TLP **17** (2): 121–147, 2017. © Cambridge University Press 2016
doi:10.1017/S1471068416000168 First published online 1 July 2016

*Default rules for Curry**

SERGIO ANTOY
Computer Science Dept., Portland State University, Oregon, USA
(*e-mail:* antoy@cs.pdx.edu)

MICHAEL HANUS
Institut für Informatik, CAU Kiel, Kiel, Germany.
(*e-mail:* mh@informatik.uni-kiel.de)

submitted 31 January 2016; revised 13 April 2016; accepted 2 May 2016

Abstract

In functional logic programs, rules are applicable independently of textual order, i.e., any rule can potentially be used to evaluate an expression. This is similar to logic languages and contrary to functional languages, e.g., Haskell enforces a strict sequential interpretation of rules. However, in some situations it is convenient to express alternatives by means of compact default rules. Although default rules are often used in functional programs, the non-deterministic nature of functional logic programs does not allow to directly transfer this concept from functional to functional logic languages in a meaningful way. In this paper, we propose a new concept of default rules for Curry that supports a programming style similar to functional programming while preserving the core properties of functional logic programming, i.e., completeness, non-determinism, and logic-oriented use of functions. We discuss the basic concept and propose an implementation which exploits advanced features of functional logic languages.

KEYWORDS: functional logic programming, semantics, program transformation

1 Motivation

Functional logic languages combine the most important features of functional and logic programming in a single language (see (Antoy and Hanus 2010; Hanus 2013) for recent surveys). In particular, the functional logic language Curry (Hanus (ed.) 2016) conceptually extends Haskell with common features of logic programming, i.e., non-determinism, free variables, and constraint solving. Moreover, the amalgamated features of Curry support new programming techniques, like *deep* pattern matching through the use of *functional patterns*, i.e., evaluable functions at pattern positions (Antoy and Hanus 2005).

For example, suppose that we want to compute two elements x and y in a list l with the property that the distance between the two elements is n, i.e., in l there are

* This is an extended version of a paper presented at the international symposium on Practical Aspects of Declarative Languages (PADL 2016), invited as a rapid communication in TPLP. The authors acknowledge the assistance of the conference program chairs Marco Gavanelli and John Reppy.

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

$n-1$ elements between x and y. We will use this condition in the n-queens program discussed later. Of course, there may be many pairs of elements in a list satisfying the given condition ("++" denotes the concatenation of lists):

```
dist n (_++[x]++zs++[y]++_) | n == length zs + 1 = (x,y)
```

Defining functions by case distinction through pattern matching is a very useful feature. Functional patterns make this feature even more convenient. However, in functional logic languages, this feature is slightly more delicate because of the possibility of functional patterns, which typically stand for an infinite number of standard patterns, and because there is no textual order among the rules defining an operation. The variables in a functional pattern are bound like the variables in ordinary patterns.

As a simple example, consider an operation `isSet` intended to check whether a given list represents a set, i.e., does not contain duplicates. In Curry, we might think to implement it as follows:

```
isSet (_++[x]++_++[x]++_) = False
isSet _                   = True
```

The first rule uses a functional pattern: it returns `False` if the argument matches a list where two identical elements occur. The intent of the second rule is to return `True` if no identical elements occur in the argument. However, according to the semantics of Curry, which ensures completeness w.r.t. finding solutions or values, *all* rules are tried to evaluate an expression. Therefore, the second rule is always applicable to calls of `isSet` so that the expression `isSet [1,1]` will be evaluated to `False` *and* `True`.

The unintended application of the second rule can be avoided by the additional requirement that this rule should be applied only if no other rule is applicable. We call such a rule a *default rule* and mark it by adding the suffix `'default` to the function's name (in order to avoid a syntactic extension of the base language). Thus, if we define `isSet` with the rules

```
isSet (_++[x]++_++[x]++_) = False
isSet'default _           = True
```

then `isSet [1,1]` evaluates only to `False` and `isSet [0,1]` only to `True`.

In this paper, we propose a concept for default rules for Curry, define its precise semantics, and discuss implementation options. In the next section, we review the main concepts of functional logic programming and Curry. Our intended concept of default rules is informally introduced in Section 3. Some examples showing the convenience of default rules for programming are presented in Section 4. In order to avoid the introduction of a new semantics specific to default rules, we define the precise meaning of default rules by transforming them into already known concepts in Section 5. Options to implement default rules efficiently are discussed and evaluated in Section 6. Some benchmarking of alternative implementations of default rules are shown in Section 7 before we relate our proposal to other work and conclude.

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

2 Functional logic programming and Curry

Before presenting the concept and implementation of default rules in more detail, we briefly review those elements of functional logic languages and Curry that are necessary to understand the contents of this paper. More details can be found in recent surveys on functional logic programming (Antoy and Hanus 2010; Hanus 2013) and in the language report (Hanus (ed.) 2016).

Curry is a declarative multi-paradigm language combining in a seamless way features from functional, logic, and concurrent programming (concurrency is irrelevant as our work goes, hence it is ignored in this paper). The syntax of Curry is close to Haskell (Peyton Jones 2003), i.e., type variables and names of defined operations usually start with lowercase letters and the names of type and data constructors start with an uppercase letter. $\alpha \to \beta$ denotes the type of all functions mapping elements of type α into elements of type β (where β can also be a functional type, i.e., functional types are curried), and the application of an operation f to an argument e is denoted by juxtaposition ($f\,e$). In addition to Haskell, Curry allows *free* (*logic*) *variables* in conditions and right-hand sides of rules and expressions evaluated by an interpreter. Moreover, the patterns of a defining rule can be non-linear, i.e., they might contain multiple occurrences of some variable, which is an abbreviation for equalities between these occurrences.

Example 1

The following simple program shows the functional and logic features of Curry. It defines an operation "++" to concatenate two lists, which is identical to the Haskell encoding. The second operation, `dup`, returns some list element having at least two occurrences[1]

```
(++) :: [a] → [a] → [a]
[]     ++ ys = ys
(x:xs) ++ ys = x : (xs ++ ys)

dup :: [a] → a
dup xs | xs == _ ++ [x] ++ _ ++ [x] ++ _
       = x
       where x free
```

Operation applications can contain free variables. They are evaluated lazily where free variables as demanded arguments are non-deterministically instantiated. Hence, the condition of the rule defining `dup` is solved by instantiating `x` and the anonymous free variables "`_`". This evaluation method corresponds to narrowing (Slagle 1974; Reddy 1985), but Curry narrows with possibly non-most-general unifiers to ensure the optimality of computations (Antoy *et al.* 2000).

Note that `dup` is a *non-deterministic operation* since it might deliver more than one result for a given argument, e.g., the evaluation of `dup [1,2,2,1]` yields the values `1` and `2`. Non-deterministic operations, which are interpreted as mappings

[1] Note that Curry requires the explicit declaration of free variables, as `x` in the rule of `dup`, to ensure checkable redundancy.

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

from values into sets of values (González-Moreno *et al.* 1999), are an important feature of contemporary functional logic languages. Hence, there is also a predefined *choice* operation

```
x ? _  =  x
_ ? y  =  y
```

Thus, the expression "`0?1`" evaluates to `0` and `1` with the value non-deterministically chosen.

Some operations can be defined more easily and directly using *functional patterns* (Antoy and Hanus 2005). A functional pattern is a pattern occurring in an argument of the left-hand side of a rule containing defined operations (and not only data constructors and variables). Such a pattern abbreviates the set of all standard patterns to which the functional pattern can be evaluated (by narrowing). For instance, we can rewrite the definition of `dup` as

```
dup (_++[x]++_++[x]++_) = x
```

Functional patterns are a powerful feature to express arbitrary selections in tree structures, e.g., in XML documents (Hanus 2011). Details about their semantics and a constructive implementation of functional patterns by a demand-driven unification procedure can be found in (Antoy and Hanus 2005).

Set functions (Antoy and Hanus 2009) allow the encapsulation of non-deterministic computations in a strategy-independent manner. For each defined operation f, $f_{\mathscr{S}}$ denotes the corresponding set function. $f_{\mathscr{S}}$ encapsulates the non-determinism caused by evaluating f except for the non-determinism caused by evaluating the arguments to which f is applied. For instance, consider the operation `decOrInc` defined by

```
decOrInc x = (x-1) ? (x+1)
```

Then "`decOrInc`$_{\mathscr{S}}$ `3`" evaluates to (an abstract representation of) the set $\{2,4\}$, i.e., the non-determinism caused by `decOrInc` is encapsulated into a set. However, "`decOrInc`$_{\mathscr{S}}$ `(2?5)`" evaluates to two different sets $\{1,3\}$ and $\{4,6\}$ due to its non-deterministic argument, i.e., the non-determinism caused by the argument is not encapsulated. This property is desirable and essential to define and implement default rules by a transformational approach, as shown in Section 5. In the following section, we discuss default rules and their intended semantics.

3 Default rules: Concept and informal semantics

Default rules are often used in both functional and logic programming. In languages in which rules are applied in textual order, such as Haskell and Prolog, loosely speaking every rule is a default rule of all the preceding rules. For instance, the following standard Haskell function takes two lists and returns the list of corresponding pairs, where excess elements of a longer list are discarded:

```
zip (x:xs) (y:ys) = (x,y) : zip xs ys
zip _      _      = []
```

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

The second rule is applied only if the first rule is not applicable, i.e., if one of the argument lists is empty. We can avoid the consideration of rule orderings by replacing the second rule with rules for the patterns not matching the first rule:

```
zip (x:xs) (y:ys) = (x,y) : zip xs ys
zip (_:_)  []     = []
zip []     _      = []
```

In general, this coding is cumbersome since the number of additional rules increases if the patterns of the first rule are more complex (e.g., we need three additional rules for the operation `zip3` combining three lists). Moreover, this coding might be impossible in conjunction with some functional patterns, as in the first rule of `isSet` above. Some functional patterns conceptually denote an infinite set of standard patterns (e.g., `[x,x]`, `[x,_,x]`, `[_,x,_,x]`, ...) and the complement of this set is infinite too.

In Prolog, one often uses the "cut" operator to implement the behavior of default rules. For instance, `zip` can be defined as a Prolog predicate as follows:

```
zip([X|Xs],[Y|Ys],[(X,Y)|Zs]) :- !, zip(Xs,Ys,Zs).
zip(_,_,[]).
```

Although this definition behaves as intended for instantiated lists, the completeness of logic programming is destroyed by the cut operator. For instance, the goal `zip([],[],[])` is provable, but Prolog does not compute the answer {`Xs=[],Ys=[],Zs=[]`} for the goal `zip(Xs,Ys,Zs)`.

These examples show that neither the functional style nor the logic style of default rules is suitable for functional logic programming. The functional style, based on textual order, curtails non-determinism. The logic style, based on the *cut* operator, destroys the completeness of some computations. Thus, a new concept of default rules is required for functional logic programming if we want to keep the strong properties of the base language, in particular, a simple-to-use non-determinism and the completeness of logic-oriented evaluations. Before presenting the exact definition of default rules, we introduce them informally and discuss their intended semantics.

We intend to extend a "standard" operation definition by one default rule. Hence, an operation definition with a default rule has the following form ($\overline{o_k}$ denotes a sequence of objects $o_1 \ldots o_k$)[2]

$$
\begin{array}{l}
f\ \overline{t_k^1}\ \texttt{|}\ c_1\ \texttt{=}\ e_1 \\
\vdots \\
f\ \overline{t_k^n}\ \texttt{|}\ c_n\ \texttt{=}\ e_n \\
f\texttt{'default}\ \overline{t_k^{n+1}}\ \texttt{|}\ c_{n+1}\ \texttt{=}\ e_{n+1}
\end{array}
$$

We call the first n rules *standard rules* and the final rule the *default rule* of f. Informally, the default rule is applied only if no standard rule is applicable, where

[2] We consider only conditional rules since an unconditional rule can be regarded as a conditional rule with condition `True`.

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

a rule is applicable if the pattern matches and the condition is satisfied. Hence, an expression $e = f\,\overline{s_k}$, where $\overline{s_k}$ are expressions, is evaluated as follows:

(1) The arguments $\overline{s_k}$ are evaluated enough to determine whether a standard rule of f is applicable, i.e., whether there exists a standard rule whose left-hand side matches the evaluated e and the condition is satisfied (i.e., evaluable to `True`).
(2) If a standard rule is applicable, it is applied; otherwise the default rule is applied.
(3) If some argument is non-deterministic, the previous points apply independently for each non-deterministic choice of the combination of arguments. In particular, if an argument is a free variable, it is non-deterministically instantiated so that every potentially applicable rule can be used.

As usual in a non-strict language like Curry, arguments of an operation application are evaluated as they are demanded by the operation's pattern matching and condition. However, any non-determinism or failure during argument evaluation is not passed inside the condition evaluation. A precise definition of "inside" is in (Antoy and Hanus 2009, Definition 3). This behavior is quite similar to set functions to encapsulate internal non-determinism. Therefore, we will exploit set functions to implement default rules.

Before discussing the advantages and implementation of default rules, we explain and motivate the intended semantics of our proposal. First, it should be noted that this concept distinguishes non-determinism outside and inside a rule application. This difference is irrelevant in purely functional programming but essential in functional logic programming.

Example 2
Consider the operation `zip` defined with a default rule:

```
zip (x:xs) (y:ys) = (x,y) : zip xs ys
zip'default _ _   = []
```

Since the standard rule is applicable to `zip [1] [2]`, the default rule is ignored so that this expression is solely reduced to `(1,2):zip [] []`. Since the standard rule is not applicable to `zip [] []`, the default rule is applied and yields the value `[]`. Altogether, the only value of `zip [1] [2]` is `[(1,2)]`. However, if some argument has more than one value, we use the evaluation principle above for each combination. Thus, the call `zip ([1] ? []) [2]` yields the two values `[(1,2)]` and `[]`.

These considerations are even more relevant if the evaluation of the condition might be non-deterministic, as the following example shows:

Example 3
Consider an operation to look up values for keys in an association list

```
lookup key assoc   | assoc == (_ ++ [(key,val)] ++ _)
                   = Just val
                   where val free
lookup'default _ _ = Nothing
```

Note that the condition of the standard rule can be evaluated in various ways. In particular, it can be evaluated (non-deterministically) to `True` and `False` for a fixed

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

association list and key. Therefore, using if-then-else (or an `otherwise` branch as in Haskell) instead of the default rule might lead to unintended results.

If we evaluate `lookup 2 [(2,14),(3,17),(2,18)]`, the condition of the standard rule is satisfiable so that the default rule is ignored. Since the condition has the two solutions $\{\texttt{val} \mapsto 14\}$ and $\{\texttt{val} \mapsto 18\}$, we yield the values `Just 14` and `Just 18`. If we evaluate `lookup 2 [(3,17)]`, the condition of the standard rule is not satisfiable but the default rule is applicable so that we obtain the result `Nothing`.

On the other hand, non-deterministic arguments might trigger different rules to be applied. Consider the expression `lookup (2?3) [(3,17)]`. Since the non-determinism in the arguments leads to independent evaluations of the expressions `lookup 2 [(3,17)]` and `lookup 3 [(3,17)]`, we obtain the results `Nothing` and `Just 17`.

Similarly, free variables as arguments might lead to independent results since free variables are equivalent to non-deterministic values (Antoy and Hanus 2006). For instance, the expression `lookup 2 xs` yields the value `Just v` with the binding $\{\texttt{xs} \mapsto \texttt{(2,v):_}\}$, but also the value `Nothing` with the binding $\{\texttt{xs} \mapsto \texttt{[]}\}$ (as well as many other solutions).

The latter desirable property also has implications for the handling of failures occurring when arguments are evaluated. For instance, consider the expression `lookup 2 failed` (where `failed` is a predefined operation which always fails whenever it is evaluated). Because the evaluation of the condition of the standard rule demands the evaluation of `failed` and the subsequent failure comes from "outside" the condition, the entire expression evaluation fails instead of returning the value `Nothing`. This is motivated by the fact that we need the value of the association list in order to check the satisfiability of the condition and, thus, to decide the applicability of the standard rule, but this value is not available.

Example 4

To see why our design decision is reasonable, consider the following contrived definition of an operation that checks whether its argument is the unit value `()` (which is the only value of the unit type):

```
isUnit x | x == () = True
isUnit'default _   = False
```

In our proposal, the evaluation of "`isUnit failed`" fails. In an alternative design (like Prolog's if-then-else construct), one might skip any failure during condition checking and proceed with the next rule. In this case, we would return the value `False` for the expression `isUnit failed`. This is quite disturbing since the (deterministic!) operation `isUnit`, which has only one possible input value, could return two values: `True` for the call `isUnit ()` and `False` for the call `isUnit failed`. Moreover, if we call this operation with a free variable, like `isUnit x`, we obtain the single binding $\{\texttt{x} \mapsto \texttt{()}\}$ and value `True` (since free variables are never bound to failures). Thus, either our semantics would be incomplete for logic computations or we compute too many values. In order to get a consistent behavior, we require that failures of arguments demanded for condition checking lead to failures of evaluations.

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

4 Examples

To show the applicability and convenience of default rules for functional logic programming, we sketch a few more examples in this section.

Example 5

Default rules are important in combination with functional patterns, since functional patterns denote an infinite set of standard patterns which often has no finite complement. Consider again the operation `lookup` as introduced in Example 3. With functional patterns and default rules, this operation can be conveniently defined

```
lookup key (_ ++ [(key,val)] ++ _) = Just val
lookup'default _  _                = Nothing
```

Example 6

Functional patterns are also useful to check the deep structure of arguments. In this case, default rules are useful to express in an easy manner that the check is not successful. For instance, consider an operation that checks whether a string contains a float number (without an exponent but with an optional minus sign). With functional patterns and default rules, the definition of this predicate is easy

```
isFloat (("-" ? "") ++ n1 ++ "." ++ n2)
        | (all isDigit n1 && all isDigit n2) = True
isFloat'default _ = False
```

Example 7

In the classical n-queens puzzle, one must place n queens on a chess board so that no queen can attack another queen. This can be solved by computing some permutation of the list `[1..`n`]`, where the ith element denotes the row of the queen placed in column i, and check whether this permutation is a safe placement so that no queen can attack another in a diagonal. The latter property can easily be expressed with functional patterns and default rules where the non-default rule fails on a non-safe placement:

```
safeDiag (_++[x]++zs++[y]++_) | abs (x-y) == length zs + 1 = failed
safeDiag'default xs = xs
```

Hence, a solution can be obtained by computing a safe permutation:

```
queens n = safeDiag (permute [1..n])
```

This example shows that default rules are a convenient way to express negation-as-failure from logic programming.

Example 8

This programming pattern can also be applied to solve the map coloring problem. Our map consists of the states of the Pacific Northwest and a list of adjacent states:

```
data State = WA | OR | ID | BC

adjacent = [(WA,OR),(WA,ID),(WA,BC),(OR,ID),(ID,BC)]
```

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

Furthermore, we define the available colors and an operation that associates (non-deterministically) some color to a state:

```
data Color = Red | Green | Blue

color x = (x, Red ? Green ? Blue)
```

A map coloring can be computed by an operation `solve` that takes the information about potential colorings and adjacent states as arguments, i.e., we compute correct colorings by evaluating the initial expression

```
solve (map color [WA,OR,ID,BC]) adjacent
```

The operation `solve` fails on a coloring where two states have an identical color and are adjacent, otherwise it returns the coloring

```
solve (_++[(s1,c)]++_++[(s2,c)]++_) (_++[(s1,s2)]++_) = failed
solve'default cs _ = cs
```

Note that the compact definition of the standard rule of `solve` exploits the ordering in the definition of `adjacent`. For arbitrarily ordered adjacency lists, we have to extend the standard rule as follows:

```
solve (_++[(s1,c)]++_++[(s2,c)]++_) (_++[(s1,s2)?(s2,s1)]++_)
  = failed
```

5 Transformational semantics

In order to define a precise semantics of default rules, one could extend an existing logic foundation of functional logic programming (e.g., (González-Moreno *et al.* 1999)) to include a meaning of default rules. This approach has been partially done in López-Fraguas and Sánchez-Hernández (2004) but without considering the different sources of non-determinism (inside versus outside) which is important for our intended semantics, as discussed in Section 3. Fortunately, the semantic aspects of these issues have already been discussed in the context of encapsulated search (Antoy and Hanus 2009; Christiansen *et al.* 2013) so that we can put our proposal on these foundations. Hence, we do not develop a new logic foundation of functional logic programming with default rules, but we provide a transformational semantics, i.e., we specify the meaning of default rules by a transformation into existing constructs of functional logic programming.

We start the description of our transformational approach by explaining the translation of the default rule for `zip`. A default rule is applied only if no standard rule is applicable (because the rule's pattern does not match the argument or the rule's condition is not satisfiable). Hence, we translate a default rule into a regular rule by adding the condition that no other rule is applicable. For this purpose, we generate from the original standard rules a set of "test applicability only" rules where the right-hand side is replaced by a constant (here: the unit value "`()`"). Thus, the single standard rule of `zip` produces the following new rule:

```
zip'TEST (x:xs) (y:ys) = ()
```

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

Now, we have to add to the default rule the condition that `zip'TEST` is not applicable. Since we are interested in the failure of attempts to apply `zip'TEST` to the actual argument, we have to check that this application has no value. Furthermore, non-determinism and failures in the evaluation of actual arguments must be distinguished from similar outcomes caused by the evaluation of the condition.

All these requirements call for the encapsulation of a search for values of `zip'TEST` where "inside" and "outside" non-determinism are distinguished and handled differently. Fortunately, set functions (Antoy and Hanus 2009) (as sketched in Section 2) provide an appropriate solution to this problem. Since set functions have a strategy-independent denotational semantics (Christiansen *et al.* 2013), we will use them to specify and implement default rules. Using set functions, one could translate the default rule into

```
zip xs ys | isEmpty (zip'TEST𝒮 xs ys) = []
```

Hence, this rule can be applied only if all attempts to apply the standard rule fail. To complete our example, we add this translated default rule as a further alternative to the standard rule so that we obtain the transformed program

```
zip'TEST (x:xs) (y:ys) = ()

zip (x:xs) (y:ys) = (x,y) : zip xs ys
zip xs ys | isEmpty (zip'TEST𝒮 xs ys) = []
```

Thanks to the logic features of Curry, one can also use this definition to generate appropriate argument values for `zip`. For instance, if we evaluate the equation `zip xs ys == []` with the Curry implementation KiCS2 (Braßel *et al.* 2011), the search space is finite and computes, among others, the solution {`xs=[]`}.

Unfortunately, this scheme does not yield the best code to ensure optimal computations. To understand the potential problem, consider the following operation:

```
f 0 1 = 1
f _ 2 = 2
```

Intuitively, the best strategy to evaluate a call to `f` starts with a case distinction on the second argument, since its value determines which rule to apply. If the value is 1, and only in this case, the strategy checks the first argument, since its value determines whether to apply the first rule. A formal characterization of operations that allow this strategy (Antoy 1992) and a discussion of the strategy itself will be presented in Section 6.2. In this example, the pattern matching strategy is as follows:

(1) Evaluate the second argument (to head normal form).
(2) If its value is `2`, apply the second rule.
(3) If its value is `1`, evaluate the first argument and try to apply the first rule.
(4) Otherwise, no rule is applicable.

In particular, if `loop` denotes a non-terminating operation, the call `f loop 2` evaluates to `2`. This is in contrast to Haskell (Peyton Jones 2003) which performs pattern matching from left to right so that Haskell loops on this call. This strategy, which is optimal for the class of programs referred to as *inductively sequential* (Antoy 1992) for which it is intended, has been extended to functional logic computations (*needed*

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

narrowing (Antoy *et al.* 2000)) and to overlapping rules (Antoy 1997) in order to cover general functional logic programs.

Now consider the following default rule for `f`:

```
f'default _ x = x
```

If we apply our transformation scheme sketched above, we obtain the following Curry program:

```
f'TEST 0 1 = ()
f'TEST _ 2 = ()

f 0 1 = 1
f _ 2 = 2
f x y | isEmpty (f'TEST𝒻 x y) = y
```

As a result, the definition of `f` is no longer inductively sequential since the left-hand sides of the first and third rule overlap. Since there is no argument demanded by all rules of `f`, the rules could be applied independently. In fact, the Curry implementation KiCS2 (Braßel *et al.* 2011) loops on the call `f loop 2` (since it tries to evaluate the first argument in order to apply the first rule), whereas it yields the result `2` without the default rule.

To avoid this undesirable behavior when adding default rules, we could try to use the same strategy for the standard rules and the test in the default rule. This can be done by translating the original standard rules into an auxiliary operation and redefining the original operation into one that either applies the standard rules or the default rules. For our example, we transform the definition of `f` (with the default rule) into the following functions:

```
f'TEST 0 1 = ()
f'TEST _ 2 = ()

f'INIT 0 1 = 1
f'INIT _ 2 = 2

f'DFLT x y | isEmpty (f'TEST𝒻 x y) = y

f x y = f'INIT x y ? f'DFLT x y
```

Now, both `f'TEST` and `f'INIT` are inductively sequential so that the optimal needed narrowing strategy can be applied, and `f` simply denotes a choice (without an argument evaluation) between two expressions that are evaluated optimally. Observe that at most one of these expressions is reducible. As a result, the Curry implementation KiCS2 evaluates `f loop 2` to `2` and does not run into a loop.

The overall transformation of default rules can be described by the following scheme (its simplicity is advantageous to obtain a comprehensible definition of the semantics of default rules). The operation definition

$$
\begin{array}{l}
f\ \overline{t_k^1}\ \texttt{|}\ c_1\ \texttt{=}\ e_1 \\
\vdots \\
f\ \overline{t_k^n}\ \texttt{|}\ c_n\ \texttt{=}\ e_n \\
f\texttt{'default}\ \overline{t_k^{n+1}}\ \texttt{|}\ c_{n+1}\ \texttt{=}\ e_{n+1}
\end{array}
$$

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

is transformed into (where f'TEST, f'INIT, f'DFLT are new operation identifiers):

$$f\texttt{'TEST}\ \overline{t_k^1}\ \texttt{|}\ c_1\ \texttt{=}\ \texttt{()}$$
$$\vdots$$
$$f\texttt{'TEST}\ \overline{t_k^n}\ \texttt{|}\ c_n\ \texttt{=}\ \texttt{()}$$

$$f\texttt{'INIT}\ \overline{t_k^1}\ \texttt{|}\ c_1\ \texttt{=}\ e_1$$
$$\vdots$$
$$f\texttt{'INIT}\ \overline{t_k^n}\ \texttt{|}\ c_n\ \texttt{=}\ e_n$$

$$f\texttt{'DFLT}\ \overline{t_k^{n+1}}\ \texttt{|}\ \texttt{isEmpty}\ (f\texttt{'TEST}_{\mathscr{S}}\ \overline{t_k^{n+1}})\ \texttt{\&\&}\ c_{n+1}\ \texttt{=}\ e_{n+1}$$

$$f\ \overline{x_k}\ \texttt{=}\ f\texttt{'INIT}\ \overline{x_k}\ \texttt{?}\ f\texttt{'DFLT}\ \overline{x_k}$$

Note that the patterns and conditions of the original rules are not changed. Hence, this transformation is also compatible with other advanced features of Curry, like functional patterns, "as" patterns, non-linear patterns, local declarations, etc. Furthermore, if an efficient strategy exists for the original standard rules, the same strategy can be applied in the presence of default rules. This property can be formally stated as follows:

Proposition 1
Let $\mathscr{R}$ be a program without default rules, and $\mathscr{R}'$ be the same program except that default rules are added to some operations of $\mathscr{R}$. If $\mathscr{R}$ is overlapping inductively sequential, so is $\mathscr{R}'$.

Proof
Let f be an operation of $\mathscr{R}$. The only interesting case is when a default rule of f is in $\mathscr{R}'$. Operation f of $\mathscr{R}$ produces four different operations of $\mathscr{R}'$: f, f'DFLT, f'INIT, and f'TEST. The first two are overlapping inductively sequential since they are defined by a single rule. The last two are overlapping inductively sequential when f of $\mathscr{R}$ is overlapping inductively sequential since they have the same definitional tree as f modulo a renaming of symbols. □

The above proposition could be tightened a little when operation f is non-overlapping. In this case, three of the four operations produced by the transformation are non-overlapping as well. Proposition 1 is important for the efficiency of computations. In overlapping inductively sequential systems, needed redexes exist and can be easily and efficiently computed (Antoy 1997). If the original system has a strategy that reduces only needed redexes, the transformed system has a strategy that reduces only needed redexes. This ensures that optimal computations are preserved by the transformation regardless of non-determinism.

This result is in contrast to Haskell (or Prolog), where the concept of default rules is based on a sequential testing of rules, which might inhibit optimal evaluation and prevent or limit non-determinism. Hence, our concept of default rules is more powerful than existing concepts in functional or logic programming (see also Section 8).

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

We now relate values computed in the original system to those computed in the transformed system and vice versa. As expected, extending an operation with a default rule preserves the values computed without the default rule.

Proposition 2
Let $\mathscr{R}$ be a program without default rules, and $\mathscr{R}'$ be the same program except that default rules are added to some operations of $\mathscr{R}$. If e is an expression of $\mathscr{R}$ that evaluates to the value t w.r.t. $\mathscr{R}$, then e evaluates to t w.r.t. $\mathscr{R}'$.

Proof
Let $f\ \overline{t_k} \to u$ w.r.t. $\mathscr{R}$, for some expression u, a step of the evaluation of e. The only interesting case is when a default rule of f is in $\mathscr{R}'$. By the definitions of f and f`'INIT` in $\mathscr{R}'$, $f\ \overline{t_k} \to f$`'INIT` $\overline{t_k} \to u$ w.r.t. $\mathscr{R}'$. A trivial induction on the length of the evaluation of e completes the proof. □

The converse of Proposition 2 does not hold because $\mathscr{R}'$ typically computes more values than $\mathscr{R}$—that is the reason why there are default rules. The following statement relates values computed in $\mathscr{R}'$ to values computed in $\mathscr{R}$.

Proposition 3
Let $\mathscr{R}$ be a program without default rules, and $\mathscr{R}'$ be the same program except that default rules are added to some operations of $\mathscr{R}$. If e is an expression of $\mathscr{R}$ that evaluates to the value t w.r.t. $\mathscr{R}'$, then either e evaluates to t w.r.t. $\mathscr{R}$ or some default rule of $\mathscr{R}'$ is applied in $e \overset{*}{\to} t$ in $\mathscr{R}'$.

Proof
Let A denote an evaluation $e \overset{*}{\to} t$ in $\mathscr{R}'$ that never applies default rules. For any operation f of $\mathscr{R}$, the steps of A are of two kinds: (1) $f\ \overline{t_k} \to f$`'INIT` $\overline{t_k}$ or (2) f`'INIT` $\overline{t_k} \to t'$, for some expressions $\overline{t_k}$ and t'. If we remove from A the steps of kind (1) and replace f`'INIT` with f, we obtain an evaluation of e to t in $\mathscr{R}$. □

In Curry, by design, the textual order of the rules is irrelevant. A default rule is a constructive alternative to a certain kind of failure. For these reasons, a single default rule, as opposed to multiple default rules without any order, is conceptually simpler and adequate in practical situations. Nevertheless, a default rule of an operation f may invoke an auxiliary operation with multiple ordinary rules, thus, producing the same behavior of multiple default rules of f.

6 Implementation

The implementation of default rules for Curry based on the transformational approach is available as a preprocessor. The preprocessor is integrated into the compilation chain of the Curry systems PAKCS (Hanus *et al.* 2016) and KiCS2 (Braßel *et al.* 2011). In some future version of Curry, one could also add a specific syntax for default rules and transform them in the front end of the Curry system.

The transformation scheme shown in the previous section is mainly intended to specify the precise meaning of default rules (similarly to the specification of the

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

meaning of guards in Haskell (Peyton Jones 2003)). Although this transformation scheme leads to a reasonably efficient implementation, the actual implementation can be improved in various ways. In the following, we present two approaches to improve the implementation of default rules.

6.1 Avoiding duplicated condition checking

Our transformation scheme for default rules generates from a set of standard rules the auxiliary operations *f*`'TEST` and *f*`'INIT`. *f*`'TEST` is used in the condition of the translated default rule to check the applicability of a standard rule, whereas *f*`'INIT` actually applies a standard rule. Since both alternatives (standard rules or default rule) are eventually tried for application, the pattern matching and condition checking of some standard rule might be duplicated. For instance, if a standard rule is applicable to some call and the same call matches the pattern of the default rule, it might be tried twice: (1) the standard rule is applied by *f*`'INIT`, and (2) its pattern and condition is tested by *f*`'TEST` in order to test the (non-)emptiness of the set of all results. Although the amount of duplicated work is difficult to assess accurately due to Curry's lazy evaluation strategy (e.g., to check the non-emptiness in the condition of *f*`'DFLT`, it suffices to compute at most one element of the set), there is some risk for operationally complex conditions or patterns, e.g., functional patterns.

This kind of duplicated work can be avoided by a more sophisticated transformation scheme where the common parts of the definitions of *f*`'TEST` and *f*`'INIT` are joined into a single operation. This operation first tests the application of a standard rule and, in case of a successful test, returns a continuation to proceed with the corresponding rule. For instance, consider the rules for `zip` presented in Example 2. The operations `zip'TEST` and `zip'INIT` generated by our first transformation scheme can be joined into a single operation `zip'TESTC` by the following transformation:

```
zip'TESTC (x:xs) (y:ys) = \_ → (x,y) : zip xs ys

zip'DFLT _ _ = []

zip xs ys = let cs = zip'TESTC𝒫 xs ys
            in if isEmpty cs then zip'DFLT xs ys
                             else (chooseValue cs) ()
```

Now, the standard rule is translated into a rule for the new operation `zip'TESTC` where the rule's right-hand side is encapsulated into a lambda abstraction to avoid its immediate evaluation if this rule is applied. The actual implementation of `zip` first checks whether the set of all such lambda abstractions is empty. If this is the case, the standard rule is not applicable so that the default rule is applied. Otherwise, we continue with the right-hand sides of all applicable standard rules collected as lambda abstractions in the set `cs`.[3]

[3] The operation `chooseValue` non-deterministically chooses some value of the given set.

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

System:	PAKCS 1.14.0 (Hanus *et al.* 2016)				
Operation: Arguments:	`isSet` [1..1000]	`isSet` [1000,1..1000]	`lookup` 5001 [(1,.)..(5000,.)]	`lookup` 5000 [(1,.)..(5000,.)]	`queens` 6
Sect. 5:	7.04	4.78	3.56	3.54	0.23
Sect. 6.1:	2.27	2.28	1.81	3.57	0.23

Fig. 1. Performance comparison of different transformation schemes.

The general transformation scheme to obtain this behavior is defined as follows: An operation definition of the form

$$
\begin{array}{l}
f\ \overline{t_k^1}\ \texttt{|}\ c_1\ \texttt{=}\ e_1\\
\vdots\\
f\ \overline{t_k^n}\ \texttt{|}\ c_n\ \texttt{=}\ e_n\\
f\texttt{'default}\ \overline{t_k^{n+1}}\ \texttt{|}\ c_{n+1}\ \texttt{=}\ e_{n+1}
\end{array}
$$

is transformed into:

$$
\begin{array}{l}
f\texttt{'TESTC}\ \overline{t_k^1}\ \texttt{|}\ c_1\ \texttt{=}\ \texttt{\textbackslash_}\ \rightarrow\ e_1\\
\vdots\\
f\texttt{'TESTC}\ \overline{t_k^n}\ \texttt{|}\ c_n\ \texttt{=}\ \texttt{\textbackslash_}\ \rightarrow\ e_n\\
\\
f\texttt{'DFLT}\ \overline{t_k^{n+1}}\ \texttt{|}\ c_{n+1}\ \texttt{=}\ e_{n+1}\\
\\
f\ \overline{x_k}\ \texttt{=}\ \texttt{let cs =}\ f\texttt{'TESTC}_{\mathscr{S}}\ \overline{x_k}\\
\qquad\texttt{in if isEmpty cs then}\ f\texttt{'DFLT}\ \overline{x_k}\\
\qquad\qquad\qquad\texttt{else (chooseValue cs) ()}
\end{array}
$$

Obviously, this modified scheme avoids the potentially duplicated condition checking in standard rules, but it is more sophisticated since it requires the handling of sets of continuations. Depending on the implementation of set functions, this might be impossible if the values are operations. If the results computed by set functions are actually sets (and not multi-sets), this scheme cannot be applied since sets require an equality operation on elements in order to eliminate duplicated elements.

Fortunately, this scheme is applicable with PAKCS (Hanus *et al.* 2016), which computes multi-sets as results of set functions so that it does not require equality on elements. Thus, we compare the run times of both schemes for some of the operations shown above which contain complex applicability conditions (functional patterns). All benchmarks were executed on a Linux machine (Debian Jessie) with an Intel Core i7-4790 (3.60 Ghz) processor and 8 GB of memory. Figure 1 shows the run times (in seconds) to evaluate some operations with both schemes. These benchmarks indicate that the new scheme might yield a reasonable performance

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

gain, although this clearly depends on the particular example. A further alternative transformation scheme is discussed in the following section.

6.2 Transforming default rules into standard rules

In some situations, the behavior of a default rule can be provided by a set of standard rules. Almost universally, standard rules are more efficient. An example of this situation is provided with the operation `zip`. In Example 2, this operation is defined with a default rule. A definition using standard rules is shown at the beginning of Section 3. The input/output relations of the two definitions are identical. In this section, we introduce a few concepts to describe how to obtain, under sufficient conditions, a set of standard rules that behave as a default rule.

The programs considered in this section are constructor-based (O'Donnell 1977) (the extension to functional patterns is discussed later). Thus, there are disjoint sets of *operation* symbols, denoted by $f, g, \ldots$, and *constructor* symbols, denoted by $c, d, \ldots$ An *f-rooted pattern* is an expression of the form $f\,\overline{t_n}$ where f is an operation symbol of arity n, each t_i is an expression consisting of variables and/or constructor symbols only, and $f\,\overline{t_n}$ is *linear*, i.e., there are no repeated occurrences of some variable. A *pattern* is an f-rooted pattern for some operation f. A pattern is *ground* if it does not contain any variable. A *program rule* has the form $l = r$ where the left-hand side l is a pattern (the extension to conditional rules is discussed later). Given a redex t and a step $t \rightarrow u$, u is called a *contractum* (of t). Although Curry allows non-linear patterns for the convenience of the programmer, they are transformed into linear ones through a simple syntactic transformation.

In the following, we first consider a specific class of programs, called *inductively sequential*, where the rules of each operation can be organized in a *definitional tree* (Antoy 1992).

Definition 1 (*Definitional tree*)
The symbols *rule*, *exempt*, and *branch*, appearing below, are uninterpreted functions for classifying the nodes of a tree. A *partial definitional tree* with an f-rooted pattern p is either a rule node $rule(p = r)$, an exempt node $exempt(p)$, or a branch node $branch(p, x, \overline{\mathscr{T}_k})$, where x is a variable in p (also called the *inductive variable*), $\{c_1, \ldots, c_k\}$ is the set of all the constructors of the type of x, the substitution σ_i maps x to $c_i\,\overline{x_{a_i}}$ (where $\overline{x_{a_i}}$ are all fresh variables and a_i is the arity of c_i), and $\mathscr{T}_i$ is a partial definitional tree with pattern $\sigma_i(p)$ (for $i = 1, \ldots, k$). A *definitional tree* $\mathscr{T}$ of an operation f is a finite partial definitional tree with pattern $f\,\overline{x_n}$, where n is the arity of f and $\overline{x_n}$ are pairwise different variables, such that $\mathscr{T}$ contains all and only the rules defining f (up to variable renaming). In this case, we call f *inductively sequential*.

Definitional trees have a comprehensible graphical representation. For instance, the definitional tree of the operation "++" defined in Example 1 is shown in Figure 2. In this graphical representation, the pattern of each node is shown. The root node is a *branch* and its children are *rule* nodes. The inductive variable of the *branch* is the left operand of "++". Referring to Definition 1, σ_1 maps this variable to "[]" and σ_2

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

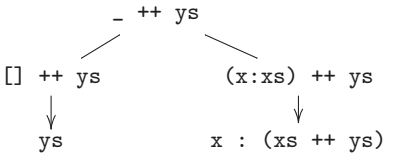


Fig. 2. A definitional tree of the operation "++".

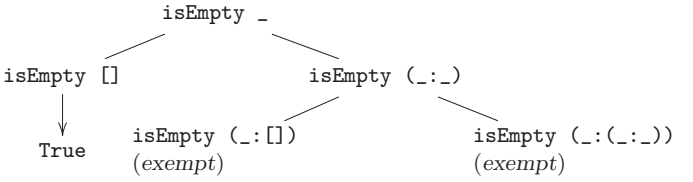


Fig. 3. A non-minimal definitional tree of the operation `isEmpty`.

to $(x : xs)$. For rule nodes, the right-hand side of the rule is shown below the arrow. Exempt nodes are marked by the keyword *exempt*, as shown in Figures 3 and 4.

For the sake of completeness, we sketch how definitional trees are used by the evaluation strategy. The details can be found in Antoy (1997). We discuss how to compute a rewrite of an expression rooted by an operation. More general cases are reduced to that. It can be shown that any step so computed is needed. Thus, let t be an expression rooted by an operation f and $\mathscr{T}$ a definitional tree of f. A traversal of $\mathscr{T}$ finds the deepest node $\mathscr{N}$ in $\mathscr{T}$ whose pattern p matches t. Such a node, and pattern, exist for every t. If $\mathscr{N}$ is a *rule* node, then t is a redex and is reduced. If $\mathscr{N}$ is an *exempt* node, then the computation is aborted because t has no value as in, e.g., `head []`, the head of an empty list. If $\mathscr{N}$ is a *branch* node, then the match of the inductive variable of p is an expression t' rooted by some operation and the strategy recursively seeks to compute a step of t'.

Before presenting our transformation, we state an important property of definitional trees.

Definition 2 (*Mutually exclusive and exhaustive patterns*)
Let f be an operation symbol and S a set of f-rooted patterns. We say that the patterns of S are *mutually exclusive* iff for any ground f-rooted pattern p, no two distinct patterns of S match p, and we say that the patterns of S are *exhaustive* iff for any ground f-rooted pattern p, there exists a pattern in S that matches p.

Lemma 1 (*Uniqueness*)
Let f be an operation defined by a set of standard rules. If $\mathscr{T}$ is a definitional tree of the rules of f, then the patterns in the leaves of $\mathscr{T}$ are exhaustive and mutually exclusive.

Proof
Let p be any ground f-rooted pattern, π the pattern in a node N of $\mathscr{T}$, and suppose that π matches p. Initially, we show that if N is not a leaf of $\mathscr{T}$, there is exactly one child N' of N such the pattern π' of N' matches p. Let x be the inductive variable of π and q the subexpression of p matched by x. Since p is ground and q is a

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

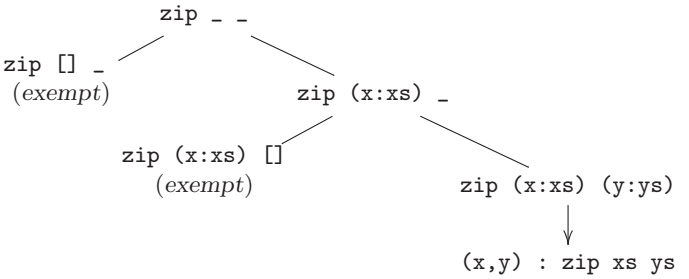


Fig. 4. A definitional tree of the standard rule of operation `zip` defined in Section 3.

proper subexpression of p, q is rooted by some constructor symbol c. Let $\{c_1, \ldots c_k\}$ be the set of all the constructors of the type defining c and let a_i be the arity of c_i, for all appropriate i. By Definition 1, N has k children with patterns $\sigma_i(\pi)$, where $\sigma_i = \{x \mapsto c_i\, \overline{x}_{a_i}\}$ and $\overline{x}_{a_i}$ is a fresh variable, for all appropriate i. Hence, exactly one of these patterns matches p since $c_i\, \overline{x}_{a_i}$ matches q iff $c_i = c$. Going back to the proposition's claim, since the pattern in the root of $\mathscr{T}$ matches p, by induction on the depth of $\mathscr{T}$, there is exactly one leaf whose pattern matches p. □

Inductive sequentiality is sufficient, but not necessary for a set of exhaustive and mutually exclusive patterns. We will later show a non-inductively sequential operation with exhaustive and mutually exclusive patterns. Nevertheless, inductive sequentiality supports a constructive method to transform default rules. Since not every definitional tree is useful to define our transformation, we first restrict the set of definitional trees.

Definition 3 (*Minimal definitional tree*)
A definitional tree is *minimal* iff there is some *rule* node below any *branch* node of the tree.

For example, consider the operation `isEmpty` defined by the single rule

```
isEmpty [] = True
```

Figure 3 shows a non-minimal tree of the rules defining `isEmpty`. The right child of the root is a *branch* node that has no *rule* node below it. In a minimal tree of the rules defining `isEmpty`, the right child would be an *exempt* node.

We now investigate sufficient conditions for the equivalence between an operation defined with a default rule and an operation defined by standard rules only.

Definition 4 (*Replacement of a default rule*)
Let f be an operation defined by a set of standard rules and a default rule $f\, \overline{x_k} = t$, where $\overline{x_k}$ are pairwise different variables and t some expression, and let $\mathscr{T}$ be a minimal definitional tree of the standard rules of f. Let $N_1, N_2, \ldots N_n$ be the exempt nodes of $\mathscr{T}$, $\overline{t_k^i}$ the pattern of node N_i and σ_i the substitution $\{\overline{x_k} \mapsto \overline{t_k^i}\}$, for $1 \leqslant i \leqslant n$. The following set of standard rules of f is called a *replacement* of the default rule of f:

$$\sigma_i(f\, \overline{x_k} = t), \qquad \text{for } 1 \leqslant i \leqslant n \tag{1}$$

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

Figure 4 shows a minimal definitional tree of the single standard rule of operation `zip` defined at the beginning of Section 3. The right-most leaf of this tree holds this rule. Since this leaf is below both branch nodes, the definitional tree is minimal according to Definition 3. The remaining two leaves hold the patterns that match all and only the combinations of arguments to which the default rule would be applicable. These patterns are more instantiated than that of the default rule, but we will see that any expression reduced by these rules does not need any additional evaluation with respect to the default rule.

Lemma 2 (*Correctness*)
Let f be an operation defined by a set S of standard rules and a default rule r of the form $f\ \overline{x_k} = t$, where each x_i is a variable, for all appropriate i, and some expression t, and let R be the replacement of r. For any ground f-rooted pattern p, p is reduced at the root to some q by the default rule r iff p is reduced at the root to q by some rule of R.

Proof
The proof is done in two steps. First, we prove that p is reduced by r iff p is reduced by some rule of R. Then, we prove that the contracta by the two rules are the same. By Lemma 1, the patterns in the rules of $S \cup R$ are exhaustive and mutually exclusive. Therefore, p is reduced by r if and only if p is not reduced by any rule of S if and only if p is reduced by some rule of R. We now prove the equality of the contracta. In the remainder of this proof, all the substitutions are restricted to $\overline{x_k}$, the argument variables of r. If p is reduced by r with some match σ, then $p = \sigma(f\ \overline{x_k})$ and $q = \sigma(t)$. Pattern p is also reduced by some rule of R which, by Definition 4, is of the form $\sigma_i(r)$, for some substitution σ_i. Consequently, $p = \sigma'(\sigma_i(f\ \overline{x_k}))$ for some match σ'. Since p is ground, $\sigma = \sigma' \circ \sigma_i$. Thus, the contractum of p by the rule of R is $\sigma'(\sigma_i(t)) = \sigma(t) = q$. □

In Definition 4, the replacement of a default rule is constructed for a minimal definitional tree. The hypothesis of minimality is not used in the proof of Lemma 2. The reason is that the lemma claims a property of f-rooted ground patterns. During the execution of a program, the default rule may be applied to some f-rooted expression e that may neither be a pattern nor ground. The hypothesis of minimality ensures that, in this case, no additional evaluation of e is required when a replacement rule is applied instead of the default rule. This fact is counter intuitive and non-trivial since the pattern of the default rule matches any f-rooted expression, whereas the patterns in the replacement rules do not, except in the degenerate case in which the set of standard rules is empty. However, a default rule is applicable only if no standard rule is applicable. Therefore, expression e must have been evaluated "enough" to determine that no standard rule is applicable. The following lemma shows that this evaluation is just right for the application of a replacement rule.

Lemma 3 (*Evaluation*)
Let e be an f-rooted expression reduced by the default rule of f according to the transformational semantics of Section 5. Let $\mathscr{T}$ be a minimal definitional tree of (the standard rules of) f. There exists an *exempt* node of $\mathscr{T}$ whose pattern matches e.

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

Proof
First note that the standard rules of f and the rules of f'TEST, as defined in Section 5, have identical left-hand sides. Hence, $\mathscr{T}$ is also a minimal definitional tree of the rules of f'TEST, which are used to check the applicability of the default rule.

To prove the claim, we construct a path $N_0, N_1, \ldots N_p$ in the definitional tree $\mathscr{T}$ of f with the following invariant properties: (a) the pattern π_i of each N_i unifies with e, and, (b) if the last node N_p is a leaf of $\mathscr{T}$, N_p is an *exempt* node. Establishing the invariant: N_0 is the root node of $\mathscr{T}$. By definition, its pattern π_0 is $f\ \overline{x_k}$, where $\overline{x_k}$ are fresh distinct variables. Hence, π_0 unifies with e so that invariant (a) holds. Furthermore, if N_0 is a leaf of $\mathscr{T}$, then it cannot be a *rule* node, otherwise e would never be reduced by a default rule. Hence, N_0 is an *exempt* node, i.e., invariant (b) holds. Maintaining the invariant: We assume that the invariant (a) holds for node N_k, for some $k \geqslant 0$. If N_k is a leaf of $\mathscr{T}$, then, as in the base case, N_k must be an *exempt* node. Hence, we assume N_k is a *branch* of $\mathscr{T}$ and show that invariant (a) can be extended to some child N_{k+1} of N_k. Since N_k is a *branch* node, e and π_k unify. For each child N' of N_k, let π' be the pattern of N', and let v be the inductive variable of the branch node N_k. By the definition of $\mathscr{T}$, $\pi' = \sigma'(\pi_k)$, where $\sigma' = \{v \mapsto c\ \overline{x_{a_c}}\}$, c is a constructor symbol of arity a_c, and x_i is a fresh variable for any appropriate i. Let σ be the match of π_k to e and $t = \sigma(v)$. If t is a variable, then any child of N_k satisfies invariant (a). Otherwise, t must be rooted by some constructor symbol, say d, for the following reasons. Because $\mathscr{T}$ is minimal, there are one or more rule nodes below N_k. The pattern in any of these rules is an instance of π_k that has some constructor symbol in the position matched by v. Hence, unless t is constructor-rooted, it would be impossible to tell which, if any, of these rules reduces e, hence it would be impossible to say whether e must be reduced by a standard rule or the default rule. Hence, N_{k+1} is the child in which v is mapped to $d\ \overline{x_{a_d}}$ so that invariant (a) also holds for N_{k+1}. □

We define the replacement of a default rule by a set of standard rules under four assumptions. We assess the significance of these assumptions below.

6.2.1 Inductive sequentiality

The standard rules are inductively sequential. This is a very mild requirement in practice. For instance, every operation of the Curry *Prelude*, except for the non-deterministic choice operator "?" shown in Section 2, is inductively sequential. Non-inductively sequential operations are problematic to evaluate efficiently. For example, the following operation, adapted from (Berry 1976, Proposition II.2.2), is defined by rules that do not admit a definitional tree

```
f False True  x     = ...
f x     False True  = ...
f True  x     False = ...
```

To apply f, the evaluation to constructor normal form of two out of the three arguments is both necessary and sufficient. No *practical* way is known to determine which these two arguments are without evaluating all three. Furthermore, since

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

the evaluation of an argument may not terminate, the three arguments must be evaluated concurrently (but see (Antoy and Middeldorp 1996)).

6.2.2 Most general pattern

We assumed in our transformation that the pattern of the default rule is most general, i.e., the arguments of the operation are all variables. Choosing the most general pattern keeps the statement of Lemma 3 simple and direct. With this assumption, no extra evaluation of the arguments is needed for the application of a replacement rule. To relax this assumption, we can modify Definition 4 as follows. If the left-hand side of the default rule is $f\,\overline{u_k}$, we look for a most general unifier, say σ_i, of $\overline{u_k}$ and $\overline{t_k^i}$. Then, rule $\sigma_i(f\,\overline{u_k} \to t)$ is in the replacement of the default rule iff such a σ_i exists.

6.2.3 Unconditional rules

Both standard rules and the default rule are unconditional. Adding a condition to the default rule is straightforward, similar to the transformation shown in Section 5. The condition of a default rule is directly transferred to each replacement rule by extending display (1) in Definition 4 with the condition. By contrast, conditions in standard rules require some care. With a modest loss of generality, assume that the standard rules have a definitional tree where each leaf node has a conditional rule of the form

$$f\,\overline{t_k} \mid c = t, \tag{2}$$

where c is a Boolean expression and t is any expression. Lemma 1 proves that if p is any f-rooted ground pattern matched by $f\,\overline{t_k}$ no other standard rule matches p. Hence, p is reduced at the root by the default rule of f iff c is not satisfied by p. Therefore, we need the following rule in the replacement of the default rule

$$f\,\overline{t_k} \mid \neg c = t, \tag{3}$$

where $\neg c$ denotes the "negation" of c, i.e., the condition satisfied by all the patterns matched by $f\,\overline{t_k}$ that do not satisfy c. In the spirit of functional logic programming, c is evaluated non-deterministically. For example, consider an operation that takes a list of colors, say `Red`, `Green`, and `Blue`, and removes all `Red` occurrences from the list:

```
data Color = Red | Green | Blue

remred cs | cs == x++[Red]++y
          = remred (x++y)
          where x,y free
remred'default cs = cs
```

The first rule is applied if there exist `x` and `y` that satisfy the condition. For example, for the list `[Red,Green,Red,Blue]` there are two such combinations of `x` and `y`. Thus, the "negation" of this condition must negate the existence of *any* such `x` and `y`. This can be automatically done according to the transformational semantics presented

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

in Section 5, but applied to a single rule. This example's replacement of the default rule is shown below

```
remred cs | isEmpty (remred'TEST℘ cs) = cs
remred'TEST cs | _++[Red]++_ == cs = ()
```

6.2.4 Constructor patterns

The standard rules defining an operation have constructor patterns. Curry also provides functional patterns, presented in Section 2. Rules defined by functional patterns can be transformed into ordinary rules (Antoy and Hanus 2005, Definition 4) by moving the functional pattern matching into the condition of a rule. Hence, the absence of functional patterns from our discussion is not an intrinsic limitation. Since functional patterns are quite expressive, operations defined with functional patterns often consist of a single program rule and a default rule (as in all examples shown in in Section 4). For instance, the previous operation `remred` can be defined with a functional pattern as follows:

```
remred (x++[Red]++y) = remred (x++y)
remred'default    cs = cs
```

Hence, the improved transformation scheme presented in Section 6.1 is still useful and should be applied in combination with the transformation shown in this section.

7 Benchmarking

To show the practical advantage of the transformation described in the previous section, we evaluated a few simple operations defined in a typical functional programming style with default rules. For instance, the Boolean conjunction can be defined with a default rule:

```
and True  True  = True
and'default _ _ = False
```

The replacement of the default rule consists of two rules so that the transformation yields the following standard rules:

```
and True  True  = True
and True  False = False
and False _     = False
```

Similarly, the computation of the last element of a list can be defined with a default rule:

```
last [x] = x
last'default (_:xs) = last xs
```

Our final example extracts all values in a list of optional (`Maybe`) values:

```
catMaybes []             = []
catMaybes (Just x : xs)  = x : catMaybes xs
catMaybes'default (_:xs) = catMaybes xs
```

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

System:	PAKCS 1.14.0 (Hanus et al. 2016)			
Operation / call size:	`zip` / 1000	`and` / 100000	`last` / 2000	`catMaybes` / 2000
Sect. 5:	3.66	8.46	2.53	2.45
Sect. 6.2:	0.01	0.25	0.01	0.01
System:	KiCS2 0.5.0 (Braßel et al. 2011)			
Operation / call size:	`zip` / 10^6	`and` / 10^6	`last` / 10^5	`catMaybes` / 10^6
Sect. 5:	2.72	1.35	0.38	0.40
Sect. 6.2:	0.04	0.08	0.01	0.01

Fig. 5. Performance comparison of different schemes for different compilers for some operations discussed in this section.

With the introduction of default rules, the order of evaluation may become more arbitrary, even though only needed steps are executed. For example, in the first definition of operation `and` both arguments must be evaluated, in any order, for the application of the standard rule. If the evaluation of one argument does not terminate and the other one evaluates to `False`, the order in which the two arguments are evaluated becomes observable. This situation is not directly related to the presence of a default rule. There are two "natural" inductive definitions of operation `and`, one evaluates the first argument first, as in the second definition of `and`, and another evaluates the second argument first. From the single standard rule of `and`, we cannot say which of the two definitions was intended. If the default rule of operation `and` is replaced by a set of standard rules, as per Section 6.2, the resulting definition, which is inductively sequential, will explicitly and arbitrarily encode which of the two arguments is to be evaluated first.

As discussed earlier, functional logic computations execute narrowing steps, i.e., steps in which some variable of an expression is instantiated and the rule reducing the expression depends on the instantiation of the variable. For example, consider again the `and` operation for its simplicity. The evaluation of `and x True`, where `x` is a free variable, narrows `x` to `True` to apply the standard rule and narrows `x` to `False` to apply the default rule. In a narrowing step, a variable is instantiated by the unification of the expression being evaluated and the left-hand side of a rule. This does not work with a default rule, since the arguments in the left-hand side are themselves variables. In particular, the transformational semantics of `and` has no rule to unify `x` with `False`. To obtain the intended behavior in narrowing steps variables are instantiated by *generators* (Antoy and Hanus 2006). In the example being discussed, the Boolean generator is `True ? False`.

Figure 5 shows the run times (in seconds) to evaluate the operations discussed in this section with the different transformation schemes (i.e., the scheme of Section 5 and the replacement of default rules presented in this section) and different Curry

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

implementations (where call size denotes the number of calls to `and` and the lengths of the input lists for the other examples). The benchmarks were executed on the same machine as the benchmarks in Section 6.1. The results clearly indicate the advantage of replacing default rules by standard rules, in particular for PAKCS, which has a less sophisticated implementation of set functions than KiCS2.

8 Related work

In this section, we compare our proposal of default rules for Curry with existing proposals for other rule-based languages.

The functional programming language Haskell (Peyton Jones 2003) has no explicit concept of default rules. Since Haskell applies the rules defining a function sequentially from top to bottom, it is a common practice in Haskell to write a "catch all" rule as a final rule to avoid writing several nearly identical rules (see example `zip` at the beginning of Section 3). Thus, our proposal for default rules increases the similarities between Curry and Haskell. However, our approach is more general, since it also supports logic-oriented computations, and it is more powerful, since it ensures optimal evaluation for inductively sequential standard rules, in contrast to Haskell (as shown in Section 5).

Since Haskell applies rules in a sequential manner, it is also possible to define more than one default rule for a function, e.g., where each rule has a different specificity. This cannot be directly expressed with our default rules where at most one default rule is allowed. However, one can obtain the same behavior by introducing a sequence of auxiliary operations where each operation has one default rule.

The logic programming language Prolog (Deransart *et al.* 1996) is based on backtracking where the rules defining a predicate are sequentially applied. Similarly to Haskell, one can also define "catch all" rules as the final rules of predicate definitions. In order to avoid the unintended application of these rules, one has to put "cut" operators in the preceding standard rules. As already discussed in Section 3, these cuts are only meaningful for instantiated arguments, otherwise the completeness of logic programming might be destroyed. Hence, this kind of default rules can be used only if the predicate is called in a particular mode, in contrast to our approach. The completeness for arbitrary modes might require the addition of concepts from Curry into Prolog, like the demand-driven instantiation of free variables.

Various encapsulation operators have been proposed for functional logic programs (Braßel *et al.* 2004) to encapsulate non-deterministic computations in some data structure. Set functions (Antoy and Hanus 2009) have been proposed as a strategy-independent notion of encapsulating non-determinism to deal with the interactions of laziness and encapsulation (see (Braßel *et al.* 2004) for details). One can also use set functions to distinguish successful and non-successful computations, similarly to negation-as-failure in logic programming, exploiting the possibility to check result sets for emptiness. When encapsulated computations are nested and performed lazily, it turns out that one has to track the encapsulation level in order to obtain intended results, as discussed in Christiansen *et al.* (2013). Thus, it is not surprising

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

that set functions and related operators fit quite well to our proposal. Actually, many explicit uses of set functions in functional logic programming to implement negation-as-failure can be implicitly and more tersely encoded with our concept of default rules, as shown in Examples 7 and 8.

Default rules and negation-as-failure have been also explored in López-Fraguas and Sánchez-Hernández (2004), Sánchez-Hernández (2006) for functional logic programs. In these works, an operator, `fails`, is introduced to check whether every reduction of an expression to a head-normal form is not successful. López-Fraguas and Sánchez-Hernández (2004) proposes the use of this operator to define default rules for functional logic programming. However, the authors propose a scheme where the default rule is applied if no standard rule was able to compute a head normal form. This is quite unusual and in contrast to functional programming (and our proposal) where default rules are applied if pattern matching and/or conditions of standard rules fail, but the computations of the rules' right-hand sides are not taken into account to decide whether a default rule should be applied. The same applies to an early proposal for default rules in an eager functional logic language (Moreno-Navarro 1994). Since the treatment of different sources of non-determinism and their interaction were not explored at that time, nested computations with failures are not considered by these works. As a consequence, the operator `fails` might yield unintended results if it is used in nested expressions. For instance, if we use `fails` instead of set functions to implement the operation `isUnit` defined in Example 4, the evaluation of `isUnit failed` yields the value `False` in contrast to our intended semantics.

Finally, we proposed in Antoy and Hanus (2014) to change Curry's rule selection strategy to a sequential one. However, it turned out that this change has drawbacks w.r.t. the evaluation strategy, since formerly optimal reductions are no longer possible in particular cases. For instance, consider the operation `f` defined in Section 5 and the call `f loop 2`. In a sequential rule selection strategy, one starts by testing whether the first rule is applicable. Since both arguments are demanded by this rule, one might evaluate them from left to right (as done in the implementation (Antoy and Hanus 2014)) so that this evaluation does not terminate. This problem is avoided with our proposal which returns `2` even in the presence of a default rule for `f`. Moreover, the examples presented in Antoy and Hanus (2014) can be expressed with default rules in a similar way.

9 Conclusions

We proposed a new concept of default rules for Curry. Default rules are available in many rule-based languages, but a sensible inclusion into a functional logic language is demanding. Therefore, we used advanced features for encapsulating search to define and implement default rules. Thanks to this approach, typical logic programming features, like non-determinism and evaluating operations with unknown arguments, are still applicable with our new semantics. This distinguishes our approach from similar concepts in logic programming which simply cut alternatives.

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

Our approach can lead to more elegant and comprehensible declarative programs, as shown by several examples in this paper. Moreover, many uses of negation-as-failure, which are often implemented in functional logic programs by complex applications of encapsulation operators, can easily be expressed with default rules.

Since encapsulated search is more costly than simple pattern matching, we have also shown some opportunities to implement default rules more efficiently. In particular, if the standard rules are inductively sequential and unconditional, one can replace the default rules by a set of standard rules so that the usage of encapsulated search can be completely avoided.

Acknowledgements

The authors are grateful to Sandra Dylus and the anonymous reviewers for their suggestions to improve a previous version of this paper. This material is based in part upon work supported by the National Science Foundation under Grant No. 1317249.

References

Antoy, S. 1992. Definitional trees. In *Proc. of the 3rd International Conference on Algebraic and Logic Programming*, Springer LNCS, Berlin Heidelberg, vol. 632, 143–157.

Antoy, S. 1997. Optimal non-deterministic functional logic computations. In *Proc. of 6th Int'l Conf. on Algebraic and Logic Programming (ALP'97)*, LNCS, vol. 1298. Springer, Southampton, UK, 16–30.

Antoy, S., Echahed, R. and Hanus, M. 2000. A needed narrowing strategy. *Journal of the ACM 47*, 4, 776–822.

Antoy, S. and Hanus, M. 2005. Declarative programming with function patterns. In *Proceedings of the International Symposium on Logic-based Program Synthesis and Transformation (LOPSTR'05)*, Springer LNCS, Berlin Heidelberg, vol. 3901, 6–22.

Antoy, S. and Hanus, M. 2006. Overlapping rules and logic variables in functional logic programs. In *Proc. of the 22nd International Conference on Logic Programming (ICLP 2006)*. Springer LNCS, Berlin Heidelberg, vol. 4079, 87–101.

Antoy, S. and Hanus, M. 2009. Set functions for functional logic programming. In *Proc. of the 11th ACM SIGPLAN International Conference on Principles and Practice of Declarative Programming (PPDP'09)*, ACM Press, 73–82.

Antoy, S. and Hanus, M. 2010. Functional logic programming. *Communications of the ACM 53*, 4, 74–85.

Antoy, S. and Hanus, M. 2014. Curry without success. In *Proc. of the 23rd International Workshop on Functional and (Constraint) Logic Programming (WFLP 2014)*. CEUR Workshop Proceedings, Aachen, vol. 1335. CEUR-WS.org, 140–154.

Antoy, S. and Middeldorp, A. 1996. A sequential strategy. *Theoretical Computer Science 165*, 75–95.

Berry, G. 1976. Bottom-up computation of recursive programs. *Informatique Théorique et Applications*, vol. 10, 47–82.

Brassel, B., Hanus, M. and Huch, F. 2004. Encapsulating non-determinism in functional logic computations. *Journal of Functional and Logic Programming 2004*, 6.

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press

Brassel, B., Hanus, M., Peemöller, B. and Reck, F. 2011. KiCS2: A new compiler from Curry to Haskell. In *Proc. of the 20th International Workshop on Functional and (Constraint) Logic Programming (WFLP 2011)*, Springer LNCS, Berlin Heidelberg, vol. 6816, 1–18.

Christiansen, J., Hanus, M., Reck, F. and Seidel, D. 2013. A semantics for weakly encapsulated search in functional logic programs. In *Proc. of the 15th International Symposium on Principle and Practice of Declarative Programming (PPDP'13)*, ACM Press, New York, 49–60.

Deransart, P., Ed-Dbali, A. and Cervoni, L. 1996. *Prolog - The Standard: Reference Manual*. Springer, Berlin Heidelberg.

González-Moreno, J., Hortalá-González, M., López-Fraguas, F. and Rodríguez-Artalejo, M. 1999. An approach to declarative programming based on a rewriting logic. *Journal of Logic Programming 40*, 47–87.

Hanus, M. 2011. Declarative processing of semistructured web data. In *Technical Communications of the 27th International Conference on Logic Programming*, Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Dagstuhl (Germany) vol. 11, 198–208.

Hanus, M. 2013. Functional logic programming: From theory to Curry. In *Programming Logics - Essays in Memory of Harald Ganzinger.*, Springer LNCS, Berlin Heidelberg, vol. 7797, 123–168.

Hanus, M. Ed. 2016. Curry: An integrated functional logic language (vers. 0.9.0). Available at `http://www.curry-language.org`.

Hanus, M., Antoy, S., Brassel, B., Engelke, M., Höppner, K., Koj, J., Niederau, P., Sadre, R. and Steiner, F. 2016. PAKCS: The Portland Aachen Kiel Curry System. Available at `http://www.informatik.uni-kiel.de/~pakcs/`.

López-Fraguas, F. and Sánchez-Hernández, J. 2004. A proof theoretic approach to failure in functional logic programming. *Theory and Practice of Logic Programming 4*, 1, 41–74.

Moreno-Navarro, J. 1994. Default rules: An extension of constructive negation for narrowing-based languages. In *Proc. 11th International Conference on Logic Programming*, MIT Press, Cambridge (USA) 535–549.

O'Donnell, M. J. 1977. *Computing in Systems Described by Equations*, LNCS, vol. 58, Springer, Berlin Heidelberg.

Peyton Jones, S. Ed. 2003. *Haskell 98 Language and Libraries—The Revised Report*. Cambridge University Press, Cambridge.

Reddy, U. 1985. Narrowing as the operational semantics of functional languages. In *Proc. IEEE Internat. Symposium on Logic Programming*, IEEE-CS, Boston, 138–151.

Sánchez-Hernández, J. 2006. Constructive failure in functional-logic programming: From theory to implementation. *Journal of Universal Computer Science 12*, 11, 1574–1593.

Slagle, J. 1974. Automated theorem-proving for theories with simplifiers, commutativity, and associativity. *Journal of the ACM 21*, 4, 622–642.

https://doi.org/10.1017/S1471068416000168 Published online by Cambridge University Press