practice, as it means that the theorems can be applied without needing to prove that their arguments are terminating.

This paper concerns the expressiveness of dependently-typed core languages. Its purpose is to explain these novel features through examples, describing the problems that non-termination and call-by-value reasoning bring to full-spectrum dependent type systems, and informally describing how these features can provide solutions to these problems. This paper discusses these ideas in the context of Sep\(^3\), one of several core language designs that we are developing in the Trellys project. The Sep\(^3\) core language is a work-in-progress, and still under development. As a result it has not been subject to metatheoretic study and proofs of standard type system results (such as type soundness) are beyond the scope of this paper. In this regard, it is similar to works like that on ΠΣ, which explore a novel language design without conducting any substantial metatheoretic investigations [1]. Indeed, tools as important for research in dependent type theory as Coq and Agda lack (in their current as opposed to historic forms) a complete metatheory, and soundness bugs have been found in recent years in both tools [12, 25].\(^2\) The goal of achieving highly reliable proof assistants by stringent metatheoretic analysis, and even verification of their implementations, is one we strongly endorse. But to warrant the tremendous theoretical and engineering investment required to realize such a goal, it is necessary first to investigate language designs carefully, to ensure they are adequate for their applications in practice. It is in this exploratory spirit that we study our proposed language design from the perspective of expressivity and applicability, and defer its metatheoretic analysis to future work.

This article is an extension of an earlier paper [14]. The organization of the paper has changed slightly, and the Sep\(^3\) language definition has been mildly extended. More significant changes include:

- Section 6.1 describes a number of lightweight proof automation techniques we developed in the course of experimenting with Sep\(^3\). Sep\(^3\) is designed as a core language design, and as a result is rather verbose. The techniques described in this section demonstrate that – with relatively simple methods – we can eliminate some of the drudgery from programming with the core language.
- Section 7 describes a solution in Sep\(^3\) to a verification problem for an interpreter for SK combinators. This problem, taken from the 2012 VSTTE Verified Software competition is interesting because the central program in question (i.e., the interpreter) is non-terminating.\(^3\) Though the Sep\(^3\) solution was not submitted to the competition, in this section we contrast the Sep\(^3\) solution with published solutions of systems that did compete.

The remainder of this paper is structured as follows. Section 2 provides a brief overview of a Sep\(^3\), and outlines the key design principles we followed. In Sections 3 through 5 we detail the problems identified above and present their associated solutions. Section 6 gives a brief overview of our experiences using a prototype implementation of Sep\(^3\), identifies opportunities for proof automation, and discusses an example proof. Section 7 presents the larger case study of verification of properties of the SK-combinator interpreter. Section 8 compares Trellys with related work. We conclude and identify directions for future work in Section 9.

### 2 Language Overview

The Sep\(^3\) language is one core language that we are developing to explore the design space of dependently typed languages. This name is short for

\[ \Gamma \vdash p : P \quad \text{Proof } p \text{ shows proposition } P \]
\[ \Gamma \vdash P : \text{Formula} \quad P \text{ is a well-formed formula} \]
\[ \Gamma \vdash t : T \quad \text{Term } t \text{ has type } T \]
\[ \Gamma \vdash \text{val } t \quad \text{Term } t \text{ has a value} \]
\[ t_1 \rightsquigarrow t_2 \quad \text{Term } t_1 \text{ reduces to } t_2 \text{ in one step} \]
\[ t_1 \rightsquigarrow_* t_2 \quad \text{Term } t_1 \text{ reduces to } t_2 \text{ within } m \text{ steps} \]

\(^2\)The bug in Coq, despite being 4.5 years old, still persists in the most recent version of Coq, Coq 8.4.

\(^3\)See https://sites.google.com/site/vstte2012/compet

Author Stump was a co-organizer of the competition, and submitted the problem description upon which this section is based.
variables
w,x,y,z, f, q
natural numbers
m,n
expressions
e ::= p | t
classifiers
A ::= P | T
proof conversion context
ˆP ::= ... P ... | ¬p
program conversion context
ˆT ::= ... T ... | ¬p
typing contexts
Γ ::= ∅ | Γ, x:A | Γ, x:A\text{val}
substitution
σ ::= [e_1/x_1]e_2 | [e_1/x_1, ..., e_i/x_i]e_t

proofs
p ::= 
x | \(\lambda (x:A) \Rightarrow p \mid p \mid e\) abstraction and application
| join n equality axiom
| valα t value axiom
| termcase t \{y\} of abort \rightarrow p_1 | ! \rightarrow p_2 termination case
| ord p | ordtrans p_1 p_2 ordering axiom, transitivity
| ind f(x:T) [q]. p induction
| case t \{y\} p of C_1 \rightarrow p_1 ... C_n \rightarrow p_n case analysis
| pack x,p exists intro
| unpack p as (x,p) in p exists elim
| conv p at ˆP conversion
| let x = p_1 in p_2 let-proof
| let x \{y\} = t in p let-prog
| contra p_1 | contraabort p_1 p_2 contradiction

propositions
P ::= 
| \forall (x:A). P | \exists (x:A). P quantification
| t_1 = t_2 | t_1 < t_2 equality, ordering
| t! termination

programs and types
t,T ::= 
x | \(\lambda (x:A) \rightarrow t \mid t \mid e\) abstraction and application
| (x:A) \rightarrow T dependent function type
| Type type of types
| rec f(x:T_1). t_2 recursion
| conv t at ˆT conversion
| \[x:A] \rightarrow t | t [e] implicit abs. and app.
| [x:A] \rightarrow T implicit function type
| case t \{y\} of C_1 \rightarrow t_1 ... C_n \rightarrow t_n case analysis
| T t_1 ... t_n | C datatype, data constructor
| abort T failure
| tcast t by p termination cast
| let x = p in t let-proof
| let x \{y\} = t_1 in t_2 let-prog
unannotated programs

\[ u, U ::= \]
\[ x \mid \lambda x \to u \mid u_1 u_2 \quad \text{abstraction and application} \]
\[ (x:A) \to U \quad \text{dependent function type} \]
\[ \text{Type} \quad \text{type of types} \]
\[ \text{rec } f(x). u \quad \text{recursion} \]
\[ T u_1 \ldots u_n \mid C \quad \text{datatype, data constructor} \]
\[ \text{case } u \{y\} \text{ of } C_1 \to u_1 \ldots C_n \to u_n \quad \text{case analysis} \]
\[ \text{abort} \quad \text{failure} \]
\[ \text{tcast } u \quad \text{termination cast} \]
\[ \text{let } x = u_1 \text{ in } u_2 \quad \text{local binding} \]

unannotated program values

\[ v ::= \]
\[ \lambda x \to u \quad \text{abstraction} \]
\[ (x:A) \to U \quad \text{dependent function type} \]
\[ \text{Type} \quad \text{type of types} \]
\[ \text{rec } f(x). u \quad \text{recursion} \]
\[ C v_1 \ldots v_n \quad \text{constructions} \]
\[ T v_1 \ldots v_n \quad \text{datatype} \]
\[ \text{tcast } u \quad \text{termination cast} \]

?! 3 Sep^3 unannotated syntax

?! 4 Erasing annotated programs to unannotated programs.
\[
\begin{align*}
\Gamma & \vdash x : A & \text{TVar} \\
\Gamma & \vdash \text{val } t & \text{TValAx} \\
\Gamma & \vdash \text{valax } t : ! & \text{TValAx} \\
\Gamma & \vdash \forall \, x : A \, P & \text{TLamProof} \\
\Gamma & \vdash e : A & \text{TAppProof} \\
\Gamma & \vdash p : [e/z] P & \text{TAppProof} \\
\Gamma & \vdash |t_1| \rightarrow^n t_1' & \text{TJoin} \\
\Gamma & \vdash |t_2| \rightarrow^n t_2' & \text{TJoin} \\
\Gamma & \vdash t_1 : T_1 & \text{TJoin} \\
\Gamma & \vdash t_2 : T_2 & \text{TJoin} \\
\Gamma & \vdash \text{join } m : t_1 = t_2 & \text{TJoin} \\
\Gamma, y : t = \text{abort } t' & \vdash p_a : P & \text{TermCase} \\
\Gamma & \vdash t = t' & \text{TermCase} \\
\Gamma & \vdash p : (\Gamma | \hat{P}) & \text{TConvProof} \\
\Gamma & \vdash p : (\Gamma | \hat{P}) & \text{TConvProof} \\
\Gamma & \vdash \text{ord } p : t_1 < t_2 & \text{TOrd} \\
\Gamma & \vdash \text{ordtrans } p_1, p_2 : t_1 < t_2 & \text{TOrdTrans} \\
\Gamma & \vdash x : T & \text{TPack} \\
\Gamma & \vdash \text{pack } x, p : \exists (x : T).P & \text{TPack} \\
\Gamma & \vdash p_1 : t_1 < t_2 & \text{TUnpack} \\
\Gamma & \vdash p_2 : t_2 < t_3 & \text{TUnpack} \\
\Gamma & \vdash p_3 : t_3 & \text{TUnpack} \\
\Gamma & \vdash p : T & \text{TUnpack} \\
\Gamma & \vdash x : T & \text{TUnpack} \\
\Gamma & \vdash \text{pack } x, p : \exists (x : T).P & \text{TUnpack} \\
\Gamma & \vdash p_1 : t_1 & \text{TUnpack} \\
\Gamma & \vdash p_2 : t_2 & \text{TUnpack} \\
\Gamma & \vdash p_3 : t_3 & \text{TUnpack} \\
\Gamma & \vdash p : T & \text{TUnpack} \\
\Gamma & \vdash p : \text{Type} & \text{TUnpack} \\
\Gamma & \vdash c_1 : (y : A_1) \rightarrow \text{Type} & \text{TUnpack} \\
\Gamma & \vdash c_2 : (x : A_2) \rightarrow \text{Type} & \text{TUnpack} \\
\Gamma, w : [x/y][z/x] A_2 & \text{val} & z = t' = c_2 w & \text{TPack} \\
\Gamma & \vdash \text{case } t' \{z\} p \text{ of } c_2 w & \text{TUnpack} \\
\Gamma & \vdash p : \text{Formula} & \text{TUnpack} \\
\Gamma & \vdash p_a : t = \text{abort } T & \text{TUnpack} \\
\Gamma & \vdash p_t : T & \text{TUnpack} \\
\Gamma & \vdash \text{contraabort } p_a p_t : P & \text{TUnpack} \\
\Gamma & \vdash p_a : t = \text{abort } T & \text{TUnpack} \\
\Gamma & \vdash p_t : T & \text{TUnpack} \\
\Gamma & \vdash \text{contraabort } p_a p_t : P & \text{TUnpack} \\
\Gamma & \vdash \text{let } x = p_1 \text{ in } p_2 : P & \text{TLetProofProof} \\
\Gamma, x : T, y : x = t & \vdash p : P & \text{TLetProofProg} \\
\Gamma & \vdash t : T & \text{TLetProofProg} \\
\Gamma & \vdash \text{let } x = \{y\} = t \text{ in } p : P & \text{TLetProofProg} \\
\end{align*}
\]
\[ x : A \in \Gamma \quad \Rightarrow \quad \Gamma \vdash x : A \quad \text{TVar} \]
\[ \Gamma \vdash x : \text{Tval} \vdash t : T' \quad \Rightarrow \quad \Gamma \vdash \lambda (x : T) \rightarrow t : (x : T) \rightarrow T' \quad \text{TLam} \]
\[ \Gamma \vdash t : (x : A) \rightarrow T \quad \Rightarrow \quad \Gamma \vdash e : A \quad \Rightarrow \quad \Gamma \vdash e : [e/x]T \quad \text{TAppImp} \]
\[ x \notin \text{FV}(t) \quad \Gamma, x : A \vdash t : T' \quad \Rightarrow \quad \Gamma \vdash \lambda (x : A) \rightarrow t : [x : A] \rightarrow T' \quad \text{TLamImp} \]
\[ \Gamma \vdash t : [x : A] \rightarrow T \quad \Rightarrow \quad \Gamma \vdash e : A \quad \Rightarrow \quad \Gamma \vdash t [e] : [e/x]T \quad \text{TAppImp} \]
\[ \Gamma, x : A \vdash t : \text{Type} \quad \Rightarrow \quad \Gamma \vdash \text{Tpi} \quad \text{TPi} \]
\[ \Gamma \vdash t : (\Gamma | \top) \quad \Rightarrow \quad \Gamma \vdash \text{conv } t \text{ at } \top : (\Gamma | \top) \quad \text{TConvProg} \]
\[ \Gamma \vdash t : \text{Type} \quad \Rightarrow \quad \Gamma \vdash \text{let } x = t \text{ in } t : T \quad \text{TLetProgProof} \]
\[ \Gamma \vdash t' : C' \quad \Rightarrow \quad \Gamma \vdash C' : (y : A_1) \rightarrow \text{Type} \quad \text{Type} \]

**For each branch**
\[ C_{i, w} \rightarrow t_i \]
\[ \Gamma \vdash C_i : (y : A_1) \rightarrow \lambda (x : A_2) \rightarrow C' \quad \Rightarrow \quad \Gamma \vdash \text{case } t' \{ z \}\text{ of } C_{i, w} \rightarrow t_i : t \quad \text{TCaseProg} \]
\[ \Gamma \vdash x : T_1, y : x = t_1 : T_2 \quad \Rightarrow \quad \Gamma \vdash let x \{ y \} = t_1 \text{ in } t_2 : T_2 \quad \text{TLetProgProg} \]

**Evaluation contexts**
\[ E ::= \square | E \text{ u } v E | \text{ case } E \text{ of } C_1 x_0 \ldots x_n \mapsto u_1 | \text{ let } x = E \text{ in } u \]

\[ E[u] \mapsto E[u'] \quad \text{CxtStep} \]
\[ E[\text{abort}] \mapsto \text{abort} \quad \text{EAbort} \]
\[ E[(\lambda x \rightarrow t) v] \mapsto E[v/x]T \quad \text{EBeta} \]
\[ E[\text{let } x = v \text{ in } u] \mapsto E[v/x]u \quad \text{Let} \]
\[ E[\text{rec } f (x). u] \mapsto E[\text{rec } f (x). u/f, v/x]u \quad \text{Rec} \]
\[ E[\text{tcast } v] \mapsto E[v] \quad \text{TCast} \]
\[ E[\text{case } C_j v_0 \ldots v_n | C_j x_0 \ldots x_n \mapsto u_j] \mapsto E[v_0/x_0, \ldots, v_n/x_n]u_j \quad \text{CaseTerm} \]

?16 Program typing rules

?17 CBV Operational Semantics for Programs
“separation of proof and program”, indicative of the syntactic separation in the language between proofs and programs (and similarly, between propositions and types). Proofs can mention programs without invoking them. We dub this capability “Freedom of Speech”. Conversely, programs can use proofs to help demonstrate to the type checker that invariants expressed using dependent types hold. A central feature design decision is that all proofs are computationally irrelevant: they are erased prior to reduction. We also allow programmers to mark certain parts of programs as compile-time. They will also be erased prior to reduction.

In the exposition and judgments that follow, we use a few conventions to make the syntactic distinction clear. The metavariable \( p \) refers to proofs, and \( P \) to propositions. Propositions classify proofs (i.e. \( p : P \)), in the sense that types classify programs in the programming language. The programming language, in contrast, has a collapsed syntax where programs and types are taken from a single syntactic category. We will use metavariables \( t \) and \( T \) for expressions when we wish informally to emphasize the roles of subject (\( t \)) and classifier (\( T \)), respectively. We will use the metavariable \( e \) to represent either proofs (\( p \)) or programs (\( t \)) and the metavariable \( A \) to represent either propositions (\( P \)) or program types (\( T \)). The metavariables \( x, y, f, \) and \( q \) range over both program and proof variables.

The syntax of the features of Sep\(^3\) that we discuss in this paper is summarized in Figure 1. Various judgments used in the paper are summarized in Figure 2. Typing rules for proofs are given in Figure 5, and for programs in Figure 6. Typing is defined for annotated programs (and proofs). We also define the syntax of unannotated programs in Figure 3. The language’s call-by-value (CBV) operational semantics is defined on unannotated programs in Figure 7. Annotated programs erase to unannotated ones (with similar names), using an eraser function defined in Figure 4. This function is also invoked in the TLamImp rule, to ensure that a compile-time input is not used in the erasure of the body of the compile-time abstraction \( \lambda [x:A] \rightarrow t \). In the rules, we assume all typing contexts are well-formed; that is, for any \( x : A \) in a context \( \Gamma \) the type \( A \) is kindable. In the proof fragment this requires the kinding judgment \( \Gamma \vdash P : \text{Formula} \) over propositions. We elide the definition.

As noted above, Sep\(^3\) is intended as a core language, so we are willing to require significant annotations in order to avoid complicating the design with complex machinery for type inference or term reconstruction. As the examples later in this paper make apparent, writing proofs and programs directly in the core language can be burdensome. However, we plan to eventually reduce this load with surface language features – such as proof tactics and integration with automated reasoning tools – that can automate the insertion of core language annotations.

**Annotations and Erasure** The use of annotations in Sep\(^3\) ensures typing is algorithmic. These annotations do not have computational content, as they are erased prior to reduction. For example, changing the type of a program requires a programmer to explicitly insert the cast (\( \text{conv} \)) into the program. But such casts are dropped by erasure. Erasure also drops type annotations on abstractions, and all proofs. This is because proofs in Sep\(^3\) are purely specification, and do not have any run-time behavior.

Sep\(^3\) provides a mechanism for a programmer to specify which programs should be preserved across the eraser, by allowing abstractions and applications to be annotated as compile-time or run-time. For example, type arguments to polymorphic functions generally do not contain computational content and are annotated as compile-time. In the concrete syntax of the language presented in this paper, compile-time abstractions are marked by wrapping the abstraction variable and type annotation with square brackets, as shown in the polymorphic identity function:

\[
\lambda [a:\text{Type}] \rightarrow \lambda (x:a) \rightarrow x
\]

Similarly, applications to compile-time arguments are marked with square brackets, as in the \( \text{Nat} \) type argument to the identity function:

\[
(\lambda [a:\text{Type}] \rightarrow \lambda (x:a) \rightarrow x) \ [\text{Nat}] \ Z
\]

Compile-time applications and abstractions are erased before executing a program and when proving programs equal with \( \text{join} \). Erasing compile-time elements allow proofs of equality to be constructed without reasoning about specification data. For example, two constructions of the empty vector \( \text{VNil} \) (defined in the datatype section below) that differ
only in the representations of the compile-time length parameter (e.g., Z vs. plus Z Z) can be proved equal without reasoning about equalities of addition since the length index is marked compile-time. This approach adapts ideas on erasure from several previous works [33, 17, 21].

Equality Formulas The Sep³ proof language includes a primitive formula representing the propositional equality of two programs, written \( t_1 = t_2 \). A proof of an equality is given by the expression \( \text{join} \ n \) where \( n \) is a meta-level natural number serving as an upper bound on the number of reductions steps the type checker will use when attempting to decide joinability of the respective programs. The typing rule for \( \text{join} \) is shown as T.Join in Figure 5. In this rule and throughout the sequel we denote \( \alpha \)-equivalence between programs \( t_1 \) and \( t_2 \) as \( t_1 \equiv t_2 \). T.Join makes use of the eraser function (Figure 4). If a program in an equality proved by \( \text{join} \) reduces to a normal form in fewer steps than the bound given as an argument to \( \text{join} \), then we compare that normal form (modulo \( \alpha \)-equivalence) with the program resulting from reducing the other side of the equality in a similar manner. Equalities can be proved between programs that are non-terminating, provided that those programs are joinable in a number of steps less than or equal to the given bounds.

Conversion Sep³ programs and proofs can utilize equalities to change the type of a given proof or program. The typing rule for conversion is shown as rules TConvProof and TConvProg in Figures 5 and 6, respectively.

Syntactically, a conversion has the form \( \text{conv} t \ \text{at} \ \hat{T} \) (for programs) or \( \text{conv} v \ \text{at} \ \hat{P} \) (for proofs). This form casts the type (or proposition) of the program (or proposition) to the left of the \text{at} to the type indicated by the conversion context to the right of the \text{at} keyword. A conversion context has the syntactic form for introducing local program bindings includes an additional variable representing a proof that equates the bound name with its definition. If the locally-bound name is captured in the type of the body of the let term (due to a dependent type), then it is necessary to cast the type of the body to remove the bound variable. The additional variable in the let binding provides the equation necessary to perform this cast.

Conversion in Sep³ is not automatically inferred, so must be supplied by the programmer. This is because conversion is based on equality between programs, which is undecidable in Sep³, since the (programming) language is not normalizing.

The let form for introducing local program bindings includes an additional variable representing a proof that equates the bound name with its definition. If the locally-bound name is captured in the type of the body of the let term (due to a dependent type), then it is necessary to cast the type of the body to remove the bound variable. The additional variable in the let binding provides the equation necessary to perform this cast.

Termination Reasoning Termination of programs is expressed with the termination formula, \( t! \), indicating a program \( t \) is total. Termination proofs can be constructed in two ways. First, programs which can be judged (see Section 3) to be values, can trivially be proved terminating using the valax (pronounced “value axiom”) form. This includes programs which are syntactic values. Furthermore, termination proofs can be introduced using a termination case proof construct, termcase, that case-splits non-constructively on the termination behavior of a program (see Section 5). Finally, with contraabort we can combine clashing proofs of termination and divergence for the same program to reason using contradiction.

Recursion and Induction The Sep³ programming language includes a rec form for defining general recursive functions. This construct does not constrain the arguments to recursive calls, potentially allowing diverging computation. The proof language, in contrast, provides an ind form for induction over
programs. The TInd typing rule for this form requires the argument to recursive calls to be strictly decreasing in size. Recursive invocations of the formula must provide a proof of this, written \( t' < t \), constructed using the ord form. A structural ordering statement of this form can only be proved between programs \( t_1 \) and \( t_2 \) when both are terminating and \( t_1 \) is an application of a constructor to arguments including \( t_2 \). This ensures that the value of \( t_1 \) is structurally larger than the value of \( t_2 \), so the induction is well founded. An ordtrans expresses transitivity of the structural ordering, thereby allowing ind to be used for course-of-values induction. The typing rules for ind, ord, and ordtrans are shown in Figure 5, and the TRec typing rule for rec is shown in Figure 6.

One could certainly imagine strengthening this ordering, but it is worth recalling that in the presence of higher types, it is already quite powerful. For example, terminating recursion based on a lexicographic combination of the natural number ordering with itself is subsumed by natural-number recursion at higher-type. So one can easily prove that Ackermann’s function, for example, is terminating using induction with our structural ordering over natural numbers. The proof is simply a nested natural-number induction, where the outer induction proves a quantified statement which is itself proved by an inner induction, in the cases of the outer induction. Similarly, mutual induction of multiple theorems can be encoded as a single induction yielding a conjunction of proofs of the constituent theorems.

**Data Types** Sep\(^3\) includes support for algebraic data types and indexed type families. These types are purely programmatic data, and may include diverging programs. We use the term “inductive datatype” to cover both.

An example non-indexed family is that of natural numbers, shown below. This definition is elaborated to the core language by introducing constants Nat : Type, \( Z : \text{Nat} \), and \( S : (x: \text{Nat}) \rightarrow \text{Nat} \).

```plaintext
data Nat : Type where  
  Z : Nat  
  S : (x: Nat) \rightarrow \text{Nat} 
```

When case splitting on a program that yields an element of an algebraic data type, in each branch we get a proof that equates the scrutinee with the pattern of the branch alternative. This proof is given a name that is supplied, in the case syntax, in braces following the scrutinee program. Moreover, when case-splitting in a proof, we are required to supply a proof that the scrutinee terminates. This is because, as just noted, (programmatic) datatypes may be inhabited by diverging programs, and hence we must know that the scrutinee does not diverge, in order to case split safely on its form. To illustrate the case typing rules, Figure 8 shows an example instantiation of TCaseProof for Nat.

The TContra rule allows us to reason using contradictions arising from equations between dissimilar constructor expressions. For example, given an proof of the \( S \ Z = Z \), we can use \( \text{contra} \) to prove any proposition.

Dependently typed languages typically include the ability to define indexed type constructors, where the index may vary in the result type of each constructor. Sep\(^3\) data types only support parameters to types, where the range of every constructor must the same. Indexed types can be simulated, however, by having constructors for the datatype accept proofs (as additional arguments) of equations that constrain the type constructor parameters appropriately. These proofs are then used to refine the type of the data constructor when case-splitting on values of the data type.

The encoding of the archetypal indexed data type of length-indexed vectors is shown below. This defines a polymorphic data type that carries its length as a type constructor parameter. The type of \( \text{VNil} \) requires a proof that the length \( n \) is equal to \( Z \). Similarly, the type of \( \text{VCons} \) constructor takes an argument \( m \) representing the length of the tail of the vector, as well as a proof that \( n \) is the successor of \( m \).

```plaintext
data Vec : [a: Type] [n: Nat] \rightarrow \text{Type where}  
  VNil : [q: n = Z] \rightarrow \text{Vec a n}  
  VCons : [m: Nat] \rightarrow [q: n = S m] \rightarrow  
        (x:a) \rightarrow (xs:\text{Vec a m}) \rightarrow \text{Vec a n} 
```

The elaboration of type constructors to the core language has a subtle interaction with erasure. The elaboration of Vec results in a constant Vec : (a : Type) \rightarrow (n: Nat) \rightarrow \text{Type}. The compile-time arguments \( [a: \text{Type}] \) and \( [n: \text{Nat}] \) become run-time arguments for the type constructor Vec. Without this restriction, one can construct examples violating type soundness, where we use a (provable) equation like Vec [bool] [n] = Vec [nat] [n] for unsoundly casting a vector of booleans to a vector of
Case instantiation for natural numbers

1. The operational semantics of the proof language defined for meta-theoretical study would use a call-by-value $\beta$-reduction rule. The soundness of the proof fragment would depend on termination of the program fragment, but allowing non-termination of the program fragment is an explicit goal of the Sep$^3$ language design.

2. The second possibility is to only instantiate theorems about known terminating programs, perhaps by requiring a syntactic value restriction on applications of theorems quantified over terminating programs. This restriction, while sound, reduces the expressiveness of the language, as there are many theorems that are true regardless of whether it is possible to re-
\[ x : A^{\text{val}} \in \Gamma \quad \frac{}{\Gamma \vdash \text{val} \ x} \quad \text{ValVar} \]
\[ \Gamma \vdash \text{val} \ (x : A) \rightarrow t \quad \text{ValLam} \]
\[ \forall i, \Gamma \vdash \text{val} \ t_i \quad \frac{}{\Gamma \vdash \text{val} \ c \ t_0 \ldots t_n} \quad \text{ValCons} \]
\[ \Gamma \vdash \text{val} \ t \text{cast} \ t \text{ by } p \quad \text{ValTCast} \]

? 9 Selected Value Judgment Rules

duce the programs those theorems range over to values. For example, \( \text{join} \ m : \text{plus} \ x \ N = x \) regardless of whether \( x \) terminates. If \( x \) diverges then both sides of the equality diverge, and are hence still equal.

For Sep\(^3\), we modify the definition of the set of programs which are values. In addition to a simple syntactic definition, we utilize a judgmental definition of value, allowing the context to be used when determining whether a term is a value. Later, we will identify a class of syntactic values to be used only by reduction. In the case of abstraction in the proof fragment (Figure 5, rule TLamProof), the variable is added to the context without a value annotation. Conversely, in the program fragment (Figure 6, rule TLamProg) the variable is added with an additional val annotation on the typing assumption.

The value judgment for programs (Figure 9) includes a rule, ValVar, which identifies a variable as a value if it occurs in the context with a val annotation. The value judgment includes a number of axioms for each syntactic value form. We also show a representative subset of syntactic forms that are axiomatically judged values; others not depicted include dependent products, recursive functions, and the classifier Type. Programs that can be judged to be values can be proved to be terminating using the \( \text{valax} \) construct (Figure 5). For example, we can prove \( \text{S} \ N \) terminates (represented by the proposition \( \text{p} : \text{S} \ N \) !) because \( \text{S} \ N \) is judgmentally a value. More interestingly, if we have \( x : \text{Nat}^{\text{val}} \) in the typing context we can prove \( \text{S} \ x ! \) using \( \text{valax} \).

The ValCons rule identifies a constructor application to arguments, each judged as values, to be judged as a value. We utilize this rule when typing case expressions in proofs scrutinizing program values. When case-splitting on a program value in a proof, it is necessary to supply a proof to the case expression that the scrutinee terminates, to ensure that divergence of programs does not leak into the proof language. We are guaranteed (by virtue of the termination proof) that the scrutinee will terminate, so we are also guaranteed that the normal form of the scrutinee will be a constructor applied to arguments that are values. Thus, when performing reasoning within a case branch, the pattern variables for the case branch will necessarily be instantiated with values. Consequently, when adding those variable to the context when type checking a case branch, we mark those variables as values, as shown in the TCaseProof rule in Figure 5.

4 Evaluation with Termination Cast

In Sep\(^3\), propositional equalities are proved using the \( \text{join} \) construct, which forms equalities by evaluating the erasure of each side of the equality to a normal form or a maximum number of steps and then comparing the resulting programs modulo \( \alpha \)-equality and dropping any tcast constructs. This means that \( \text{join} \) depends on call-by-value reduction of programs, where \( \beta \)-reduction can only be performed in a context where the function argument is a value. However, reasoning in proofs is often on open programs, where variables occurring in the propositional equalities range over programs which need not be values. Variables ranging over programs are not treated as values due to the freedom-of-speech principle. This principle was designed to allow proofs to quantify over programs, including those that diverge.

Unfortunately, quantifying over all programs, not just values, causes difficulty. Evaluating expressions which include free program variables introduced by quantifiers in proofs will result in a stuck term whenever such a variable occurs in the argument position of a \( \beta \)-redex. We want to reason about programs, but our most powerful tool (reasoning about equality under \( \beta \)-reduction), is restricted because the programs we wish to reason about might possibly diverge.

Consider a theorem that expresses that the application of a polymorphic identity function to any argument (including diverging arguments) can always be substituted with the argument itself.

The identity function and proof definitions are de-
fined in the syntax of the Sep⁴ tool, which allows top-level program declarations to be defined by combining a type annotation and a program definition.⁴ For example, the Program declaration below defines a name, id, and term \( \text{id} : [a:] \rightarrow (y:a) \rightarrow y \), with type \([a:] \rightarrow (y:a) \rightarrow a\). Similarly, \text{id_is_id} defines a proof with an associated quantified proposition.

Program \text{id} : [a:] \rightarrow a := x
Theorem \text{id_is_id} :
  forall (a: Type)(x:a). id [a] x = x :=
  join 100

This proof fails because the two programs are not joinable. The reduction sequence below shows the problem.

\[ id [a] x \quad \{ \text{by erasure} \} \]
\[ id x \quad \{ \text{by def. of id} \} \]
\[ (y.y) x \not\rightarrow \]

The β-redex \((y.y) x\) is blocked because the Sep³ programming language has a call-by-value semantics, and the variable \(x\) is not a value, as it was introduced by a proof. The equality proved by \text{join} is too fine, because it can only equate programs based on joinability using the call-by-value reduction semantics of the programming language.

Consider the above example if a call-by-name evaluation strategy were used when proving equalities using \text{join}. The blocked β-redex \((y.y) x\) would step, because it is not necessary to reduce the argument to a value prior to β-reduction. Making such a change would be unsafe, however, because it could allow us to observe different termination behaviors of the program, depending on whether the program is being reduced inside a proof (using call-by-name) or during actual execution (using call-by-value).

To illustrate, consider the term \((\lambda x: \text{Nat}.Z) \, \text{loop}\), where \text{loop} stands for any diverging computation. Inside a proof, using a call-by-name semantics we can prove \((\lambda x: \text{Nat}.Z) \, \text{loop} = Z\), while at run time using call-by-value the program would diverge.

On the other hand, if the argument \(t\) to an application is known to be terminating, because a proof \(p:t!\) is available, then the termination behavior of such an application will remain the same, regardless of whether it is reduced using call-by-name or call-by-value β-reduction.

Sep³ provides a termination cast construct, \text{tcast}, that allows a programmer to mark expressions as known to be terminating. The \text{tcast} construct takes a program and a proof that the program has a value. The TTCast typing rule for \text{tcast} is shown in Figure 6.

To allow reasoning over expressions including \text{tcast} constructs, the semantics of the programming language is augmented to allow an application with a \text{tcast} argument to step, despite the subject of the termination cast not necessarily being reduced to a value. In effect, \text{tcast} allows a programmer to posit a hypothetical value that the expression will reduce to, and then continue reduction based on that hypothesized value. This allows the language to prove more equalities than would be possible if \text{tcast} were not included.

Using \text{tcast}, a weaker form of the \text{id_is_id} theorem can be proved. The theorem is weakened to only hold for terminating arguments, by adding an additional parameter to the theorem that proves \(x\) is terminating. The proof is shown below.

Theorem \text{id_is_id_term} :
  forall (a: Type)(x:a)(x_term :x !).
  (id [a] x = x) :=
  join 100 :
  (id [a] (tcast x by x_term) = x)

In Section 3 we introduced a value judgment to differentiate between variables introduced by proof abstractions from those introduced by program abstractions. Using the value judgment, it is possible to redefine the β-rule to use the value judgment.

\[ \Gamma \vdash \text{val} \, u' \]
\[ \Gamma \vdash E[(\lambda x.u) \; u'] \equiv E[u'/x]u \quad \text{EBeta} \]

This rule allows reduction to reuse the value judgment, but it comes at the cost of complicating the reduction relation. No longer is reduction defined syntactically, but now it is a contextual relation. The alternative approach, and the one we take in Sep³, is to use the syntactic view of evaluation. We identify a syntactic class of values for reduction purposes that contains all of the syntactic forms which are trivially judged to be values by the value judgment, represented by the production \(v\) in Figure 3. Also included in this syntactic category are programs wrapped with termination casts. Variables are not classified as values, as before. If a term can be judged a value, then

---

⁴The type annotation is necessary because the typechecker implementation uses a bidirectional algorithm that combines type checking with type synthesis.
it is possible to extract a term in this class of syntactic values using a termination cast. For example, if a variable \( x \) is judged a value (but would not be syntactically classified a value), then \( \text{tcast} \ x \) by \( \text{valax} \ x \) is the associated syntactic value. In this way, we simplify the reduction relation to use standard call-by-value \( \beta \)-reduction over a class of syntactic values including \( \text{tcast} \) constructs, while still using the value judgment for typing.

When constructing a proof of equality using \( \text{join} \), the programs being equated are erased, as described previously. However, the \( \text{tcast} \) constructs are preserved (the termination proof is dropped, as it has no computational content). Additionally, reduction allows a \( \text{tcast} \) to be dropped when the expression being cast is a syntactic value. This prevents \( \text{tcast} \) from blocking a redex, as would be the case with \( \text{tcast} \ \lambda x.x \) by \( p \) \( v \).

Preserving termination casts during \( \text{join} \) reduction is necessary to allow \( \text{join} \) to construct equalities between programs involving expressions that do not normalize to values. After programs are reduced with \( \text{join} \), they are compared modulo termination casts and renaming of bound variables.

Preserving termination casts when reasoning about programs using \( \text{join} \) may cause a term to take more reduction steps inside proofs, because the EBeta reduction step may duplicate \( \text{tcast} \) constructs requiring a non-zero number of reduction steps to normalize to values. This need not introduce inefficiencies in compiled programs, because termination casts only occur in proofs, which are erased during compilation.

5 Termination Case

Sep² includes a \( \text{termcase} \) (Figure 5) construct that allows a proof to case-split on whether a scrutinized program terminates or diverges, with \( y \) bound as a proof of the corresponding termination assumption in each branch. In the terminates \( (! \) branch, \( y \) is a proof that the scrutinee terminates. In the diverges \( (\operatorname{abort}) \) branch, \( y \) is a proof that the scrutinee is equal to \( \operatorname{abort} \), signifying divergence. The typing rule TAbort, for \( \operatorname{abort} \) in the annotated language, requires an annotation \( t \) providing the type of the \( \operatorname{abort} \) term. Reduction for \( \operatorname{abort} \) is defined by the rule EAbort. If \( \operatorname{abort} \) appears in evaluation position, then the term immediately steps to \( \operatorname{abort} \). This means that all provably diverging programs (identified by equivalence to \( \operatorname{abort} \)) are contextually equivalent.

The \( \text{termcase} \) construct is non-constructive, as it axiomatizes excluded middle for termination, an undecidable property. This relies on an oracle that can determine whether any given term normalizes to a value. The construct can be viewed as an internalization of the theorem of type soundness for the program fragment, relaxed to partial correctness: if \( \vdash t : T \), then either \( t \leadsto^* v \) (where \( v \) is a syntactic value) and \( \vdash v : T \), or else \( t \) diverges. This has already been observed in Wright and Felleisen’s classic paper [36].

Using \( \text{termcase} \) we can strengthen proofs of algebraic theorems, which are subject to termination preconditions, to stronger theorems that do not require termination preconditions. As a simple example, we can return to the proof of \( \text{id_is_id} \) from Section 4. To prove the theorem above, it is necessary to have a proof of termination available to \( \text{tcast} \) the variable \( x \), so that the \( \text{join} \) proof can succeed. Nevertheless, the theorem is valid for all inputs, irrespective of termination behavior. Using \( \text{termcase} \), this theorem can be strengthened.

Theorem \( \text{id_is_id} \):

\[
\text{forall } (a:\text{Type}) (x:a). \text{id } a x = x :=
\text{termcase } x \{ x_{-\text{term}} \} \text{ of}
\begin{align*}
\text{abort} \rightarrow \\
\text{let } u_1 = \text{join } 100 \\
: \text{id } a (\text{abort } a) = (\text{abort } a)
\end{align*}
\]

This shows how we can do multiple conversion steps at one time, by including more than one splice, although in this simple example both splices refer to the proof \( x_{-\text{term}} : (\text{abort } a) = x \).

The proof does not require a separate lemma proving \( \text{id} \) to be a total function on terminating inputs. This capability – proving theorems about programs without proving those programs total – is an important proof technique enabled by \( \text{termcase} \). Although \( \text{id_is_id} \) amounts to such a proof for this particular example, in general \( \text{termcase} \) allows us to do algebraic reasoning without resorting to proving termi-
nation of the functions we are reasoning over.

6 Implementation and Experience

To gauge the feasibility of the Sep³ language design, we have implemented a prototype type checker and evaluator. This implementation has been useful in guiding the language design, and the design of Sep³ and the implementation have continued to evolve in tandem. The implementation is available online. The examples in this paper can be found in Tests/unittests/PiI.sep.

Sep³ is designed as a core language, and requires a large number of programmer annotations to get the type checker to accept a program. Requiring such a large number of annotations simplifies the design of the language and the implementation of the tools, but it complicates using the language, as the annotation burden is quite great. A second goal of the Trellys project is to design a surface language that allows many of the necessary core annotations to be synthesized by a program analysis algorithm.

6.1 Proof Automation

Sep³ programs rely heavily on the use of join to construct equations between programs, and the use of conv to change the type of a term or a proof. These tools are often irritatingly precise, as they provide a very low-level interface for proof. Quite often, constructing a proof in this language consists of the following.

1. Use join to prove a number of equations between programs.
2. Finesse the equations into just the right form.
3. Use conv, in concert with the generated equations, to cast the type of a term or the formula of a proof.

In a number of instances, we can provide simple automated support for common proof tasks. Below, we describe a collection of such automated tactics, starting with steps 2 and 3 from above, which are quite simple, and finally addressing the first step.

Theorem refl :
  forall (a:Type)(t:a). t = t :=
  join 0

Theorem sym :
  forall (a:Type)(b:Type)(t1:a)(t2:b)
  (p:t1 = t2).(t2 = t1) :=
  conv (refl a t1) at "p = t1"

Theorem trans :
  forall (a:Type)(b:Type)(c:Type)
  (t1:a)(t2:b)(t3:c)
  (p:t1 = t2)(q:t2 = t3).(t1 = t3) :=
  conv p at t1 = ~q

?10 Proofs that = is an equivalence.

which is considerably more complicated. The constructs for automated proof support will be used extensively in the examples in the remainder of this section and the next.

First, the primitive propositional formula judging two programs t1 and t2 to be joinable under call-by-value reduction, t1 = t2, is an equivalence relation. Moreover we can prove this within the language, as shown in Figure 10.

Undoubtedly, these theorems are simple enough to prove, but when it comes to using these lemmas, the story is quite different. The equality of Sep³ is heterogeneous, as the types of the two sides of an equation can be different. Consequently, in the proofs above we are required to take, as type parameters, the types of the equated programs. In the absence of general type inference support in Sep³, simply supplying those type arguments leads to very verbose programs. Because reasoning in Sep³ depends so heavily on manipulation of equality proofs, the prototype implementation provides primitive constructs trans and sym for transitivity and symmetry respectively that infer the necessary type arguments.

Even with primitives for reasoning about symmetry and transitivity, verification often requires a large number of tedious steps that amount to applications of these operators. As an additional automation capability, the prototype implementation provides an equiv form that eliminates the need to manually perform this reasoning. Operationally, equiv collects all of the (finite number of) equations present in the current typing context and takes the symmetric-transitive closure of them. Moreover, we associate with each equation in the closure the proof term built
from applications of sym and trans. When checking a proof term equiv against a particular formula \( t_1 = t_2 \), the typechecker simply looks for the expected formula in the transitive-symmetric closure and, if it is found, checks the associated proof term, or alternatively raises an error if the formula is not found.

Using equiv eliminates a great deal of the tedium of working with equations in proof. After proving just the right equalities, the next step is often to use those equations to cast the type of a term or a proof. And in Sep\(^3\), to do a cast requires the programmer to supply a conv term with a conversion context that identifies the proper places for those equations to be used. The Sep\(^3\) implementation provides an autoconv tactic that constructs a conversion context syntactically. The tactic takes a term \( t \) and an expected type \( \mathcal{S} \) – the intended type of the result of the cast – \( \mathcal{T} \), and calculates a term conv \( t \) at \( \mathcal{C} \), for some conversion context \( \mathcal{C} \). The autoconv tactic first synthesizes the actual type \( \mathcal{S} \) of \( t \), and then does a comparison between the actual type \( \mathcal{S} \) and the intended type \( \mathcal{T} \). We use the notation \( \mathcal{S}_{|p} \) to represent the subterm of \( \mathcal{S} \) occurring at position \( p \), for some subterm position \( p \) of \( \mathcal{S} \) (and similarly for \( \mathcal{T} \) and \( \mathcal{C} \)). If \( \mathcal{S}_{|p} \) and \( \mathcal{T}_{|p} \) differ, we check to see if there is an equation in the context \( \mathcal{h} : \mathcal{S}_{|p} = \mathcal{T}_{|p} \). If so, then we set \( \mathcal{C}_{|p} \) to be \( \neg \mathcal{h} \) (the escape of \( \mathcal{h} \)). If \( \mathcal{S}_{|p} \) and \( \mathcal{T}_{|p} \) are the same, then we set \( \mathcal{C}_{|p} \) to be \( \mathcal{S}_{|p} \). As with the equiv tactic, we do not need to trust the implementation of autoconv as the resutting conversion proof conv \( t \) at \( \mathcal{C} \) is subsequently checked using the core typing rules.

The equiv and autoconv tactics compact the second and third bookkeeping steps of proofs identified above, but they do nothing to address the issue of constructing initial set of equations using join. To illustrate the usability challenges with reasoning directly with join, consider the proof of the property and_commutes, capturing that the boolean \&\& function commutes. This is proved in Figure 11.

In each case split on \( x \) and \( y \), we get proofs x_eq and y_eq that respectively prove formulas \( p = x \) and \( p = y \) in each case branch, where \( p \) is the pattern (either True or False in this example) associated with the branch. Since the only tool we have to generate equalities between programs is normalization of open programs using join, the programmer is forced to prove a simpler equation using the patterns, and then insert a conversion using the proofs of equalities introduced by the case split. To simplify matters slightly, though, the False branch for \( x \) uses autoconv to generate the conversion context automatically.

In practice, using join to prove equations is even more troubling, as the desired reduction may depend on some intermediate term that we have an equation for, yet the intermediate term is not exposed in the top level formula. For example, suppose the formula to be proven is \( t_1 = t_2 \), and we have a proof \( p : e = \mathcal{C} \) in the context, where \( e \) is an arbitrary expression and \( \mathcal{C} \) a nullary constructor. For the sake of the example, assume that \( t_1 \) reduces to some intermediate term case \( e \{ e_{-eq} \} \) of \( \mathcal{C} \rightarrow t_2 \). Using join, we would first prove that \( t_1 \) equals case \( e \{ e_{-eq} \} \) of \( \mathcal{C} \rightarrow t_2 \), and then conv with the equation to get that this equals case \( \mathcal{C} \{ e_{-eq} \} \) of \( \mathcal{C} \rightarrow t_2 \). Finally, applying join again, we can get that the latter term equals \( t_2 \). The issue here is that since \( e \) is some arbitrary (and perhaps reducible) expression, we are required to fiddle with the bound on reduction steps passed to join to ensure that we don’t reduce \( e \) in the process of reducing \( t_1 \). For if we reduce \( e \) to some \( e’ \), our equation \( e = \mathcal{C} \) can no longer be applied.

The difficulty with using join in this way is that it is not aware of equations that are available in the typing context, so it forces the programmer to manually perform the task of rewriting intermediate programs using conv. A partial solution to this problem is to make the implementation of the reduction semantics aware of equations, and to perform the rewriting automatically. The Sep\(^3\) implementation provides a tactic, called morejoin to do precisely this. In our experience, this tactic has proved invaluable for constructing equations feasible for programs of any reasonable size. For example, in the proof morejoin_and_commutes in Figure 11, all of the cases are proven directly with the same invocation of morejoin.

The interface for morejoin is quite simple – it simply takes a list of proofs and the expected equation to be proven, and invokes the evaluator in the same way as join. On the other hand, the implementation is quite more involved, as it requires instrumenting
Program $(\&\&)(x:Bool)(y:Bool) \rightarrow Bool :=
    \begin{align*}
    &\text{case } x \{ x_{eq} \} \text{ of} \\
    &\quad \text{True } \rightarrow y \\
    &\quad \text{False } \rightarrow \text{False}
    \end{align*}

Theorem and_commutes :
for all $(x:Bool)(y:Bool)(x_{term}:x!)(y_{term}:y!)$. 
$(x \&\& y) = (y \&\& x) :=$
    \begin{align*}
    &\text{case } x \{ x_{eq} \} x_{term} \text{ of} \\
    &\quad \text{True } \rightarrow \\
    &\quad \quad \begin{align*}
    &\text{let } u_1 = \text{join } 100 : (\text{True } \&\& \text{True}) = (\text{True } \&\& \text{True}) \\
    &\quad \quad \quad \text{in } \text{conv } u_1 \text{ at } (^{x_{eq}} \&\& ^{y_{eq}}) = (^{y_{eq}} \&\& ^{x_{eq}})
    \end{align*} \\
    &\quad \quad \text{| False } \rightarrow \\
    &\quad \quad \begin{align*}
    &\text{let } u_1 = \text{join } 100 : (\text{True } \&\& \text{False}) = (\text{False } \&\& \text{True}) \\
    &\quad \quad \quad \text{in } \text{conv } u_1 \text{ at } (^{x_{eq}} \&\& ^{y_{eq}}) = (^{y_{eq}} \&\& ^{x_{eq}})
    \end{align*}
    \end{align*}

Theorem morejoin_and_commutes :
for all $(x:Bool)(y:Bool)(x_{term}:x!)(y_{term}:y!)$. 
$(x \&\& y) = (y \&\& x) :=$
    \begin{align*}
    &\text{case } x \{ x_{eq} \} x_{term} \text{ of} \\
    &\quad \text{True } \rightarrow \\
    &\quad \quad \begin{align*}
    &\text{let } u_1 = \text{join } 100 : (\text{False } \&\& \text{True}) = (\text{True } \&\& \text{False}) \\
    &\quad \quad \quad \text{in } \text{autoconv } u_1 \quad \text{-- conv } u_1 \text{ at } (^{x_{eq}} \&\& ^{y_{eq}}) = (^{y_{eq}} \&\& ^{x_{eq}})
    \end{align*} \\
    &\quad \quad \text{| False } \rightarrow \\
    &\quad \quad \begin{align*}
    &\text{let } u_1 = \text{join } 100 : (\text{False } \&\& \text{False}) = (\text{False } \&\& \text{False}) \\
    &\quad \quad \quad \text{in } \text{autoconv } u_1 \quad \text{-- conv } u_1 \text{ at } (^{x_{eq}} \&\& ^{y_{eq}}) = (^{y_{eq}} \&\& ^{x_{eq}})
    \end{align*}
    \end{align*}

Theorem morejoin_and_commutes :
for all $(x:Bool)(y:Bool)(x_{term}:x!)(y_{term}:y!)$. 
$(x \&\& y) = (y \&\& x) :=$
    \begin{align*}
    &\text{case } x \{ x_{eq} \} x_{term} \text{ of} \\
    &\quad \text{True } \rightarrow \\
    &\quad \quad \begin{align*}
    &\text{let } u_1 = \text{join } 100 : (\text{False } \&\& \text{True}) = (\text{True } \&\& \text{False}) \\
    &\quad \quad \quad \text{in } \text{autoconv } u_1 \quad \text{-- conv } u_1 \text{ at } (^{x_{eq}} \&\& ^{y_{eq}}) = (^{y_{eq}} \&\& ^{x_{eq}})
    \end{align*} \\
    &\quad \quad \text{| False } \rightarrow \\
    &\quad \quad \begin{align*}
    &\text{let } u_1 = \text{join } 100 : (\text{False } \&\& \text{False}) = (\text{False } \&\& \text{False}) \\
    &\quad \quad \quad \text{in } \text{autoconv } u_1 \quad \text{-- conv } u_1 \text{ at } (^{x_{eq}} \&\& ^{y_{eq}}) = (^{y_{eq}} \&\& ^{x_{eq}})
    \end{align*}
    \end{align*}
the evaluator to perform rewrites\textsuperscript{6}. The proofs supplied must either prove equations, which are treated as left-to-right oriented rewrite rules, or else termination proofs, which are used to automatically insert termination casts.

When reducing a term, if an application \( v_1 \ t_2 \) is encountered where \( v_1 \) is a value, and the list of proofs supplied to \texttt{morejoin} includes a proof of \( p : t_2 \), then a termination cast is inserted around \( t_2 \), and \( v_1 (t\texttt{cast} t_2) \) is a \( \beta \)-redex. Furthermore, if the term to be reduced is a case expression \texttt{case t1 of \{ \ldots \}}, and the list of proofs supplied to \texttt{morejoin} includes \( p : t_1 = t_2 \), then the term is rewritten to \texttt{case t2 of \{ \ldots \}}, and reduction proceeds.

Combining inserting termination casts and rewriting is quite useful, as oftentimes we will want to prove some formula \( f \ t_1 = t_2 \) where \( f \) does an immediate case-split on its argument, and we have proofs of \( t_1 \) and \( t_1 = v_1 \). Intuitively, these proofs imply that the value of \( t_1 \) is \( v_1 \), so we would like to reduce \( f \ t_1 \) and then reduce the case-split on \( v_1 \). But in general \( t_1 \) need not actually reduce to \( v_1 \) (or indeed any value), and so we need to insert a \texttt{tcast} around \( t_1 \) to enable the first reduction (of \( f \ t_1 \)). Inserting this \texttt{tcast} around \( t_1 \) prevents further reduction within \( t_1 \) (since \texttt{tcast E} is not included as a form of evaluation context in Figure 7). This means that \texttt{morejoin} will reduce \( f \ t_1 \) to a case-split on \texttt{tcast t1}, and then substitute \( v_1 \) for \( t_1 \). The \texttt{TCast} rule (of Figure 7) will then reduce this term to a case-split just on \( v_1 \), which can finally reduce using the \texttt{CaseTerm} rule.

Inserting termination casts and rewriting case scrutinees can be justified quite directly.

If \( t \rightsquigarrow \texttt{v1 t2} \), and \( p : t2! \), then \( \texttt{join m : t = v1 t2} \). If \( \texttt{join 0 : v1 t2 = v1 (tcast t2 by p)} \) using transitivity we can prove \( t = v1 (tcast t2 by p) \). Similarly, for case expressions, if \( t \rightsquigarrow \texttt{case t1 of \{ \ldots \}} \) and \( p : t_1 = t_2 \), then \( \texttt{conv (join m \at m : t = case \_p of \{ \ldots \}} \). If \( \texttt{case t2 of \{ \ldots \}} \) reduces further, we can produce a proof equating it with its contractum using \texttt{join}, and compose the proofs using \texttt{trans}.

The primitive \texttt{join} proof term provides a precise mechanism for controlling the number of steps used to reduce a term, which also serves as an upper bound on the number of reduction steps, ensuring that type checking \texttt{join} will terminate. On the other hand, it’s not clear how to count reduction steps when using rewriting, as a single rewrite may simulate some arbitrary (but finite) number of reduction steps. As a practical matter we simply set the upper bound of steps for \texttt{morejoin} to some arbitrary large constant; this has been sufficient in practice. On the other hand, the current implementation does not provide any accounting for number of rewrites. Given a poor choice of rewrite rules (for example, anything with a \texttt{refl} proof), rewriting may not terminate.

### 6.2 Example: Append is associative

Figures 12-15 show an example proof of associativity of append for lists using the prototype Sep\textsuperscript{3} implementation. The proofs liberally use \texttt{morejoin} to automatically insert conversions and termination casts when constructing equalities, rather than using the more verbose \texttt{join}.

These examples use additional notation from the Sep\textsuperscript{3} implementation to introduce top-level definitions. The \texttt{Recursive} form used by \texttt{append} elaborates to a \texttt{rec} construct directly. The body of the \texttt{rec} includes nested lambda-abstractions for each additional parameter to the \texttt{Recursive} definition.

The \texttt{Inductive} notation used with \texttt{append_term} is somewhat more complicated. In the example, we are performing on induction on the argument \( xs \). However, the type of \( xs \) refers to the argument \( a \). In the \texttt{Inductive} notation, the definition will elaborate to a core language term where all of the arguments preceding the inductive argument will be introduced by lambda abstractions. The inductive argument is indicated by being followed by a name, in braces, that represents a proof that the inductive argument terminates. This inductive argument (and the termination argument) will be introduced with a core language \texttt{ind} form. All following arguments will be bound by lambda-abstractions in the body. In the \texttt{append_term} example, the proof term will desugar to the following:

\[
\texttt{\lambda (xs:List a) \{xs_term\}.} \\
\texttt{\lambda (ys:List) \Rightarrow \texttt{\lambda (ys_term:ys) \Rightarrow \ldots} }
\]

The proof of associativity is by induction on the structure of the list argument. Because lists are programs, it is necessary to provide proofs that the ar-

---

\textsuperscript{6}In principle, the evaluator should also keep a trace of rewrites used, so that a proper proof term with the required \texttt{convs} can be reconstructed, although the current implementation simply trusts the instrumented evaluator.
data List : [a: Type] -> Type where
   Nil : List a
   | Cons : (x:a) -> (xs:List a) -> List a

Recursive append :
   [b: Type] (xs:List b) (ys:List b) -> List b :=
   case xs { xs_eq } of
      Nil -> ys
   | Cons x xs' -> Cons [b] x (append [b] xs' ys)

Inductive append_assoc_term :
   forall (a: Type) (xs : List a){ xs_term }
   append [a] xs ( append [a] ys zs ) =
   append [a] ( append [a] xs ys ) zs :=
   let term_xs_ys = append_term [a] xs xs_term ys ys_term ;
   term_ys_zs = append_term [a] ys ys_term zs zs_term 
   in case xs { xs_eq } xs_term of
      Nil ->
         let u1 = morejoin { sym xs_eq , ys_term , xs_term } : ys = append [a] xs ys ;
         u2 = morejoin { sym xs_eq , xs_term } :
                  append [a] xs ( tcast ( append [a] ys zs ) by term_ys_zs ) =
                  append [a] ys zs ;
         u3 = morejoin { sym xs_eq , xs_term , ys_term } :
                  ys = append [a] xs ys
      in conv u2 at append [a] xs ( append [a] ys zs ) = append [a] ~ u3 zs
   | Cons x xs' ->
      let unroll_app = morejoin { sym xs_eq ,xs_term , term_ys_zs } :
                     append [a] xs ( append [a] ys zs ) =
                     Cons [a] x ( append [a] xs' ( append [a] ys zs ));
      ih = append_assoc_term [a] xs' ( ord xs_eq ) ys ys_term zs zs_term ;
      u1 = conv unroll_app at
                     append [a] xs ( append [a] ys zs ) = Cons [a] x ~ih ;
      u2 = morejoin { sym xs_eq , xs_term , ys_term } :
                     append [a] xs ys = Cons [a] x ( append [a] xs' ys ) ;
      term_xs'_ys = append_term [a] xs' ( value xs' ) ys ys_term ;
      u3 = morejoin { sym xs_eq , xs_term , ys_term } :
                     ( append [a] ( append [a] xs ys ) zs ) =
                     ( append [a] ( Cons [a] x ( append [a] xs' ys ) )) zs ;
      u4 = morejoin { zs_term , ys_term , value x , term_xs'_ys } :
                     append [a] ( Cons [a] x ( append [a] xs' ys ) ) zs
                     = Cons [a] x ( append [a] ( append [a] xs' ys ) zs ) ;
      u5 = trans u3 u4 :
                     append [a] ( append [a] xs ys ) zs
                     = Cons [a] x ( append [a] ( append [a] xs' ys ) zs )
      in conv u1 at append [a] xs ( append [a] ys zs ) = ~( sym u5 )
Inductive append_term :
(append [a] xs ys)! :=
case xs (xs_term) of
| Nil -> let u1 = morejoin {sym xs_eq, xs_term, ys_term} : append [a] xs ys = ys
| Cons x xs' ->
| append [a] xs ys' ->
| Cons x xs' ys' : x < xs
| append_term [a] xs' ys' = Cons [a] x (append [a] xs' ys');
| x_term = value x : x !;
| unroll_app = morejoin {sym xs_eq, xs_term, ys_term}
| append [a] xs ys = Cons [a] x (append [a] xs' ys);
u1 = value (Cons [a] ~x_term~ih)
in conv u1 at "(sym unroll_app)" !

Proof that append terminates on terminating inputs

Theorem append_assoc :
append [a] xs (append [a] ys zs) =
append [a] (append [a] xs ys) zs :=
termcase xs (xs_term) of
| abort ->
| append [a] xs (append [a] (abort (List a)) ys zs) = (abort (List a));
| append [a] xs (append [a] ys (abort (List a))) zs;
u1 = trans aleft aright
in conv u1 at
| append [a] xs (append [a] ys zs) =
append [a] (append [a] xs ys) zs
| ! ! ->
termcase ys (ys_term) of
| abort ->
| append [a] ys (append [a] (abort (List a)) zs) =
append [a] (append [a] xs (abort (List a))) zs;
u1 = trans aleft aright
in conv u1 at
| append [a] ys (append [a] xs zs) =
append [a] (append [a] xs ys) zs
| ! ! ->
termcase zs (zs_term) of
| abort ->
| append [a] zs (append [a] (abort (List a))) zs =
a_x_y_term = append_term a xs xs_term ys ys_term;
| append [a] xs (append [a] y (abort (List a))) zs;
u1 = trans aleft aright
in conv u1 at
| append [a] zs (append [a] xs zs) =
append [a] (append [a] xs ys) zs
| ! ! ->
append_assoc_term [a] xs xs_term ys ys_term zs zs_term

Generalizing associativity of list append to non-terminating arguments.
arguments to append terminate. The weak form of the theorem is shown in Figure 13. The proof proceeds by induction on \( xs \), so it requires a proof that \( xs \) is terminating.

In the syntax of Sep\(^3\), the presence of curly braces and lack of type ascription for the parameter \( xs \_term \) indicates that the preceding argument \( xs \) is the induction variable. When applying \( \text{append\_assoc\_term} \) from outside the proof, the argument in \( xs \_term \) position should be a proof of \( xs \)!. However, when we appeal to the induction hypothesis for a subterm \( xs' \) of \( xs \) within the body of \( \text{append\_assoc\_term} \), we supply a proof that \( xs' < xs \), as shown in the \( \text{ih} \) binding in the proof.

Within the body of the proof, we prove equalities involving the variables \( ys \) and \( zs \). These variables are introduced by a proof abstraction, so they range over expressions. It is necessary to use termination casts, inserted by \text{morejoin} using the supplied termination proofs, to reduce programs involving these variables.

The proof uses an additional lemma (Figure 14) that proves \( \text{append} \) total on terminating inputs. This is a convenience, to simplify presentation, and could be avoided with additional reasoning using \text{termcase}. For more complex functions, where the proof of totality is not so straightforward, using \text{termcase} may be preferable.

Because the programs \( xs \), \( ys \), and \( zs \) are all used in strict positions on both sides of the equality, the formula can be strengthened to an equality over all programs producing lists, regardless of whether they terminate. Figure 15 shows the generalization of the proof of associativity of append to potentially non-terminating arguments.

The proof uses \text{termcase} to consider the termination behavior of each argument in turn. In each \text{abort} branch, the EAbort rule allows us to join an application of append to the diverging argument with abort, demonstrating that both sides of the associativity formula join with abort. In the final terminates branch, the context contains proofs \( xs \_term \), \( ys \_term \), and \( zs \_term \) that prove the associated arguments are terminating. With these proofs available, the weaker \text{append\_assoc\_term} lemma can be invoked.

7 Example: Combinators

The \text{append\_assoc} example of the previous section provides a small flavor of proving system properties within Sep\(^3\). While the proof does not rely on the \text{append} function terminating, it is easy to prove this inductively. Examples of programs that we \textit{cannot} prove total yet we wish to perform external verification upon abound.

Below, we describe the Sep\(^3\) solution to a problem posed as part of the 2012 VSTTE verified software competition. The problem involves proofs of properties of an interpreter for a SK combinator language. Because the SK calculus is Turing complete, any interpreter of terms over this language will necessarily be partial.

Following the VSTTE competition, a number of participants made their submitted solutions available. Two particular solutions [8, 27] were developed in verification environments based on normalizing programs. In the description of the Sep\(^3\) solution below, we use these solutions as points of comparison to the Sep\(^3\) approach, where we support external verification of non-terminating functions.

7.1 Problem definition

The description for the SK combinator problem is available from the VSTTE 2012 program verification competition website\(^7\). In the interests of making the description of the Sep\(^3\) solution self-contained, we reproduce parts of the problem as necessary.

The SK combinator calculus consists of a simple term language defined as follows:

\[
t := S | K | t \ t,
\]

where terms are made up of constants \( S \) and \( K \), along with left-associative application. Values \( v \) are a subset of the language defined by the grammar:

\[
v := K | S | K v | S v | S v v
\]

Reduction is call-by-value. Terms can be decomposed to an evaluation context, given by the grammar:

\[
C = \emptyset | C t | v C.
\]

The operation \( C[t] \) produces a term by substituting \( \emptyset \) with \( t \) in the evaluation context \( C \). It is defined by the following set of equations:

\(^7\)https://sites.google.com/site/vstte2012/compet
\[ \box[t] = t \]
\[ (\forall C)[t] = \forall C[t] \]
\[ (C \ t)[t'] = C[t'][t'] \]

Finally, we have a single-step reduction relation defined as follows:

\[ C[K t1 t2] \rightarrow C[t1] \]
\[ C[S t1 t2 t3] \rightarrow C[(t1 t2) (t1 t3)] \]

We call the transitive-reflexive closure \( \rightarrow^* \) of \( \rightarrow \) the reduction relation. Finally, for a term \( t \), if there does not exist a term \( t' \) such that \( t \rightarrow t' \), then we write \( t \not\rightarrow \).

The programming task is to define a function \( \text{reduction} \) that takes a term and returns a term \( t' \) such that \( t \rightarrow^* t' \) and \( t' \not\rightarrow \). If there does not exist such a \( t' \), then \( \text{reduction} \) diverges. The verification problem consists of three parts. In the first, we are required to prove \( \text{reduction} \) correctly implements its specification. In the second, we are to show that reduction on a term containing no \( S \) subterms always terminates. The third requires us to demonstrate a property of the reduction of the terms consisting of left-recursive nested applications of \( K \).

In the remainder of this section, we will focus on the solution to the programming task and the verification of the second problem, as these serve to demonstrate the ability to perform external verification over potentially non-terminating programs in Sep³.

### 7.2 Sep³ solution

We begin by defining a datatype for terms, corresponding to the term grammar.

```haskell
data Term : Type where
  S : Term
  | K : Term
  | App : Term -> Term -> Term
```

Moreover, we define a program \( \text{isValue} \), defined over terms. In normalizing dependent type theories, a common practice when defining a predicate is to define an inductive proposition representing the predicate. In Sep¹, we cannot do this directly, as data types are exclusively programmatic, and may be inhabited by diverging programs. Therefore, a logical interpretation of an element of an inductive type is not valid. In Sep³, we define the \( \text{isValue} \) function directly as a recursive function over terms, returning a \( \text{Bool} \).

The definition of \( \text{isValue} \), along with a proof of termination (on terminating input) for \( \text{isValue} \), is given in Figure 16. The proof of \( \text{isValue}_\text{term} \) closely mirrors the structure of the program \( \text{isValue} \). This is to be expected, as the proof largely follows from equations constructed with \( \text{join} \). A more sophisticated surface language may be able to derive such termination proofs for a subclass of recursive functions using syntactic methods such as those employed in tools like Coq and Agda.

Similarly to the definition of \( \text{isValue} \), we define a recursive function \( \text{isRedex} \) (Figure 17) which returns a \( \text{Bool} \). This function is used to determine if a term is a reducible expression. As with \( \text{isValue} \), we can prove by induction that \( \text{isRedex} \) terminates on all terminating inputs, although we elide the proof.

The formulation of the small-step reduction relation in the problem description is expressed in terms of evaluation contexts. Rather than implement this directly, we instead define a function \( \text{step} \) only on redexes instead of all reducible terms. We then define a function \( \text{decompose} \) for identifying a redex and evaluation context of a term, as well as a function \( \text{plug} \) performing the substitution of a term for \( \square \) in an evaluation context.

The \( \text{step} \) function is defined only on redexes, so it takes a proof \( p: \text{isRedex} t = \text{True} \). The signature for \( \text{step} \) is

```haskell
Program step : (t:Term)[p:isRedex t = True] -> Term
```
Recursive value predicate

Recursive `isValue : (t:Term) -> Bool` :=

```
case t {t_eq} of
  K -> True
  |S -> True
  |App l r -> (case l {l_eq} of
    K -> isValue r
    |S -> isValue r
    |App l' r' -> (case l' {l'_eq} of
      S -> isValue r' && isValue r
      |K -> False
      |App a b -> False))
```

-- `isValue` terminates on terminating terms.
-- Note that this pretty much duplicates the code from `isValue`.

Inductive `isValue_term : forall (t:Term) {t_term}. isValue t ! :=

```
case t {t_eq} t_term of
  K -> let u1 = morejoin {t_term, sym t_eq}
       : True = isValue t
       in conv valax True at "u1" !
  |S -> let u1 = morejoin {t_term, sym t_eq}
       : True = isValue t
       in conv valax True at "u1" !
  |App l r ->
    case l {l_eq} (valax l) of
      K ->
        let u1 = morejoin {t_term, sym t_eq}
            : isValue r !
        u2 = morejoin {sym t_eq, sym l_eq, t_term, valax r}
            : isValue r = isValue t
        in conv u1 at "u2" !
      |S -> let u1 = isValue_term r (ord t_eq)
           : isValue r !
    u2 = morejoin {sym t_eq, sym l_eq, t_term, valax r}
        : isValue r = isValue t
        in conv u1 at "u2" !
    |App l' r' ->
      case l' {l'_eq} valax l' of
        S ->
          let ih_r = isValue_term r (ord t_eq : r < t); ih_r' = isValue_term r'
            (ordtrans (ord l_eq : r' < l) (ord t_eq : l < t));
          u1 = and_term (isValue r') (isValue r) ih_r' ih_r;
          u2 = morejoin {sym t_eq,sym l_eq,sym l'_eq, t_term, valax l, valax l'}
            : ((isValue r' && isValue r) = isValue t)
          in conv u1 at "u2" !
        |K ->
          let u1 = morejoin {sym t_eq,sym l_eq,sym l'_eq,t_term, valax l, valax l'}
            : False = isValue t
          in conv valax False at "u1" !
        |App a b ->
          let u1 = morejoin {sym t_eq,sym l_eq,sym l'_eq,t_term, valax l, valax l'}
            : False = isValue t
          in conv valax False at "u1" !
```

?!!16 `isValue`
Although the `step` function is parameterized over all terms, the precondition `isRedex t = True` identifies a subset of those terms which actually are redexes – that is, they have the forms

- `App (App K t1) t2`
- `App (App (App (S t1) t2) t3)`

Case-splitting on any term that is not of this form will yield a contradiction, as we will be able to prove that `isRedex t = False`. To prove this contradiction directly in the definition of `step` requires us to effectively inline the definition of `isRedex` in the body of `step`. Branches where `isRedex t = False` can be proved are unreachable, yet we would still have to return some result, for example `abort` or some designated `Term`. Whatever the choice, it may not be immediately obvious to a programmer reading the code that the case branch is unreachable.

Our solution to unreachable case branches draws from experience with normalizing type theories that include inductive propositions. We define a programmatic type `RedexProp t` representing the inductive proposition and then case splits on it. The associated `step_term` function shows that `step` is total, directly appealing to the `redexPropTerm` lemma rather than by inductively reasoning on the input term `t`.

```plaintext
Program step : (t:Term)[p:isRedex t = True] -> Term :=
    case redexProp t [p] {redex_eq} of
    RedexK t1 t2        -> App (App t1 t2) (App t1 t3)
    RedexS t1 t2 t3     -> App (App t1 t2) (App t1 t3)

Theorem step_term : forall (t:Term)(p:isRedex t = True). step t [p] !
```

The `step` function operates on redexes, but the small-step reduction relation is defined over reducible terms, including those containing a redex in a subterm. The `Ctx` type captures evaluation contexts and is defined by the following inductive data type.

```plaintext
data Ctx : Type where
    Box : Ctx
| C1 : Ctx -> Term -> Ctx
| C2 : (t:Term) -> [p:isValue t = True] -> Ctx -> Ctx
```

We relate terms to evaluation contexts by way of the `decompose` and `plug` functions (Figure 19). Since `decompose` must return a pair of `Ctx` and `Term`, we define a type `Decomp` to represent this pair.

The machinery for performing reduction is now in place – `step` relates a redex with its contractum. Using `decompose` separates a term into an evaluation context and possible redex. Composing `step` with `decompose`, followed by `plug` with the evaluation context and the contractum, gives us the small-step reduction relation `→`. Finally, recursively reducing the resulting term yields the transitive-reflexive closure of the small-step reduction relation. The `reduce` function, shown below, captures this process.

```plaintext
Recursive reduction :
    (t:Term) -> Term :=
    case decompose t {dec_t} of
    Dec c t' ->
```

Given a proof `p:isRedex t = True`, we can define a recursive program `redexProp` that constructs a term of type `RedexProp t`. Moreover, we can prove that `redexProp` is total on all inputs. In effect, the combination of the `redexProp` and `redexPropTerm` allow us to migrate programmatic data to the proof language. We can soundly case-split on a `RedexProp` resulting from a terminating application of `redexProp` in the proof language if we know that the `RedexProp` is a value.

A portion of the definitions of `redexProp` and `redexPropTerm` are shown in Figure 18. Note that in the definition of `redexProp`, we return `abort` when a pattern match leads to a contradiction. The associated termination proof makes this contradiction apparent, using the `contra` primitive.
Program redexProp : (t:Term)[p:isRedex t = True] -> RedexProp t :=
  case t {t_eq} of
    K -> abort (RedexProp t) -- Contradiction, since isRedex t = True
  | S -> abort (RedexProp t) -- Contradiction, since isRedex t = True
  | App f1 t1 ->
    case f1 {f1_eq} of
      K -> abort (RedexProp t) -- Contradiction, since isRedex t = True
    | S -> abort (RedexProp t) -- Contradiction, since isRedex t = True
    | App f2 t2 ->
      case f2 {f2_eq} of
        K -> let [u1] = (conv sym t_eq at t = App ~(sym f1_eq) t1)
            : t = App (App f2 t2) t1;
        [u2] = (conv u1 at t = App (App ~(sym f2_eq) t2) t1)
            : t = App (App K t2) t1;
        [u3] = morejoin {u2, valax t2, valax t1}
            : isRedex t = (isValue t1 && isValue t2);
        [u4] = trans (sym p) u3
            : True = (isValue t1 && isValue t2)
        in RedexK t t2 t1 [u2]
          [and_right (isValue t1) (isValue t2) (sym u4)]
          [and_left (isValue t1) (isValue t2) (sym u4)]
    | S -> abort (RedexProp t) -- Contradiction, as isRedex t = True
    | App f3 t3 -> ...

Theorem redexPropTerm :
  forall(t:Term)(p:isRedex t = True). redexProp t [p] ! :=
termscase t {t_term} of
  abort ->
    let isredex_t_aborts = aborts (isRedex "t_term"
        : ((abort Bool) = (isRedex t));
    isredex_t_terminates = (conv valax True at ~(sym p) !)
        : (isRedex t) !
    in contraabort isredex_t_aborts isredex_t_terminates
  | ! ->
    case t {t_eq} t_term of
      K -> let u1 = morejoin {t_term, sym t_eq} : False = isRedex t
          in contra (equiv 3 : False = True)
    | S -> ...
    | App f1 t1 -> ...

?! 18 Redex Inductive Proposition
data Decomp : Type where
Dec : (c: Ctx) -> (t: Term) -> Decomp

Recursive decompose:
(t: Term) -> Decomp :=
case isRedex t { redex_t } of
True -> Dec Box t |
False ->
case t { e_eq } of
K -> Dec Box t |
S -> Dec Box t |
App x y ->
case isValue x { x_val } of
True -> (case decompose y { y_eq } of
Dec c' t' -> Dec (C2 x [sym x_val] c') t') |
False -> (case decompose x { x_eq } of
Dec c' t' -> Dec (C1 c' y) t')

Recursive plug:
(c: Ctx)(t: Term) -> Term :=
case c { c_eq } of
Box -> t |
C1 c' t' -> App (plug c' t) t' |
C2 v [pf] c' -> App v (plug c' t)

?!19 Decomposition and Context Plugging

case isRedex t' { red_t' } of
True -> reduction (plug c (step t' [sym red_t'])) |
False -> plug c t'

The definition of the reduction function follows the definition of the reduction relation \( \rightarrow^* \) faithfully. The primary deviation is the use of the decompose function to identify an evaluation context and possible redex; a relationship between terms and evaluation contexts that is left implicit in the specification of the plug function. Indeed, as part of the development of reduction, we proved the connection between decompose and plug:

Inductive plug_decompose_inv :
forall (t: Term)(t_term : t!)(t': Term)(t'_term : t'!)(c: Ctx)(c_term)
(p: decompose t = Dec c t').
plug c t' = t

Discussion The reduction function is intentionally defined to match the specification of the reduction relation. However, this particular implementation decision has its drawbacks, particularly in efficiency. In each small-step reduction, a term is decomposed to an evaluation context and redex, the redex is contracted, and then the contractum plugged back into the evaluation context. In the transitive closure of this implementation of the small-step reduction relation, the reduction will repeatedly plug a contractum into an evaluation context and then immediately in the recursive call decompose the result, undoing the work of the plug function. A more efficient implementation may eliminate this overhead by continuing reduction immediately upon producing a new redex when plugging the contractum into the context.

Unfortunately, the resulting reduction function differs substantially from the specification of the reduction relation, because it interleaves the plug function with reduction, effectively defining a large-step reduction semantics. This in turn requires additional verification effort to prove that such a large-step interpreter simulates the small-step interpreter.

It is illustrative to compare the Sep3 definition to solutions in other verification environments, graciously posted online following the competition [8, 27]. We examine two top-scoring solutions implemented using ACL2 [13] and PVS [26]. Both ACL2 and PVS are language-based verification tools based on a logic of total functions. In particular, we see that the handling of non-termination is central to the definition of the reduction function.

The PVS solution defines a terminating reduce function implementing large-step reduction, which is later proven to simulate the small-step reduction relation. This function can be proved terminating in PVS by a measure function defined on the structure of the input term, much in the same way as the Sep3 _<_ formula is used in inductive proofs.

More interesting is the definition of the reduction function implementing the transitive reflexive closure of small-step reduction. First, iter_reduce – a bounded definition of reduction – is defined to take, in addition to a term to reduce, a maximum bound on the number of reduction steps required to yield a value. Next, reduction is defined by appealing to an oracle supplying an appropriate number of reduction steps.

iter_reduce(n)(R): RECURSIVE term =
( IF n = 0
  THEN R
  ELSE
    (LET Q = reduce(R)
      IN IF R = Q
        THEN R
        ELSE iter_reduce(n-1)(R)))
This solution uses a predicate subtype \( \text{reducible?} \) to ensure that \( \text{iter\_reduce} \) is only defined on normalizing terms – those for which there exists an appropriate finite number of reduction steps to yield a value, so \( \text{reduction} \) is not defined on all combinator terms.

Sep\(^3\) provides a similar oracle by virtue of the \text{termcase} construct, yet this is purely a proof language construct. Because proofs are erased from programs prior to execution, there is no need to implement such an oracle.

The ACL2 solution uses a similar approach, utilizing a \text{steps\_required} oracle to generate a bound \( n \) on the number of steps required to produce a value term. As commentary on the solution describes, this introduces an axiom that \text{steps\_required} always returns some value, but doesn’t indicate how to calculate that value. That is, some entity from ACL2’s untyped universe is returned, but it is not possible to reason about this entity in the ACL2 universe.

Next, the \( \text{reduction} \) function is defined. In contrast to the PVS solution, this definition is separated into two parts – a logical specification that uses the \text{steps\_required} oracle and an executable definition that simply recursively invokes the terminating small-step reduction function. The definition furthermore asserts that the specification and the implementation are equal. The separation between logical specification and executable implementation is similar to the proof/program separation in Sep\(^3\).

The comparison between the Sep\(^3\) implementation to the PVS and ACL2 solutions is intended to highlight the differences between language-based verification in a logic of total functions from Sep\(^3\), which internalizes the notion of termination within the logic. This allows the language to relax the constraint that all programs terminate. ACL2 and PVS are both mature logics and tools, and the above is not intended to be an exhaustive comparison with Sep\(^3\).

### 7.3 Reasoning about reduction

One of the verification exercises for the combinator problem involves proving termination of combinator terms which do not contain any \( S \) subterms. Intuitively, this is because any redexes in such a term will always produce a smaller term.

The \text{s\_free} predicate is defined recursively on the structure of the input term \( t \).

#### Recursive \text{s\_free}:

\[
(\text{defchoose \ steps\_required \ (n) \ (x)} \quad (b* (((\text{mv terminates} \&) \ (\text{reduce\_n} \ n \ x)))) \quad \text{terminates})
\]

\[
(\text{defchoose \ steps\_required \ (n) \ (x)} \quad (b* (((\text{mv terminates} \&) \ (\text{reduce\_n} \ n \ x)))) \quad \text{terminates})
\]

Next, the \( \text{reduction} \) function is defined. In contrast to the PVS solution, this definition is separated into two parts – a logical specification that uses the \text{steps\_required} oracle and an executable definition that simply recursively invokes the terminating small-step reduction function. The definition furthermore asserts that the specification and the implementation are equal. The separation between logical specification and executable implementation is similar to the proof/program separation in Sep\(^3\).

The comparison between the Sep\(^3\) implementation to the PVS and ACL2 solutions is intended to highlight the differences between language-based verification in a logic of total functions from Sep\(^3\), which internalizes the notion of termination within the logic. This allows the language to relax the constraint that all programs terminate. ACL2 and PVS are both mature logics and tools, and the above is not intended to be an exhaustive comparison with Sep\(^3\).
Theorem \texttt{plug\_preserves\_ord} that expresses the decrease in size of the term using the \texttt{\_ < \_} formula.

\begin{verbatim}
Inductive plug_preserves_ord :
  forall (t1:Term)(t1_term:t1) (t2:Term)(t2_term:t2)
  (p:t1 < t2)(c:Ctx){c_term}.
plug c t1 < plug c t2 := <omitted>
\end{verbatim}

The remainder of the proof is directly reasoning using computation and application of the inductive hypothesis.

8 Related Work

It is surprising that relatively few works and systems are concerned with external reasoning for call-by-value, general-recursive higher-order functional programs. Indeed, we are aware of no prior theorem proving systems which exactly address this very natural problem! NuPRL might be the closest, since it supports external reasoning about higher-order functional programs with general recursion, but it appears that the semantics is lazy rather than call-by-value [7]. The logics of theorem provers like Coq and Isabelle require all functions to be terminating, and then need not (and do not) include a particular reduction strategy as part of their semantics [24, 34]. ACL2 also requires totality of functions [19]. As mentioned above, there are methods for defining and reasoning about general-recursive functions, but these require a non-trivial encoding, for example, using co-inductive data types, domain predicates, or domain theory [15, 16, 5, 18]. Systems or theories for direct reasoning about general-recursive functions seem to be less widely used or known. LTC supports explicit reasoning about totality, conversion, and typing for (untyped) PCF programs (for a recent work on LTC, see [4]). Equality is based on conversion, rather than reduction, and hence no reduction strategy is privileged in the axiomatization of the theory. VeriFun supports reasoning about general-recursive, possibly undefined functions [35]. The language of VeriFun does have call-by-value semantics and polymorphic types, but only first-order functions. Feferman’s System W is a logical theory intended for the formalization of mathematics [9, Chapter 13]. Its language for function definition uses a (generally non-computable) search operator in place of general recursion, and its theory, like LTC, is based on conversion rather than reduction. The CFML tool automatically extracts a formula from an OCaml program that can be used for verification in Coq. It’s used for external verification only, and not for a dependently-typed language [6].

On the semantic side, several recent works are concerned with axiomatizing fixed-point operators (including call-by-value ones) that arise in various categorical structures [11, 31]. These works are focused on foundations, and propose general axioms applicable to a wide range of specific structures, including models of call-by-value computation with effects. In contrast, our interest is in applied reasoning in a specific pure call-by-value functional language. A major technical difference is that the notion of equality we have adopted here is not extensional, and hence would not (it seems) be able to validate the axioms proposed by, say, Hasewaga and Kakutani, which are expressed as equalities between denotationally equivalent higher-order terms [11].

The Ynot system, based on Hoare Type Theory, is a generalization of Hoare Logic to higher-order functional programs with general recursion, state, and call-by-value semantics [23, 22]. Thus, Ynot provides an internal verification solution to the problem of interest in this paper, and indeed to the further difficult matter of reasoning about state. But, to our knowledge, Ynot is not intended for external reasoning about programs; rather, it uses a monad indexed by pre- and post-conditions on the imperative state in order to perform internal verification of programs. Previous work of Stump and co-authors on Guru has similar goals as Sep³, with a similar language design separating proofs and programs, and using termination casts with a judgmental notion of value [32, 33]. But in those works, quantifiers range only over values, rather than arbitrary programs; there is no construct for termination case; and the issue of call-by-value \( \beta \)-reduction is not addressed. Indeed, the Guru implementation unsoundly allows \( \beta \)-reduction with non-value arguments\(^8\).

Somewhat less closely related are works based on the Edinburgh Logical Framework (LF) (also known as the \( \lambda^H \) type theory) [10]. Systems like Beluga and Delphin add the ability to write functional programs operating over datatypes described in LF [29, 30]. Like the type theories mentioned above, these systems support terminating recursion over indexed

\(^8\)This issue with the Guru implementation, discovered in the course of the current research on Sep³, remains to be repaired.
9 Conclusion

Trellys is a research project investigating the design of a dependently typed programming language with call-by-value semantics and general recursion. Sep$^3$ is a core language design for Trellys, and occupies, to the best of our knowledge, a unique position in the language design space. Sep$^3$ supports internal and external verification, while not requiring a programmer to resort to indirect encodings to implement general recursive functions.

Sep$^3$ uses a syntactic distinction between the proof and programming languages to isolate non-termination in the programming language from the proof language. Despite the syntactic distinction proofs can mention possibly diverging programs without inheriting their divergence, a capability dubbed “Freedom of Speech”.

Reasoning about programs with general recursion in a dependently typed language requires a number of modifications to the logic to ensure soundness while maintaining expressiveness. Variables can range, depending on context, over values or expressions, so Sep$^3$ includes a value judgment to differentiate the two. Equality proofs are constructed by reducing open programs, so termination casts are added to allow the programming language to soundly extend call-by-value reduction over non syntactic values. Many theorems are valid regardless of the termination behavior of the programs the theorems quantify over, so a termination case expression allows us to express those theorems, and furthermore allows us to reason about possibly diverging programs without proving termination.

Trellys remains a work in progress, and Sep$^3$ represents one attempt at defining a core language to support the desired goal of combining dependent types and a call-by-value language including general recursion. While the principal language design includes the concepts presented here as well as many other features, much work remains, most importantly the analysis of the meta-theoretical properties of the language design. Sep$^3$ depends on a syntactic separation between the proof and programming fragments of the language. The Trellys team continues investigation into methods to remove this syntactic distinction, including an internalized type representing the proof/program classification of a term. This allows terms to safely be migrated from the proof language to the programming language, and vice-versa.

Acknowledgements This work was supported by the National Science Foundation (NSF grant 0910500).

References


321–332, New York, NY, USA, 2010. ACM.


