A Constructor For
Applicative Multiprogramming*

Daniel P. Friedman
David S. Wise

Computer Science Department
Indiana University
Bloomington, Indiana 47405

TECHNICAL REPORT NO. 80
A CONSTRUCTOR FOR
APPLICATIVE MULTIPROGRAMMING

DANIEL P. FRIEDMAN
DAVID S. WISE
(REVISED) NOVEMBER, 1979

*Research reported herein was supported (in part) by the
National Science Foundation under grants numbered MCS75-06678 A01
and MCS77-22325.
A Constructor for Applicative Multiprogramming*

Daniel P. Friedman
David S. Wise

Computer Science Department
Indiana University
Bloomington, Indiana

Abstract

A new structure, the nondeterministically ordered multiset, and a new structure builder for it, \textit{frong}, are defined as convenient primitives for applicative languages like pure LISP. Since the order of a multiset will be determined as the structure is turned into a list (rather than as the multiset is constructed), functionality or reproducibility of results must be sacrificed in a limited way. This paper is oriented toward the motivation for imbedding nondeterminism in data structures—as opposed to control structure or parameter-passing mechanisms. The heart of the airline reservation system is included.

*Research reported herein was supported (in part) by the national Science Foundation under grants numbered MCS75-06678 A01 and MCS77-22325.
Introduction

Nondeterministic programming is not easily treated in any higher level language. Yet it plays an important role in the design of operating systems, real-time control, and asynchronously parallel systems. It has attracted much attention from software engineers, much programming talent from software shops, and much aggravation from irreproducible events. With processors and parallelism becoming cheaper and ubiquitous, the control of this sort of programming process has become very important. Various attempts to refine the control of interprocess communication \([5,8,9,20,21,34]\) and to extend program proving techniques \([10,17,26,27,32]\) indicate the importance and the limits of this activity under the von Neumann \([2]\) style languages. It has become more apparent that current solutions are not working out and that the impasse is not the fault of the problem. An applicative approach to all programming problems has recently been advanced by Backus \([2]\) in his Turing lecture. The interested reader is referred to his thorough review. His arguments, running orthogonal to current machine architecture, emphasize a purely functional style of programming in which the only binding is of positions within structures to values and the only control structure is function application.

Using only lambda-bindings, function application, conditional expressions, and trivial pimitives, we developed a new perspective on the constructor function which had been used to build all data structures \([11]\). This paper introduces a new constructor, frons, which allows for the postponement of both the content (as in cons) and the order of a structure. Here the same accessing functions
are used to probe into the unordered structure as into an ordered sequence; in fact the user cannot distinguish between such types once it is built. The probing of an unordered structure not only causes the content to appear (as with \texttt{cons}), but it also begins to produce some order within the structure; probing makes a structure behave as if it were ordered.

This nondeterministic structure is called a "multiset" \cite{24}. From the coarse characterization of the multiset above, it is clear that the eventual order might not meet the requirements of stringent functionality; that is, an identically constructed structure may assume a different order. We ask the indulgence of our more theoretically-minded readers during our separation of applicative data structure construction from the rigors of functional programming. (The distinction has more to do with the standard perspective of data structures than with our view of functionality.) The apparent absence of functionality is quite benign and a fair implementation of the constructor exists \cite{15}.

The remainder of this paper is in five parts. The first emphasizes syntax, but offers a rough axiomitization of known primitive operations in our notation. The second introduces the \texttt{multiset} as a broad concept and motivates its application with a simple example. It is followed by a detailed definition of the constructor, \texttt{frons}, which is used to construct multisets. The fourth section presents two examples of its application; the more important one is a real-time interrupt handler which is the heart of a solution for the airline reservation system. The final section reviews the relation of this work with that of others in applicative languages. An appendix is attached which offers an implementation of \texttt{frons} in terms of our syntax and in that of LISP.
Syntax

Let \( \mathbf{1} \) (read "bottom") denote an ill-defined or computationally divergent value. The notation \( x^* \) is used to indicate that evaluation of \( x \) converges in the recursive function theoretical sense; others would write \( x \neq \mathbf{1} \) for \( x^* \). Later we shall require a strict \([11,35]\) two argument "identity" function in order to recover some control over the order of evaluation:

\[
\text{strictify} : \langle x \ y \rangle = y \text{ if } x^* .
\]

We present here some syntactic conventions. A sequence \( s \) of elements \( e_1, e_2, \ldots, e_k \) is denoted by \( \langle e_1 \ e_2 \ldots \ e_k \rangle \) \((k \geq 0)\). A sequence evaluates to a list of the values of its respective elements; infinite sequences \([73,23]\) are also possible. An application is denoted by an infix colon which is taken to be a right-associative operator. A function (or functional expression) occurs to the left and its argument appears at its right. We introduce this notation and primitives familiar from pure LISP \([29]\) in the prototype examples below; take \( s \) as the sequence given above with \( k > 0 \).

\[
\begin{align*}
\text{first} : & s = e_1; \quad \text{first} : <> = \mathbf{1}; \\
\text{rest} : & s = \langle e_2 \ldots e_k \rangle; \quad \text{rest} : <> = \mathbf{1}; \\
\text{cons} : & \langle e \ s \rangle = \langle e \ e_1 \ e_2 \ldots e_k \rangle; \quad \text{cons} : \langle e \ <> \rangle = \langle e \rangle; \\
\text{null} : & s = \text{false}; \quad \text{null} : <> = \text{true}; \\
\text{atom} : & s = \text{false}; \quad \text{atom}\text{null} : s = \text{true}; \\
\text{eq} : & \langle \text{atom}\text{null} : s \ \text{null} : s \rangle = \text{false}; \quad \text{successor} : \langle 6 \rangle = 7; \\
\text{if} : & \langle \text{null} : s \ \text{banana} \ \text{first} : s \rangle = \text{if} \ \text{null} : s \ \text{then} \ \text{banana} \ \text{else} \ \text{first} : s = e_1.
\end{align*}
\]
The last line indicates that arguments are called-by-need [37] or
called-by-delayed-value [36] (or that we are using a lazy evaluator
[8,43]) so that unneeded/undefined ones like banana need not be a
bother (banana is an unbound variable); it also indicates a syntactic
convenience for conditionals which we adopt in the examples below.
We include the starred form of functional combination [12] which
will be extended for multisets later. This is a purely syntactic
convention which effects LISP's MAP functions or Backus's α (apply
to all) operator [2] when a structure appears in the functional
position.

\[
<\text{successor}^*>:<<1 \ 2 \ 3>> = <2 \ 3 \ 4>;
\]
\[
<f^*>:<<a \ b \ c\ <x \ y \ z>> = <f:\langle a \ x\ f:\langle b \ y\ f:\langle c \ z>>>
\]
\[
<f^*>:<<s \ \text{rest}:s\ >= <f:\langle e_1 \ e_2\ f:\langle e_2 \ e_3\ ... f:\langle e_{k-1}\ e_k>>>
\]

A starred function is spread across its arguments, and because we
allow infinite structures, an infinite spreading is possible:

\[
\text{naturals} = \text{cons}:\langle 1 \ <\text{successor}^*>:\langle\text{naturals}>> = <1 \ 2 \ 3 \ 4 \ ... 
\]

Rather than use explicit "lambda" conventions we shall define new
functions from old ones by a pattern equivalent to a prototype
invocation:

\[
\text{or}:\text{disjuncts} \equiv
\begin{align*}
\text{if} & \text{null:disjuncts then false} \\
\text{else if} & \text{first:disjuncts then true} \\
\text{else} & \text{or:rest:disjuncts} .
\end{align*}
\]

Thus, \[
\text{or}:\langle\text{atom:s} \ \text{null:s}>> = \text{false};
\]
\[
\text{or}:\langle\text{atom:s} \ \text{null:s} \ \text{atom:null:s}>> = \text{true};
\]
and \[
\text{or}:\langle\text{atom:s} \ \text{first}::\langle\text{null}::\langle\rangle = 1 .
\]

Multisets

The braced sequence \( m = \{e_1, e_2, \ldots, e_k\} \) denotes a multiset of elements \( e_i \) for \( 1 \leq i \leq k \) where it is not necessary that \( e_i \) for every \( i \). In light of this possibility we want to define first and rest on multisets as a generalization of this behavior on sequences.

\[
\begin{align*}
\text{first}:m &= e_j \text{ where } 1 \leq j \leq k \text{ is chosen so that } e_j \neq \text{ every } i; \quad \text{first}:\{\} = 1; \\
\text{rest}:m &= \{e_1, e_2, \ldots, e_{j-1}, e_{j+1}, \ldots, e_k\} \text{ where } j \text{ is chosen as above.} \\
\text{rest}:\{\} &= 1.
\end{align*}
\]

Note that we do not consider whether or not \( e_i \neq \) for \( i \neq j \); \( j \) is chosen without regard to that.

\[
\text{null}:m = \text{false}; \quad \text{null}:\{\} = \text{true}; \quad \text{atom}:m = \text{false}.
\]

These last rules demonstrate that "\( \{\} \)" = "\( <> \)" within the user's semantics specified by these primitives.

(The reader is referred to a more complete presentation of these semantics \([15]\) for a definition of the behavior of \texttt{cons} and \texttt{frons} (i.e. multisets) in building a structure which is something between a sequence and a multiset.) From these prototypes we can derive that

\[
\begin{align*}
\text{or}:\{\text{atom}:s \text{ null}:s \text{ atom:}\text{null}:s\} &= \text{true}; \\
\text{or}:\{\text{atom}:s \text{ first}::<> \text{ null}::<>\} &= \text{true}; \\
\text{and} \quad \text{or}:\{\text{rest}::<> \text{ first}::<> \text{ rest}::<>\} &= 1.
\end{align*}
\]

These results arise without changing the definition of \texttt{or}. The old definition \([30]\) is sufficient to define "symmetric" \texttt{or} \([31]\) which converges whenever any disjunct is \texttt{true} or when all are \texttt{false}; we just changed the argument.

\[\text{or}^+\]

The behavior depends much on the binding of the variable \texttt{disjuncts} to a single multiset, whence \texttt{first} and \texttt{rest} must choose the same \( j \). See below.
The angle-bracketed combinator is now simply extended to a braced combinator; the brackets merely determine whether the result is taken as a sequence or as a multiset.

\[
\{f^*: <a\ b\ c> <x\ y\ z> \} = \{f: <a\ x> f: <b\ y> f: <c\ z>\};
\{\text{successor}*\}: <\langle\ 2\ 3\rangle> = \{2\ 4\ 3\}.
\]

At this point functionality is set aside. We wish to relax the demands of the semantic sense of functionality ever so carefully so that we can handle nondeterministic problems. This does not mean to imply that we can afford to drop all the trappings of such languages. On the contrary, we want to preserve as many of these properties as possible because the practical benefits of this style of programming may be attributed to the lack of side-effects and the preservation of environments. Therefore, as we introduce nondeterminism we "encapsulate" it so that the desirable properties of determinism in other parts of the language are preserved.
The constructor, frons

Like the sequence constructor cons, we have a multiset constructor frons. In the same way that we understand that

\[ s = \text{cons} : \langle e_1 \text{ cons} : \langle e_2 \ldots \text{ cons} : \langle e_k \ldots \rangle \ldots \rangle \]

we use the new constructor to build the multiset \( m \) from above

\[ \text{frons} : \langle e_1 \text{ frons} : \langle e_2 \ldots \text{ frons} : \langle e_k \{\} \ldots \rangle \ldots \rangle \]

but since order doesn’t matter we also might build \( m \) as

\[ \text{frons} : \langle e_k \ldots \text{ frons} : \langle e_2 \text{ frons} : \langle e_1 \{\} \rangle \ldots \rangle \]

or in many other ways. The rules for coarse interaction (that is, in the absence of bindings) between frons and the other primitives follow.

\[
\begin{align*}
\text{first} : \text{frons} : \langle x \ y \rangle &= \begin{cases} 
  x \text{ if } x^\downarrow; \\
  \text{first} : \text{frons} : \langle x \ y \rangle \text{ if } \text{first} : \text{frons} : \langle x \ y \rangle^\downarrow.
\end{cases} & (1) \\
\text{rest} : \text{frons} : \langle x \ y \rangle &= \begin{cases} 
  y \text{ if } x^\downarrow; \\
  \text{frons} : \langle x \ \text{rest} : \langle y \rangle \text{ if } \text{first} : \text{frons} : \langle x \ \text{rest} : \langle y \rangle \rangle^\downarrow.
\end{cases} & (2) \\
\text{null} : \text{frons} : \langle x \ y \rangle &= \text{false}; & \text{atom} : \text{frons} : \langle x \ y \rangle &= \text{false}.
\end{align*}
\]

Because frons behaves so much like cons, the user has no way of distinguishing how a structure was built after it is built; for access purposes a multiset behaves as if it were a sequence. Indeed, the rules for frons are an extension of the rules for cons \([29,11]\) since cons is defined by alternatives (1) and (3) only; cons has no choice. In the definitions above there is a potentially dangerous freedom of choice when both \( x^\downarrow \) and \( \text{first} : \text{frons} : \langle x \ y \rangle^\downarrow \), because the choice of alternatives may be different in the application of first and rest. That is, we might choose \( \text{first} : \text{frons} : \langle x \ y \rangle = \text{first} : \text{frons} : \langle x \ y \rangle^\downarrow \)
and \text{rest:frons:}<x \ y> = y \ (3) \ so \ that \ x \ might \ not \ be \ in \ the \ multi\text{set} \ \text{frons:}<\text{first:frons:}<x \ y> \ \text{rest:frons:}<x \ y>> \ ! \ The \ picture \ is \ far \ different, \ however, \ if \ the \ multi\text{set} \ has \ already \ been \ constructed \ and \ bound \ to \ a \ variable, \ because \ such \ choices \ on \ a \ bound \ variable \ are \ immutable. \n
Let \ the \ binding \ of \ the \ variable \ v \ be \ \text{frons:}<x \ y> \ (i.e. \ we \ fix \ an \ occurrence \ of \ \text{frons:}<x \ y>). \ Then \ we \ define \n
\begin{align*}
\text{first:} v &= \begin{cases} 
x & \text{if } x^+ \text{ and the other alternative has not yet been chosen for } v; 
\text{first:} y & \text{if } \text{first:} y^+ \text{ and the other alternative has not yet been chosen for } v. 
\end{cases} \quad (5) \\
\text{first:} y &= \begin{cases} 
\text{first:} y & \text{if } \text{first:} y^+ \text{ and the other alternative has not yet been chosen for } v. 
\end{cases} \quad (6) \\
\text{rest:} v &= \begin{cases} 
y & \text{if } x^+ \text{ and the other alternative has not yet been chosen for } v; 
\text{frons:}<x \ \text{rest:} y> & \text{if } \text{first:} y^+ \text{ and the other alternative has not yet been chosen for } v. 
\end{cases} \quad (7) \\
\text{rest:} y &= \begin{cases} 
\text{frons:}<x \ \text{rest:} y> & \text{if } \text{first:} y^+ \text{ and the other alternative has not yet been chosen for } v. 
\end{cases} \quad (8)
\end{align*}

In the case that both \( x^+ \) and \( \text{first:} y^+ \), the choice of alternatives is not so free as above because any previous probing of \( v \) may have already determined it. This predetermination of \( \text{first:} v \ ((5) \ or \ (6)) \) also arises if the other alternative ((7) or (8), respectively) had been chosen as a result of a \( \text{rest:} v \) probe, and vice versa. The net result is that every variable (or field of a structure) bound to a single multi\text{set} must yield consistent results under \( \text{first} \) and \( \text{rest} \) (i.e. \( \text{first} \) and \( \text{rest} \) behave functionally as long as they are invoked over the same bindings). This must extend to all other variables or fields bound to the same multi\text{set} in this or other environments. (Bindings arise from argument-parameter assignment upon function application; they may not be based on lexical matching when \text{frons} is involved.)
This convention on binding applies to braced structures as well: the choice of \( j \) for first:\( m \) must be made consistently with the choice for rest:\( m \) (in the example which introduced braces) because of the binding of \( m \). Similarly, the lambda-binding of disjuncts causes the symmetric or example to work properly. Operationally, every incarnation of a multiset is obliged to keep track of which alternative has been chosen after the choice is made, because its behavior must be immutable. The difference between these rules (5-8) and those above (1-4) is that in the earlier case the explicit constructor created a new (and unshared) multiset so that there were no previous choices possible. The appendix presents an implementation of \textbf{frons} which guarantees this behavior. Like our definition of \textbf{cons} which allows specification of infinite sequences, that definition allows specification of infinite multisets.
Examples

In order to demonstrate the facility of nondeterministic programming with the new constructor \texttt{frons}, we present a few example programs which solve some problems of nondeterminism in an applicative style. In reading these examples the reader should notice how the nondeterminism is isolated into the data structure, so that the program is rather simple. First we consider the problem of flattening a multiset of sequences, $M = \{s_1 s_2 \ldots s_k\}$ into a sequence. In this example the argument is much like a matrix except that we allow for infinite bounds (i.e. the number of rows and the number of elements in each row may be infinite), in which case the first two lines of \texttt{merge} are meaningless.

\begin{verbatim}
merge:M =
  if null:M then M
  else if null: first:M then merge:rest:M
  else cons:<first: first:M
\end{verbatim}

The use of the two constructor functions in \texttt{merge} is particularly interesting. Assuming that the argument is defined appropriately, we can interpret the four possible substitutions of the two constructors in those positions. If both were \texttt{cons}, then \texttt{merge} would append [30] all the rows of M in the order they are presented in M; if the first row of M is infinite then that row is copied.
With the constructors as in the definition of `merge`, the effect is to interleave the various rows of \( M \); so the order of each one is preserved, but elements from other rows may be interspersed in the final result. If both constructors were `frons`, the effect would be to allow any mixture of all the elements of the array as an ordering in the result. In the unusual case that the first constructor was `frons` and the second was `cons`, a similar mixture would result but elements in the result would be restricted to rows only up through the first infinite one in \( M \).

The interleaving behavior is what we desire for the example below. We would like to write a nondeterministic input driver for a time-sharing system. Specifically, we want to solve the input problem for the airline reservation system \([7,40]\). In that problem we have an arbitrary number of remote agents' terminals each producing an infinite stream of characters. Each stream forms a row of an input matrix. Thus every row is infinite (as time passes); and there are an indeterminate number of rows (new terminals may be activated at any time). The problem is to write an applicative program which will accept these characters as soon as they are typed (regardless of the inactivity of another terminal) and interleave them into a single input stream with each character identified according to its source.
We require an auxiliary function which will transform a file--a sequence of characters--into a sequence of pairs--a character and the signature of the file. Furthermore, that sequence of pairs, and all its suffixes, should be strict in the convergence of their first characters. That strictness precludes convergence of such sequences until their first character has been typed at the corresponding remote terminal.

\[
\text{identify}:\langle \text{file id} \rangle \equiv \\
\quad \text{if null:file then file} \\
\quad \text{else strictify:}\langle \text{first:file cons:} \langle \text{first:file id} \rangle \\
\quad \quad \text{identify:}\langle \text{rest:file id} \rangle \rangle .
\]

(Even though \text{identify} specifies a full computation over \text{file}, the reader should satisfy himself that each step is suspended until it is needed.)

It is the multiset of identified files which must be merged. Let us assume that \text{files} is a sequence of the sequential files to be interleaved. Then we may invoke the function \text{fanin} upon \text{files} and \text{naturals} in order to generate the desired stream of agents' communication:

\[
\text{fanin}:\langle \text{files signatures} \rangle \equiv \text{merge:}\{\text{identify}\}:\langle \\
\text{files} \\
\quad \text{signatures} \rangle .
\]

The braced combinator is used to convert the sequence of files and signatures into a multiset of "strictified-identified" files. Thus, the application of \text{first} in \text{merge} can only yield a result from an active teletype.
Relation to other work

There have been many other attempts to introduce nondeterminism in a controlled way into programming languages and models of computation. Of these, we mention only those that are defined for use in languages with "applicative" flavor*. McCarthy [29] suggested a nondeterministic "ambiguity" operator which is clearly not a function.

\[ \text{amb:<x y> = } \begin{cases} 
  x & \text{if } x \downarrow; \\
  y & \text{if } y \downarrow. 
\end{cases} \]

The \text{amb} operator implements pure, mindless nondeterminism; results are not necessarily reproducible. Its failing is that there is no way of knowing which argument was chosen and/or recovering the unchosen one in order to resume its computation later. Others have developed variations on \text{amb} (e.g. Ward [39] developed a function either [3] which allowed for \text{amb} with a finite number of choices) and Hennessey and Ashcroft [19] introduced parameter passing mechanisms for when the choice was to be made: call-time choice and run-time choice. Kosinski [25] introduced an \text{arbiter} into the data flow approach [6]. The arbiter merges streams [4, 25] nondeterministically resulting in a single new stream. Using the arbiter, Dennis [7] fabricates an airline reservation system. The arbiter is behaviorally similar to the program \text{fanin}; but since it is postulated as a hardware primitive, it may only have a finite number of input streams. Arvind and Gostelow [1] have adopted a similar primitive in their data flow model.

* Waldinger and Levitt [30] have developed an unordered structure called a \text{bag} (see also [33]), but a \text{bag} contains only convergent elements.
We have planted all nondeterminism inside data structures. A formalist may not perceive the significance of this choice, but any programmer will; we want to encapsulate nondeterminism where the programmer may plant it and ignore it easily according to his preconceived notions about programming style. A significant contribution of the applicative style to this development is the experience with an applicative regimen which allows the simple introduction of nondeterminism as a trivial twist in programming style. A good programmer follows a few simple rules when he works: don't compute the same thing twice; never build the same structure twice; etc. The use of the second rule is applied when a good programmer borrows a reference to a shared structure rather than copying it. The point here is that the manipulation of data structures has engendered precisely the same programming practice which is required of properly used nondeterminism in an applicative programming language.

Acknowledgement: We are grateful for the contributions of several people to this work, particularly Mitchell Wand for his creative criticisms, Steve Johnson for his insightful implementation [22], and Steve Smoliar for carrying the development of multisets into the practice of software engineering [35].
REFERENCES


Appendix

This appendix presents an implementation of the \texttt{frons} constructor assuming only the \texttt{amb} and \texttt{strictify} primitives defined in this paper, and an understanding of call-by-need \cite{37} or call-by-delayed-value \cite{36} parameter linkages. In particular the reader is cautioned that the list constructor \texttt{cons} from LISP and its accessing functions \texttt{first} and \texttt{rest} (i.e. \texttt{car} and \texttt{cdr}) build structures whose contents are only evaluated once, but such an evaluation is postponed until the structure is probed for the first time in a particular field \cite{11}, just as the first use of a formal parameter causes evaluation of the corresponding arguments. Johnson \cite{22} has implemented such an interpreter for the language defined here in PASCAL for the CDC CYBER computers.

Thus, the function \texttt{insulate} is a bit more than an identity function on structures.

\begin{verbatim}
insulate:<z> ≡ cons:<first:z rest:z> .
\end{verbatim}

Because the argument for the parameter is called-by-need, and because \texttt{cons} does not need it in order to converge, this function always converges before evaluating its argument. Evaluation is postponed until either field of the resulting structure is probed, and then it is only evaluated once.
Then

\[ \text{frons}:<x \ y> = \]
\[ \text{insulate}:<\text{amb}:<\]
\[ \text{strictify}:<x \ \text{cons}:<x \ y>> \]
\[ \text{strictify}:<\text{first}:y \ \text{cons}:<\text{first}:y \ \text{frons}:<x \ \text{rest}:y>> >>.\]

Because \text{insulate} is called immediately from \text{frons}, \text{frons} always converges to yield a \text{cons} data structure. The content of that structure is not determined until it is probed because \text{amb} is not invoked until then. \textbf{Furthermore}, the \text{amb} expression is evaluated at most once for each invocation of \text{frons}. At the time of that probe (by either \text{first} or \text{rest}) one of the choices in \text{amb} is made and that determines whether Lines (5) and (7) or Lines (6) and (8) will be used to interpret the multiset bound once to \text{z}.

In addition we offer a \textbf{similar} definition for a LISP interpreter which uses call-by-need. An example of such an interpreter appears in the appendix to our earlier paper [11]. The only operations which need be added to McCarthy's pure LISP interpreter [30, Ch. 1] are \text{amb} [29] and \text{progl} [30], which happens to coincide (for the wrong reasons) with \text{strictify}. (\text{progl} is not to be implemented using call-by-need although its first argument is never needed, we require strictness.) Then \text{frons} is

\[
\text{(label frons (lambda (x y))}
\*\]
\[
(((\text{lambda (z)} (\text{cons (car z) (cdr z)}))
\*\]
\[
(\text{amb (progl x (cons x y))}
\*\]
\[
(\text{progl (car y) (cons (car y) (frons x (cdr y))))}
\*\]
\[
))))) .
\]