ProTem

Eric Hehner

ProTem is a programming system that serves as both programming language and operating system, and includes a theorem prover to check each step of program composition. This document is an informal specification of ProTem. Formal specifications of the data types and program semantics can be found in the book *a Practical Theory of Programming* (with syntactic differences). “ProTem” also means “for now”, short for the Latin words “pro tempore”.

Programming languages and operating system languages have a lot of functionality in common, but differ greatly in syntax and terminology. These differences are historical, accidental, and unnecessary. They complicate a programmer's life with no benefit. For example, a file is just a variable; file update and storage are just assignment. By unifying the programming language and the operating system commands, both gain in functionality. Communication channels and file piping are as useful in programming as they are in operating systems. Directories are useful in large-scale multi-programmer programs. Conditional execution (if) and indexed loops (for) are useful operating system commands.

ProTem is also designed for easy proof of correctness, including functionality, time requirements, and space requirements. To that end, loops can be constructed by labeling any block of code with a specification, and then using the label within the block of code. For example,

\[
\langle n \geq 0 \Rightarrow n' = 0 \rangle \texttt{if } n > 0 \langle n := n-1. \langle n \geq 0 \Rightarrow n' = 0 \rangle \rangle
\]

The proof methods are the subject of the book *a Practical Theory of Programming* and paper *Specified Blocks*. They do not require preconditions, postconditions, invariants, or variants. If proof is not wanted, then an ordinary identifier can be used as label. For example,

\[
\texttt{loop } \langle \texttt{if } n > 0 \langle n := n-1. \texttt{loop} \rangle \rangle
\]

A primary design criterion is to make ProTem a small, easy-to-learn, easy-to-use language. The size of a language can be measured by the number of symbols and by the complexity of grammar structure, which can be measured by the number of nonterminals. ProTem has 8 keywords. (Python has 35, C has 36, Pascal has 36, Haskell has 37, Ada has 62, MS Basic has 205.) ProTem is presented by a Presentation Grammar, which has just the structure that a programmer needs to know, not all the structure that a compiler needs. It has 2 nonterminals (program and data) plus some informally defined kinds of names. (There is also an LL(1) grammar with 25 nonterminals and an LR(0) grammar with 14 nonterminals at the end of this document. For comparison, the Haskell grammar has 68 nonterminals, and the Python grammar has 87 nonterminals.) The design ethos demands an extremely good reason for adding a new feature to ProTem that requires a new keyword or syntax. That same design ethos will not tolerate any addition to the 2 nonterminals in the Presentation Grammar.

To judge ease of use, you need to use the language, but you may get a sense of the ease of use (and of the beauty of the language, if that is of interest) from reading example programs. For that purpose, there are example programs near the end of this document.

The design of ProTem is complete except for the following. We need to describe and compose picture and sound elements. We need to define touchpad and touchscreen gestures. We may need to define regions of documents and regions of the screen to be clickable links. An implementation of ProTem in ProTem is partly written.
Contents

Introduction
Contents
Symbols
Presentation Grammar
  Precedence
Using ProTem
Data
  Numbers
  Characters
  Binary Values
  Bunches
  Sets
  Strings
  Lists
  Conditional Data
  Functions
  Named Data
  value-Data
Quote and Unquote
Scope
Programs
  Variable Definition
  Assignment
  Constant Definition
  Data Definition
  Data Recursion
  Named Data versus Data Definition
  Constant Definition versus Data Definition
Output and Input
Sequential Composition
Concurrent Composition
if-Program
case-Program
for-Program
Program Definition

Named Program
Plan
Dictionary Definition
Measuring Unit Definition
Forward Definition
Channel Definition
Name Removal
Synonym Definition
Format
Commands
  Execute
  Faster
  Pause
  Abort
  Session
  Undo
  Edit
  Display
  Names
  Memo
  Context
  Verify
Predefined Names
Miscellaneous
Intentionally Omitted Features
Implementation Philosophy
Example Programs
  Portation Simulation
  Quote Notation Lengths
  Minimum Redundancy Codes
  Read Password
Grammars
  LL(1) Grammar
  LR(0) Grammar
Acknowledgements
Symbols

ProTem has 8 keywords, plus 4 kinds of lexeme, and 71 other symbols; altogether they are:

- case
- else
- for
- if
- new
- old
- plan
- value

number
text
name
comment

"  
«  »  `  '  ,  ,..  ;  ;;  ;..  .  !  ?  ??  :  ::  :=  =
≠<  >
\ ≤
≥#
\|\|\∞
&  %
⊤  ⊥
+  –
×  /  _
→  ↔
∧  ∨
^  ^^  @  *
~
¢  $
∈
☐
⊲  ⊳
⊨
⫤
(  )  { }  [ ]  ⟨ ⟩  ¶ ¶ ¶ ¶ ¶ ¶

Keywords and multicharacter symbols cannot have spaces between the characters.

Some of the ProTem symbols may not be on your keyboard. Here are the substitutes.

- for “ and ” use "
- for ‘ use '
- for – use –
- for ≠ use /=
- for ≤ use <=
- for ≥ use >=
- for × use ><
- for → use ->
- for ↔ use <>
- for ∧ use /\ for v use \\
- for ∈ use :
- for ⊥ use false

The names infinity, true, and false are predefined, and redefinable.

A number is formed as one or more decimal digits, with an optional decimal point between digits. A decimal point must have at least one digit on each side of it. Commas and spaces between digits are not allowed. Here are four examples.

```plaintext
0  275  27.5  0.21
```

A text begins with a left-double-quote, continues with any number of any characters (but a double-quote (left or right) within a text must be underlined), and concludes with a right-double-quote. Characters within a text are not limited to any alphabet. Here are five examples.

```
“"  “abc”  “don’t”  “Just say “no”.”  “♣♥♠”
```

A name is either simple or compound. A simple name is either plain or fancy. A plain simple name begins with a letter from an alphabet (this document uses the 26 small italic and 26 capital italic letters of the English alphabet), and continues with any number of letters and decimal digits, except that keywords cannot be names. A fancy simple name begins with « , and continues with any number of any characters not limited to any alphabet, except « and », and ends with ». A compound name is composed of two or more simple names joined with backslash \ characters. For examples:

plain simple names: x  AI  george  refStack
fancy simple names: «Rick & Margaret»  «x’ ≥ x»
compound names: ProTem\grammars\LL1  DCS\«grad recruiting»\«2016-9-8»

At each point in a program, a name is one of

- newname: a simple name that is not defined in the current scope,
- or a compound name that is not defined in its dictionary in the current scope
- oldname: a simple name that is defined in the current scope,
- or a compound name that is defined in its dictionary in the current scope.

An oldname is defined as one of: variablename, constantname, dataname, programname, channelname, unitname, or dictionaryname.

A comment begins with ` and ends at the end of the line. Characters within a comment are not limited to any alphabet. For example: `I❤ProTem
Presentation Grammar

ProTem will be explained in complete detail in the following sections. In this section, for reference, we present the grammar. There are 36 ways of forming a program. There are a few words of explanation on the right side.

new newname : data := data
define variablename with type and initial value
define constantname and evaluate data
define dataname but do not evaluate data
define programme but do not execute program
define channelname with type and initial value
define measuring unitname
define dictionaryname
define dictionaryname with definitions
define synonym
forward definition of dataname or programname
remove or hide
assign variable to value
to channel send output
from channel receive input in this pattern
from channel receive input in this pattern and echo
from channel receive corrected input and echo
define programme and execute program
execute (call) named program
sequential composition
concurrent composition
if-program
if-else-program
case-program
case-else-program
case separator in a case-program or case-else program
for-program, index is constantname
plan, parameter is constantname
plan, parameter is variablename
plan, parameter is output channelname
plan, parameter is input channelname
plan, parameter is dictionaryname
plan, data argument
plan, variable argument
plan, channel argument
plan, dictionary argument
program brackets, scope

There are 60 ways of expressing data. Examples and pronunciations are shown on the right side.

0  1.2
infinity, the infinite number
complex number  \( x + iy \)
percentage  \( x\% = x/100 \)
plus, identity
minus, negation, not
plus, addition

number
\( \infty \)
data & data
\( x \& y = x + iy \)
data %
data + data
\( x\% = x/100 \)
data - data
\( x\% = x/100 \)
data + data
\( x\% = x/100 \)
data - data
\( x\% = x/100 \)
data + data
data – data  
minus, subtraction

data × data  
times, multiplication

data / data  
divided by, division

data ^ data  
to the power, exponentiation

data ^ ^ data  
scale, scientific notation \( x^{^y} = x \times 10^{^y} \)

\( \top \)  
top, true

\( \bot \)  
bottom, false

data ∧ data  
minimum, conjunction, and, set intersection

data ∨ data  
maximum, disjunction, or, set union

data = data  
equals, equation

data ≠ data  
not equals, differs from, exclusive or

data < data  
less than, strict implication, strict subset

data > data  
greater than, strict reverse implication, strict superset

data ≥ data  
greater than or equal to, reverse implication, superset

bunch union \( (0, 2), (2, 5) = 0, 2, 5 \)
bunch from (including) to (excluding) \( 0,..3 = 0,1,2 \)
bunch intersection \( (0, 2, 5) \cap (2, 5, 9) = 2, 5 \)
bunch inclusion \( 2, 5: 0, 2, 5, 9 \)
reverse bunch inclusion \( 0, 2, 5, 9:: 2, 5 \)
bunch size, bunch cardinality \( \varnothing (0,..5) = 5 \)
set size, set cardinality \( \{0, 2, 5\} = 3 \)
elements of a set \( 2, 3 \in \{0, 2, 3, 5\} \)
power \( \varnothing (0,1) = \{null\}, \{0\}, \{1\}, \{0, 1\} \)

“” “abc” “Just say “no”.”

string join \( 0; 2; 5 \)
string from (including) to (excluding) \( 0;..3 = 0;1;2 \)
string indexing \( (2; 3; 7; 3)_2 = 7 \)
string modification \( 3; 4; 5 \ 
\downarrow 1\ 
\leftrightarrow 2; 3; 7; 3 = 4 \)
definite repetition \( 3*2 = 2; 2; 2 \)

list \( [3; 6; 5] \)
list join \( [3; 6];[7; 4] = [3; 6; 7; 4] \)
list length, function size \( #[2; 3; 7; 3] = 4 \)
list index, function argument, composition \( f x \)
pointer indexing \( L @ (i; j) = L \downarrow j \)
function, parameter is constantname \( \langle x: nat \rightarrow x+1 \rangle \)
function, function space \( nat \rightarrow bin = \langle x: nat \rightarrow bin \rangle \)
domain of a list or function \( \square \langle x: nat \rightarrow x+1 \rangle = nat \)
selective union \( 2 \rightarrow 8 \ | [3; 6; 7; 4] = [3; 6; 8; 4] \)
variable name
constant name
data name and evaluate data
the most recent data read on the channel
test for written but unread data on the channel
unit name, positive finite real number constant
conditional data, if data then data else data
named data, defines dataname

value-data, defines variablename
evaluation brackets
Precedence

Here is the precedence (order of execution) of the forms of program.

0  \textbf{new} \hspace{0.5em} \textbf{old} \hspace{0.5em} ::= \hspace{0.5em} ! \hspace{0.5em} ? \hspace{0.5em} \text{programname} \hspace{0.5em} \textbf{plan} \hspace{0.5em} \textbf{if} \hspace{0.5em} \textbf{case} \hspace{0.5em} \textbf{for} \hspace{0.5em} \textbf{[ [ ] ]}

1  \textit{plan argument} \hspace{0.5em} \text{left-to-right}

2  || \hspace{0.5em} \text{left-to-right}

3  .

Program brackets \textbf{[[ ]]} can always be used to change the order of execution.

Here is the precedence (order of evaluation) of the forms of data.

0  \textit{number} \hspace{0.5em} \textit{text} \hspace{0.5em} \textit{name} \hspace{0.5em} \top \hspace{0.5em} \bot \hspace{0.5em} \infty \hspace{0.5em} ( ) \hspace{0.5em} [ ] \hspace{0.5em} \{ \} \hspace{0.5em} \langle \rangle \hspace{0.5em} \textit{value}

1  \textit{list index} \hspace{0.5em} \textit{function argument} \hspace{0.5em} \textit{postfix} \hspace{0.5em} \% \hspace{0.5em} ? ? ? \hspace{0.5em} \textit{infix} \hspace{0.5em} _@& \hspace{0.5em} \text{left-to-right}

2  pre \hspace{0.5em} \textit{fi} \hspace{0.5em} \textit{x} \hspace{0.5em} + \hspace{0.5em} - \hspace{0.5em} \% \hspace{0.5em} | | \hspace{0.5em} \# \hspace{0.5em} ~ \hspace{0.5em} \textit{infix} \hspace{0.5em} \textit{*} \hspace{0.5em} \to \hspace{0.5em} \land \hspace{0.5em} ^{^{\text{right-to-left}}}

3  \textit{infix} \hspace{0.5em} \times \hspace{0.5em} / \hspace{0.5em} \% \hspace{0.5em} \textit{left-to-right}

4  \textit{infix} \hspace{0.5em} + \hspace{0.5em} - \hspace{0.5em} ;; \hspace{0.5em} ; \hspace{0.5em} ;) \hspace{0.5em} \textit{left-to-right}

5  \textit{infix} \hspace{0.5em} , \hspace{0.5em} ,.. \hspace{0.5em} ,\textit{left-to-right}

6  \textit{infix} \hspace{0.5em} = \hspace{0.5em} \neq \hspace{0.5em} < \hspace{0.5em} > \hspace{0.5em} \leq \hspace{0.5em} \geq \hspace{0.5em} :: \hspace{0.5em} \textit{left-to-right}

7  \textit{infix} \hspace{0.5em} \models \hspace{0.5em} \textit{right-to-left}

Evaluation brackets \textbf{( )} can always be used to change the order of evaluation.

The domain of a function must be on precedence level 0 or 1. Any data expression becomes precedence level 0 by putting it in evaluation brackets \textbf{( )}. On level 6, the operators are “continuing”; this means, for example, that \(a=b=c\) neither associates to the left \((a=b)=c\) nor associates to the right \(a=(b=c)\), but means \((a=b)\land(b=c)\). Similarly \(a<b=c\leq d\) means \((a<b)\land(b=c)\land(c\leq d)\).

Using ProTem

Following the prompt \textbf{â}, key in a program, and then, to cause execution, press \textbf{[ctl x]}, which is the simultaneous combination of the control key and the letter \textit{x}. For example, you can key in \new \textbf{temp}:= 2+2

The \! means output to the screen (see \textit{Output and Input}). Then press \textbf{[ctl x]}, causing 4 to be printed on the screen. Thus ProTem can be used as a calculator.

As we will see in \textit{Constant Definition}, the program

\textbf{new \hspace{0.5em} temp}:= 2+2

followed by \textbf{[ctl x]}, saves the result of the calculation 2+2 under the name \textit{temp}, perhaps for use in further calculation. As we will see in \textit{Program Definition}, the program

\textbf{new \hspace{0.5em} myprogram} ["2+2="; \textit{temp}]

followed by \textbf{[ctl x]}, defines and saves a program named \textit{myprogram}. The saved program is not immediately executed. To execute this saved program, key in \textit{myprogram}

followed by \textbf{[ctl x]}.

A program may stretch over many lines. As you key in a program, keywords become bold, plain names become italic, and keyboard substitutes become the proper symbols. You may make changes to a program using the standard cut/copy/paste editing commands and delete (backspace) key anytime before you press \textbf{[ctl x]}. When you press \textbf{[ctl x]}, the program is checked for errors; if there are errors, you are told what the errors are, and invited to correct them; if there are no errors, your program is executed. Saved programs and data can be edited using the command \textbf{[ctl e]} (see \textit{Edit}).
**Data**

ProTem's basic data are numbers, characters, and binary values. ProTem's data structures are bunches, sets, strings, and lists. Also, there are functions, named data, and value-data.

**Numbers**

Numbers are not divided into disjoint types. A natural number is an integer number; an integer number is a rational number; a rational number is a real number; a real number is a complex number. There is also an infinite number $\infty$, greater than all other numbers, and $-\infty$, less than all other numbers, included in the reals.

The one-operand postfix operator `%` means division by 100; for examples, 99.9%, $x\%$ and $(x+y)\%$. There are two one-operand prefix operators `+` and `–`. There are nine two-operand infix operators `+`, `–`, `×`, `/`, `∧`, `∨`, `&`, `^`, `^` . Division of integers, such as `1/2`, may produce a noninteger. Exponentiation is 2-operand infix `^`; for example, $1.2 \times 10^3$ (one point two times ten to the power three), which can be written more briefly as $1.2^3$, and in general, $x^y = x \times 10^y$. The operator `∧` is minimum (arms down, does not hold water; note that `^` and `∧` are different). The operator `∨` is maximum (arms up, holds water). In addition to the number symbols, there are predefined names of numbers such as `pi` (the ratio of a circle's circumference to its diameter), `e` (the base of the natural logarithms), and `i` (the imaginary unit, a square root of $-1$). The complex number $x + iy$ can be written more briefly as `x&y`. There are predefined function names such as `abs`, `arc`, `arccos`, `arcsin`, `arctan`, `ceil`, `cos`, `cosh`, `exp`, `floor`, `im`, `lb`, `ln`, `log`, `mod`, `randInt`, `randReal`, `re`, `round`, `sin`, `sinh`, `sqrt`, `tan`, and `tanh` (see Predefined Names). Predefined names can be redefined.

**Characters**

A character is a text of length 1. We leave it to each implementation to list the characters, and to state their order. In addition to the character symbols such as “a” (small a) and “ ” (space), there are some predefined character names: `delete` (backspace), `tab`, `nl` (new line, next line, return, enter), `click`, `doubleclick`, and `end` (the end-of-file character). Predefined functions `suc` and `pre` give the successor and predecessor in the character order; “ ” (space) comes first and `end` comes last. Predefined functions `charnat` and `natchar` map between characters and their (possibly extended ASCII or unicode) numeric encodings. Some key combinations, such as `ctl x` and `ctl e` are characters and have numeric encodings.

**Binary Values**

The two binary values are $\top$ and $\bot$. Negation is $\neg$, conjunction (minimum) is $\land$, disjunction (maximum) is $\lor$. The infix two-operand operators $=$ and $\neq$ apply to all data in ProTem with a binary result; the two operands may even be of different types.

The order operators $< > \leq \geq$ apply to real numbers (including rationals, integers, and naturals), to characters, to binary values, to sets (subset, superset), to strings of ordered items, and to lists of ordered items, with a binary result. In the binary order, $\bot$ is below $\top$, so $\leq$ is implication.

The postfix operator `??` applies to channels, and has a binary result saying whether there is written but unread data on the channel.
Bunches

Any number, character, binary value, set, string of elements, and list of elements is an elementary bunch, or synonymously, an element. For example, the number 2 is an elementary bunch, or element. Every data expression is a bunch expression, though not all are elementary.

Bunch union is denoted by a comma:

\[ A , B \]  

A union B

For example,

\[ 2, 3, 5, 7 \]

is a bunch of four integers. There is also the notation

\[ x..y \]

where \( x \) and \( y \) are integers or \( \infty \) or \( -\infty \) or characters that satisfy \( x \leq y \). Note that \( x \) is included and \( y \) is excluded. For example, \( 0..3 \) consists of the first three natural numbers 0,1,2, and 5..5 is the empty bunch \( null \).

For any \( A \) and \( B \),

\[ A : B \]

\[ A : B \]

is binary. The size (or cardinality) of \( A \) is \( \varepsilon A \). For examples, \( \varepsilon (0, 1) = 2 \) and \( \varepsilon null = 0 \) and \( \varepsilon (a..b) = b - a \). The evaluation brackets ( ) are needed due to operator precedence. Bunches are equal if and only if they include the same elements, ignoring order and multiplicity.

Bunches serve as a type structure in ProTem, as the contents of sets, and other uses. There are several predefined bunch names:

- \( null \) the empty bunch
- \( nat \) all natural numbers. Examples: 0 1 2
- \( int \) all integer numbers. Examples: –2 –1 0 1 2
- \( rat \) all rational numbers. Examples: 1/2 3.4
- \( real \) all real numbers. Example: \( 2^{1/2} = 1.4142, -1.4142 \) approximately
- \( com \) all complex numbers. Examples: \( -1^{1/2} = i, -i \) 3&4 = 3+4×i
- \( char \) all characters. Examples: “a” “ ” “”
- \( bin \) both binary values: ⊤, ⊥
- \( text \) all texts (character strings). Examples: “abc” “Say “hi”.”
- \( ord \) the ordered type for which \( \& \), \( \vee \), \( < \), \( > \), \( \leq \), \( \geq \) are defined
- \( all \) all ProTem data values

In ProTem, all operators whose precedence is before that of bunch union, except \( \varepsilon \) and \( \setminus \), distribute over bunch union. Infix \( * \) distributes in its left operand only. For examples,

\[ -(3, 5) = -3, -5 \]

\[ (2, 3)+4, 5) = 6, 7, 8 \]

This makes it easy to express the plural naturals \( nat+2 \), the even naturals \( nat\times2 \), the square naturals \( nat^2 \), the natural powers of two \( 2^n \), and many other things.

Sets

A set is formed by enclosing a bunch in set brackets. For examples, \( \{0, 2, 5\} \), \( \{0,..100\} \), \( \{null\} \), \( \{nat\} \). The inverse of set formation is the content operator \( \sim \). For example, \( \sim\{0, 1\} = 0,1 \). The size (or cardinality) of a set, traditionally written \( |S| \), is \( SS \) in ProTem. For examples, \$\{0, 1\} = 2 \), \$\{null\} = 0 \), and \$\{nat\} = \infty \). The element relation is \( x \in S \). For example, \( 1, 2 \in \{0, 1, 2, 3\} \). The
union operator, traditionally $\cup$, is $\lor$ in ProTem. The intersection operator, traditionally $\cap$, is $\land$. Subset, traditionally $\subseteq$, is $\leq$; strict subset is $<$; superset is $\geq$; strict superset is $>$. The power operator $^*$ takes a bunch as operand and produces the bunch of all sets that contain only elements of the operand. For example, $(0, 1) = \{null\}, \{0\}, \{1\}, \{0, 1\}$. In $(0, 1)$, the evaluation brackets ( ) are needed due to operator precedence.

**Strings**

There is a predefined string name:

```
nil
```

the empty string

Any number, character, binary value, list, and function is a one-item string, or synonymously, an item. For example, the number 2 is a one-item string, or item.

String join is denoted by a semi-colon:

```
S; T  join  T
```

For example,

```
2; 3; 5; 7
```

is a string of four integers. There is also the notation

```
x..y  join  y
```

where $x$ and $y$ are integers or characters that satisfy $x \leq y$. Again, $x$ is included and $y$ is excluded. For examples, 0..3 = 0;1;2 and 5..5 = nil.

The length of a string is obtained by the ↔ operator. For examples, $\leftrightarrow(2; 3; 5; 7) = 4$, and $\leftrightarrow(x..y) = y-x$. The evaluation brackets ( ) are needed due to operator precedence.

A string is indexed by the _ operator. Indexing is from 0. For example, $(2; 3; 5; 7)_2 = 5$. A string can be indexed by a string. For example, $(3; 5; 7; 9)_{(2; 1; 2)} = 7; 5; 7$.

If $S$ is a string and $n$ is an index of $S$ and $i$ is any item, then $S^{<n>i}$ is a string like $S$ except that item $n$ is $i$. For example, $3; 5; 9^{<2>8} = 3; 5; 8$. This operator associates from left to right, so $3; 5; 9^{<2>8^{<1>7}} = ((3; 5; 9^{<2>8})^{<1>7} = (3; 5; 8^{<1>7} = 3; 7; 8$. And $3; 5; 9^{<2>8^{<2>7}} = ((3; 5; 9^{<2>8})^{<2>7} = (3; 5; 8^{<2>7} = 3; 5; 7$.

A text is a more convenient notation for a string of characters.

```
“abc” = “a”; “b”; “c”
```

```
“He said “Hi”. ” = “H”; “e”“”; “s”; “a”; “i”; “d”; “ ”; “’”; “H”; “i”; “’”; “’”; “”
```

```
abcdefgij”_(3..6) = “def”
```

Strings are equal if and only if they have the same length, and corresponding items are equal. They are ordered lexicographically. For examples,

```
3; 5 < 3; 5; 2 < 3; 6
```

```
(3; 5)\lor(3; 5; 2) = 3; 5; 2
```

A bunch of items is an item. Join distributes over bunch union, so

```
(3, 4); (5, 6) = 3;5, 3;6, 4;5, 4;6
```

A string is an element (elementary bunch) if and only if all its items are elements.

If $S$ is a string and $n$ is a natural number, then

```
n * S  n copies of S, or n S's
```
is a string, and
\[ * S \] strings of \( S \), or any number of \( S \)'s
is a bunch of strings. For examples,
\[
3*5 = 5;5;5 \\
3*(4, 5) = 4;4;4, 4;4;5, 4;5;4, 5;4;4, 5;4;5, 5;5;4, 5;5;5, 5;5;5;5, 5;5;5;5, \text{and so on}
\]
The * operator distributes over bunch union in its left operand only.
\[
null*5 = null \\
(2,3)*5 = 2*5, 3*5 = 5;5, 5;5;5
\]
Using this semi-distributivity, we have
\[ ^*a = nat^a \]

Lists

A list is a packaged string. It can be written as a string enclosed in list brackets. For example,
\[ [0; 1; 2] \]

Let \( L \) and \( M \) be lists, let \( n \) be a natural number, and let \( p \) be a string of natural numbers. The list operators are:
\[
\square L \quad \text{domain of } L \\
\sim L \quad \text{content of } L \\
# L \quad \text{length of } L \\
L n \quad L \text{ at } n, L \text{ at index } n \\
L @ p \quad L \text{ at } p, L \text{ at pointer } p \\
L ;; M \quad L \text{ join } M \\
L M \quad L \text{ composed with } M \\
L \mid M \quad L \text{ otherwise } M, \text{the selective union of } L \text{ and } M \\
i \rightarrow x \mid L \quad \text{index } i \text{ is item } x \text{ and otherwise } L
\]
plus the operators \( L \wedge M \), \( L \vee M \), \( L=M \), \( L\neq M \), \( L\leq M \), \( L\geq M \). For examples,
\[
\square[10; 11; 12] = 0, 1, 2 \quad \text{the domain of a list} \\
\sim[10; 11; 12] = 10; 11; 12 \quad \text{the content of a list} \\
#[10; 11; 12] = 3 \quad \text{the length of, or number of items in, a list} \\
[10;..20] 5 = 15 \quad \text{indexing starts at zero} \\
[2; 3]; 4; [5; 6; 7]] @ (2; 1; 0) = 6 \\
[0;..10];[10;..20] = [0;..20] \quad \text{joining lists} \\
[10;..20] [3; 6; 5] = [13; 16; 15] \quad \text{composition } (L \mid M)n = L(Mn)
\]

By using the \( \triangle \) operator, a string acts as a pointer to select an item from within an irregular structure. If the list \( L \mid M \) is indexed with \( n \), the result is either \( Ln \) or \( Mn \) depending on whether \( n \) is in the domain \((0;..#L)\) of \( L \). If it is, the result is \( Ln \), otherwise the result is \( Mn \).
\[
[10; 11] \mid [0;..10] = [10; 11; 2;..10] \\
1\rightarrow21 \mid [10; 11; 12] = [10; 21; 12]
\]
The index can be a string, as in
\[
(0;1) \rightarrow 6 \mid [[0; 1; 2]; [3; 4; 5]] = [[0; 6; 2]; [3; 4; 5]]
\]

When a string or list is indexed by a structure, the result has the same structure as the index.
\[
(10;..20) \_ [2; (3, 4); [5; 6; 7]] = [[12; (13, 14); (15; [16; 17])]] \\
[10;..20] [2; (3, 4); [5; 6; 7]] = [12; (13, 14); (15; [16; 17])]
\]

Let \( S = 10; 11; 12 \). Then
\[
S(0, \{1, [2; 1]; 0\}) \\
= \ S_0, \{S_1, [S_2; S_1]; S_0\} \\
= \ 10, \{11, [12; 11]; 10\}
\]

Let \( L = [10; 11; 12] \). Then
\[
L(0, \{1, [2; 1]; 0\}) \\
= \ L_0, \{L_1, [L_2; L_1]; L_0\} \\
= \ 10, \{11, [12; 11]; 10\}
\]

Lists are equal if and only if they have the same length and corresponding items are equal. They are ordered lexicographically. For examples,
\[
[3; 5] < [3; 5; 2] < [3; 6] \\
[3; 5] \lor [3; 5; 2] = [3; 5; 2]
\]

The list brackets \([\ ]\) distribute over bunch union. For example,
\[
[0, 1] = [0], [1]
\]
Thus \([10*nat]\) is all lists of length 10 whose items are natural, and \([4*[6*real]]\) is all 4 by 6 arrays of reals.

**Conditional Data**

The 3-operand expression \( x \models y = z \), pronounced “if \( x \) then \( y \) else \( z \)”, has binary operand \( x \), but \( y \) and \( z \) are of arbitrary type. For example,
\[
y \neq 0 \models x/y \models “nan”
\]
If \( y \neq 0 \) has value \( \top \), then this data expression has number value \( x/y \). If \( y \neq 0 \) has value \( \bot \), then this data expression has text value “nan”. This operator associates from right to left. For example,
\[
(a \models b \models c \models d \models e) = (a \models b \models (c \models d \models e))
\]
If \( a \) has value \( \top \), then this expression has value \( b \), with no need to evaluate \( c, d, \) or \( e \). If \( a \) has value \( \bot \), there is no need to evaluate \( b \).

**Functions**

A function defines a parameter; that is its only job. Let \( p \) (parameter) be any simple name, let \( D \) (domain) be any data expression (but not using \( p \)), and let \( B \) (body) be any data expression (possibly using \( p \) as a constant name for an element of \( D \)). Then \( \langle p: D \to B \rangle \) is a function with parameter \( p \), domain \( D \), and body \( B \). For example,
\[
\langle n: \text{nat} \to n+1 \rangle \quad \text{map \( n \) in \( \text{nat} \) to \( n+1 \)}
\]

is the successor function on the natural numbers. The parameter name begins its scope at the left function bracket \( \langle \) and ends its scope at the right function bracket \( \rangle \) (see **Scope**). Consequently, the parameter name can be any simple name, even one that has already been defined in the scope that encloses the function. The \( \Box \) operator gives the domain of a function. For example,
\[
\Box \langle n: \text{nat} \to n+1 \rangle = \text{nat}
\]
The \# operator gives the size of the function, which is the size of its domain. For examples,
\[
\#(n: (0...10) \to n+1) = 10 \\
\#(n: \text{nat} \to n+1) = \infty
\]
A function of \( n+1 \) parameters is a function of \( 1 \) parameter whose body is a function of \( n \) parameters. For example, the average function
\[
\langle x: \text{rat} \to \langle y: \text{rat} \to (x+y)/2 \rangle \rangle
\]
has two parameters. The notation for applying a function to an argument is the same as that for indexing a list: adjacency. If \( f \) is a function of two parameters, then \( f x y \) applies \( f \) to \( x \) and \( y \).
Caution: in some languages, applying \( f \) to \( x \) and \( y \) is \( f(x, y) \). In ProTem, comma is bunch union, and function application distributes over bunch union. So in ProTem, \( f(x, y) = f x, f y \).

The predefined function \( \text{realtext} \) has four parameters. The first three parameters say how to format a number, and the last is the number to be formatted. For example, \( \text{realtext} \ 4 \ 1 \ 10 \ \pi = "3.1416^{0}" \). We can define a new function

\[
\text{new myrealtext} (\text{realtext} \ 4 \ 1 \ 10)
\]

by supplying just three parameters, and then apply it to a number to be formatted:

\[
\text{myrealtext} \ \pi = "3.1416^{0}"
\]

When the body of a function does not use its parameter, there is a syntax that omits the function brackets \( \langle \rangle \), unused name, and colon. For example, \( 2 \rightarrow 3 \) means \( \langle n : 2 \rightarrow 3 \rangle \) or choose any other parameter name. And \( \text{nat} \rightarrow \text{bin} \) means \( \langle n : \text{nat} \rightarrow \text{bin} \rangle \) or choose any other parameter name.

Argumentation comes before bunch union in precedence, and so it distributes over bunch union.

\[
(f, g) (x, y) = f x, f y, g x, g y
\]

If you want to apply a function to a bunch without distributing over the elements of the bunch, you must package the bunch as a set.

Allowing the body of a function to be a bunch generalizes the function to a relation. For example, \( \text{nat} \rightarrow \text{bin} \) can be viewed in either of the following two ways: it is a function (with unused and omitted parameter) that maps each natural to \( \text{bin} \); it is all functions with domain at least \( \text{nat} \) and range at most \( \text{bin} \). As an example of the latter view, we have

\[
\langle i : \text{int} \rightarrow i \leq 10 \rangle : \text{nat} \rightarrow \text{bin}
\]

If \( f \) and \( g \) are functions, then

\[
f \mid g \quad \text{“f otherwise g”, “the selective union of f and g”}
\]

is a function that behaves like \( f \) when applied to an argument in the domain of \( f \), and otherwise behaves like \( g \).

\[
\square (f \mid g) = \square f, \square g
\]

\[
(f \mid g) x = (x: \square f \Rightarrow f x = g x)
\]

The function \( \langle f : (\text{nat} \rightarrow \text{rat}) \rightarrow f \ 2 \rangle \) is “higher order”, which means it has a function-valued parameter. It can be applied to any function with domain at least \( \text{nat} \) and range at most \( \text{rat} \).

Let \( f \) and \( g \) be functions such that \( -(f : \square g) \) (\( f \) is not in the domain of \( g \)). Then \( g f \) is the composition of \( g \) and \( f \).

\[
(x: \square (g f)) = (x: \square f) \land (f x: \square g)
\]

\[
(g f) x = g (f x)
\]

**Named Data**

Named data has the form

\[
\text{simplename} \ (\text{data})
\]

Within the data, the simplename stands for the data recursively. The simplename begins its scope at the left data bracket \( ( \) and ends its scope at the corresponding right data bracket \( ) \) (see **Scope**). Consequently, the name can be any simple name, even one that has already been defined in the scope that encloses the named data.

For example, here is a bunch of texts (a grammar).
This bunch includes the text “a+b+a–a” and many more texts. It is equivalent to
(“a”, “b”); *( (“+”, “–”); (“a”, “b”) )
Data is evaluated as needed. The preceding bunch is in
finite, but it may appear in a context that does not require its complete evaluation. For example, the binary expression
“a+b+a–a”: term ( “a”, “b”, term; “+”; term, term; “–”; term )
can be evaluated to ⊤ without fully evaluating the named data to the right of the colon.

Here is the factorial function.

fact ( 0 → 1 | n: (nat+1) → n × fact (n–1) )

If we were to fully evaluate it by applying it to all its arguments, the evaluation would take infinite time and memory. But

fact ( 0 → 1 | n: (nat+1) → n × fact (n–1) ) 5

applies the function to one argument, 5, and the evaluation takes finite time and memory.

The expression

tree ( [nil], [tree; int; tree] )
is all binary trees with integer nodes.

Here is a named data expressing the infinitely long text ∞❤ that is all hearts.

hearts ( ❤; hearts )
A named data is equal to the data with all occurrences of the name replaced by the named data. So

hearts ( ❤; hearts ) = ❤; hearts ( ❤; hearts )

❤; hearts ( ❤; hearts ) = ❤; and so on. Evaluation of hearts ( ❤; hearts ) _ 5 gives ❤.

value-Data

A value-data allows us to use a program to compute data. It has the form

value simplename : data := data [ program ]
A local variable, called the value-variable, is defined with a type and initial value. Then the program is executed. The result is the final value of the value-variable. We have not yet presented programs, but the following example, which approximates the base of the natural logarithms e, should give the idea.

value sum: rat := 1
[ new term: rat := 1.
  for i:= 1..15 [term:= term/ i. sum:= sum+term]]

There are no side effects. Nonlocal variables become constants within the program; their values may be used, but assigning them is not permitted. Input from and output to nonlocal channels are not permitted.

All the ways of expressing data can be combined arbitrarily, without restriction. Here we define fun as a function whose body is a value-data. It expresses the number of times 2 is a factor of n.

new fun f( n: (nat+1) → value f: 0..n:= 0
  [new m: 1..n+1:= n.
    loop [if even m [f:= f+1. m:= m/2. loop ]]] )

A value-variable begins its scope at the left program bracket [ and ends its scope at the
corresponding right program bracket ⟭ (see Scope). Consequently, the value-variable can be any simple name, even one that has already been defined in the scope that encloses the value-data. The type and initial value of the value-variable cannot use the value-variable.

**Quote and Unquote**

The predefined function quote takes any data and produces a text representation of the data. Roughly speaking, it quotes its argument. For examples, \( \text{quote } 123 = \text{“123”} \), \( \text{quote } \top = \text{“\top”} \), \( \text{quote } \text{“abc”} = \text{“\text{“abc”}”} \), \( \text{quote } \{0, [1; 2]\} = \text{“\{0,[1;2]\}”} \). The argument is evaluated before quoting; for examples, \( \text{quote } (2 \times 3) = \text{“6”} \), \( \text{quote } (0 < 1) = \text{“\top”} \). In a context where \( x = 3 \), \( \text{quote } x = \text{“3”} \) and \( \text{quote } (x \times 4) = \text{“12”} \). Function quote does not provide any control over the format of the resulting text. For real numbers, fine control over the format of the resulting text is provided by the predefined function realtext.

The predefined function unquote is a kind of inverse of quote. It applies to texts; roughly speaking, it unquotes its argument. For examples, \( \text{unquote } \text{“123”} = 123 \), \( \text{unquote } \text{“\top”} = \top \), \( \text{unquote } \text{“\text{“abc”}”} = \text{“abc”} \), \( \text{unquote } \{0, [1; 2]\} = \{0, [1; 2]\} \). The argument is evaluated after unquoting; for examples, \( \text{unquote } 2 \times 3 = 6 \), \( \text{unquote } 0 < 1 = \top \). In a context where \( x = 3 \), \( \text{unquote } x = 3 \) and \( \text{unquote } (x \times 4) = 12 \). In a context where \( x \) is not defined, \( \text{unquote } x \) is not defined. If the argument does not represent a ProTem data expression, the result is not defined. It is equivalent to the predefined function textreal when applied to a text representing a real number.

Whenever a data that is not a text is used in a context that requires a text, the quote function is applied automatically. For example

\[ !123 \]

places a number where a text should be. So quote is applied automatically ( !quote 123 ) resulting in a text ( !“123” ) as required for output to the screen. If quote has been redefined (see Scope), the predefined quote is used.

Whenever a text data is used in a context that requires a data that is not a text, the unquote function is applied automatically. For example,

\[ \text{“123”} + 1 = \text{unquote “123”} + 1 = 123+1 = 124 \]

If unquote has been redefined (see Scope), the predefined unquote is used.

**Scope**

A name is defined in these six ways: by the keyword new, as a named program, as a function parameter, as a plan parameter, as a for-index, or as a value-variable; we have already met function parameters and value-variables; we shall meet the others shortly. The scope of a name is the part of a program in which the name is defined. The scope of a name is limited by program brackets ⟭ ⟭, except that the scope of a function's parameter is limited by function brackets ⟨ ⟩, and the scope of the dataname of a named data is limited by data brackets ⟨ ⟩.

A name defined in a local scope using the keyword new must be new, not already defined since the most recent opening program bracket ⟭. Its scope extends from its definition through all following sequentially composed programs to the corresponding closing program bracket ⟭. But it may be covered by a redefinition in an inner scope. Using new \( x := 2 \) and new \( x := 3 \) as example definitions, and letting \( A \), \( B \), \( C \), \( D \), and \( E \) stand for arbitrary program forms (but not new or old), in

\[ [A. \text{new } x := 2. \text{[C. new } x := 3. \text{D]}. \text{E}] \]
the definition of \( x \) as the number 2 is not yet in effect in \( A \), but it is in effect in \( B, C, \) and \( E \). The definition that makes \( x \) the number 3 is in effect in \( D \). None of \( A, B, C, D, \) or \( E \) can contain a redefinition of \( x \) unless it is within further scope limiters \([ ]\) or \( \langle \rangle \) or \( (\mid) \).

A name defined by \texttt{new} can become undefined by the keyword \texttt{old}, ending its scope early. So in
\[
\texttt{new \ x := 2. A. old \ x. B}
\]
the definition of \( x \) is in effect in \( A \) but not in \( B \). Within \( B \), the name \( x \) has the same meaning (if any) that it had before the definition \texttt{new \ x := 2}. After \texttt{old \ x}, the name \( x \) is again new and available for definition. However,
\[
\texttt{new \ x := 2. [old \ x. A]}
\]
is not allowed; a scope cannot be ended by \texttt{old} within a subscope.

A scope can be nested inside another scope, which can be nested inside another, and so on. Outside all scope limiters, \([ ]\) and \( \langle \rangle \) and \( (\mid) \), is the persistent scope. A name defined by \texttt{new} in the persistent scope is called a persistent name. Persistent names serve as “files”. When you define a \texttt{new} name in the persistent scope, the name must differ from all names already in the persistent scope. The scope of a persistent name ends only with \texttt{old}. Its scope does not end with the end of a computing session (see \texttt{Session}), not even by switching off the power, which should not cut the power instantly, but should first cause the values of any variables in the persistent scope to be saved in nonvolatile memory. In the persistent scope, there is a dictionary named \texttt{predefined} that contains the predefined names.

Names must be defined before they are used (except for a Forward Definition). Whenever a name is used, it is looked up as follows: first, look in the most local scope (innermost \([ ]\) or \( \langle \rangle \) or \( (\mid) \)); then look in the next most local scope; then the next, and so on; when all local scopes have been searched, look in the persistent scope; finally, look in the \texttt{predefined} dictionary. The first occurrence found is the one that is used.

\section*{Programs}

Some programs are concerned with names: defining a name (\texttt{new}), deleting a name (\texttt{old}). Other programs are variable assignment, input, output, and a variety of ways of combining programs to form larger programs. All programs, including those that define and delete names, are executed in their turn, just like variable assignments and input and output.

\section*{Variable Definition}

Variable definition has the form
\[
\texttt{new \ newname : data := data}
\]
The newname becomes a variablename. Here is an example variable definition.

\[
\texttt{new \ x: nat:= 5}
\]
This defines \( x \) to be a variable assignable to any element in \texttt{nat}, and initially assigned to 5. There is no such thing as an “uninitialized variable” nor the “undefined value” in ProTem. In a variable definition, the data after \( : \) is called the “type” of the variable, and the data after \( := \) is called the “initial value”. The type can be anything except the empty bunch, and the initial value must be an element of the type. The type and initial value can depend on previously defined names, including variables. For example,

\[
\texttt{new \ y: 0..2*x:= x}
\]
defines \( y \) as a variable whose value can be any natural number from (including) 0 up to (excluding) twice the current value of \( x \) (the value of \( x \) at the time this definition is executed), with
initial value equal to the current value of \( x \). But the type and initial value cannot make use of the variable being defined. ProTem allows the type and initial value to be “static” (known at compile time), as in the first example (variable \( x \)), and it allows the type and initial value to be “dynamic” (not known until execution time), as in the second example (variable \( y \)).

Here are five more examples.

\begin{itemize}
  \item \texttt{new s: 10*int:= 10*0}
  \item \texttt{new \texttt{t}}: \texttt{text:= ""}
  \item \texttt{new \texttt{u}}: \texttt{(0,.20)*char:= “abcde”}
  \item \texttt{new L: [*nat]:= [0;..10]}
  \item \texttt{new A: \{3*[3*bin][]{}}:
\end{itemize}

In the first example, \( s \) is defined as a variable that can be assigned to any string of ten integers, and is initially assigned to the string of ten zeroes. In the second example, \( \texttt{t} \) is a predefined bunch equal to \( *\texttt{char} \), so \( \texttt{t} \) can be assigned to any text, and is initially assigned to the empty text. In the next example, \( \texttt{u} \) is defined as a variable that can be assigned to any text of length less than 20, and is initially assigned to the text “abcde”. In the next example, \( \texttt{L} \) is defined as a variable that can be assigned to any list of naturals, and is initially assigned to the first ten naturals. In the last example, \( \texttt{A} \) is defined as a variable that can be assigned to a \( 3 \times 3 \) array of binary values, and is initially assigned to a \( 3 \times 3 \) array of \( \top \).

**Assignment**

A variable can be reassigned by the assignment program. It has the form

\begin{itemize}
  \item \texttt{variablename := data}
\end{itemize}

Here are five examples using the definitions of the previous subsection.

\begin{itemize}
  \item \texttt{x:= \texttt{x+1}} \hspace{1cm} \text{‘now} \hspace{0.2cm} \texttt{x = 6}
  \item \texttt{s:= \texttt{s+2}:= \texttt{8}} \hspace{1cm} \text{‘now} \hspace{0.2cm} \texttt{s = 0; 0; 8; 0; 0; 0; 0; 0; 0; 0}
  \item \texttt{u:= \texttt{u(0;..3)}} \hspace{1cm} \text{‘now} \hspace{0.2cm} \texttt{u = “abc”}
  \item \texttt{L:= 3\rightarrow 5 | 5\rightarrow 7 L} \hspace{1cm} \text{‘now} \hspace{0.2cm} \texttt{L = [0; 0; 0; 5; 0; 7; 0; 0; 0; 0]}
  \item \texttt{A:= (1; 2)\rightarrow \bot \bot A} \hspace{1cm} \text{‘now} \hspace{0.2cm} \texttt{A = [\top; \top; \top]; [\top; \top; \bot]; [\top; \top; \top]}
\end{itemize}

The data on the right of \( := \) must be an element in the type (evaluated at definition) of the variable on the left of \( := \). As in the examples, the data on the right of \( := \) can make use of the variable on the left of \( := \).

**Constant Definition**

Constant definition has the form

\begin{itemize}
  \item \texttt{new newname := data}
\end{itemize}

The newname becomes a constantname. The data on the right of \( := \) cannot make use of the name on the left of \( := \). The constantname cannot be reassigned. Here are three constant definitions.

\begin{itemize}
  \item \texttt{new size:= 10.}
  \item \texttt{new piBy2:= \pi / 2.}
  \item \texttt{new range:= 0,..size}
\end{itemize}

where \( \pi \) is a predefined constant name.

A constant may use variables to express its value. For example

\begin{itemize}
  \item \texttt{new xplus1:= x+1}
\end{itemize}

The current value of variable \( x \) is used to evaluate \( x+1 \), and \( xplus1 \) expresses that value. Variable \( x \) may later be reassigned to another value, but that does not affect the value of \( xplus1 \). For example,
Data Definition

Data definition has the form

```
new newname (| data |)
```

The newname becomes a dataname. Data definition gives the data a name, but does not evaluate the data.

```
new xplus2 (| x+2 |)
```

makes the value of `xplus2` depend on the value of variable `x`. As `x` changes value, `xplus2` changes value so that `xplus2 = x+2` is always `⊤`. In the constant definition of `xplus1` earlier, `x+1` is evaluated once, at definition time. In the data definition of `xplus2`, `x+2` is not evaluated at definition time; it is evaluated every time `xplus2` is used in a context that requires its value.

```
new x: nat:= 3. ` assigns x to 3
new xplus2 (| x+2 |).
new xplus2 (| x+2 |).
```

A data definition can depend indirectly on a variable. For example,

```
new twoxplus4 (2×xplus2)
```

makes `twoxplus4` depend indirectly on the value of variable `x`. In this definition, the value of `xplus2` is not required, so it is not evaluated. If `x` currently has value 3, then `x:= twoxplus4` assigns `x` to 10. Then another `x:= twoxplus4` assigns `x` to 24.

Data Recursion

In a variable definition, the type and initial value cannot depend on the variable being defined.

```
new bad: 0,..2×bad:= bad ` illegal
```

is not allowed due to the two occurrences of `bad` to the right of the colon. Likewise a constant definition cannot be recursive.

Data definition does allow recursion. The next example defines `div` to be the integer division function for natural numbers.

```
new div ⟨〈a: nat→〈d: (nat+1)→a<d=0=1 even a=2×div (a/2) d = 1 + div (a−d) d〉〉
```

Here is a function that eats arguments until it is fed argument 0.

```
new eat ⟨〈n: nat→n=0=0=eat〉
```

So `eat 5 2 0 = 0`, and `eat 4 7 3 8 0 = 0`, and `eat 1 2 = eat` and `eat: nat→(0, eat)`.

Here is a baseless recursion. If it is evaluated, its evaluation is nonterminating.

```
new rec ⟨rec⟩
```

Named Data versus Data Definition

Since we have data definition, named data is unnecessary. For example,

```
new t: tree ⟨[text],[tree; tree]]:= ["abc"]
```

defines `t` to be a tree-valued variable (binary tree, text at the leaves). This example is almost equivalent to
new tree ([text], [tree; tree]). new t: tree:= [“abc”]. old tree

The difference occurs when tree has already been defined in the current scope; in that case, the named data is still all right, but the data definition is not.

In the next example, variable \( g \) is assigned to the greatest common divisor of 10 and 15.

\[
g := \text{gcd} \langle a: (\text{nat}+1) \rightarrow b: (\text{nat}+1) \rightarrow a=b \Leftarrow a \land a<b \Leftarrow \text{gcd } a (b-a) = \text{gcd} (a-b) b \rangle \rangle 10 \ 15
\]

This example is exactly equivalent to

\[
\text{new } \text{gcd} \langle a: (\text{nat}+1) \rightarrow b: (\text{nat}+1) \rightarrow a=b \Leftarrow a \land a<b \Leftarrow \text{gcd } a (b-a) = \text{gcd} (a-b) b \rangle \rangle.
\]

\[
g := \text{gcd} 10 15
\]

Named data is a succinct way of recursively defining some data (in this example, the \( \text{gcd} \) function), and using the data once, immediately. However, if you need to use the data (call \( \text{gcd} \)) more than once, later, you must make a data definition.

**Constant Definition versus Data Definition**

A constant definition evaluates its data once, at definition time, whereas a data definition evaluates its data each time its value is required. If the data is fully evaluated, there is no difference. For example, there is no difference between these two definitions:

new five:= 5
new five \((5)\)

When there are no variables used to express the value (neither directly nor indirectly), there is no semantic difference between data definition and constant definition, but there may be an efficiency difference. Compare these two definitions.

new six:= 5+1
new six \((5+1)\)

If the value of \( six \) is never required, the data definition \((5)\) is more efficient. If the value of \( six \) is required once, they are equally efficient. If the value of \( six \) is required two or more times, the constant definition := is more efficient. Here is a more interesting comparison.

new double:= \(\langle n: (0,..10) \rightarrow 2\times n\rangle\)
new double \(\langle n: (0,..10) \rightarrow 2\times n\rangle\)

The constant definition causes the function to be evaluated by applying it to all its arguments and storing the results. In effect, the function is evaluated to the list

\[0; 2; 4; 6; 8; 10; 12; 14; 16; 18\]

Then, when the value of \( double \) applied to an argument is required, that argument indexes the list. The data definition does not evaluate the function. Each time the value of \( double \) applied to an argument is required, the body of the function is evaluated. Which one is more efficient depends on the size of the domain, the complexity of the result, and the number of times the definition is used.

**Output and Input**

Each channel is defined to transmit a specific type of value. The output channels \( \text{screen} \), \( \text{printer} \), and \( \text{mail} \), and the input channels \( \text{keys} \) and \( \text{mail} \) are predefined to transmit text. The input channel \( \text{microphone} \) and the output channel \( \text{speaker} \) are predefined to transmit sound. We can define local channels to transmit any type of value (see **Channel Definition**).

Output has the form

channelname ! data

Channel \( \text{screen} \) accepts text, which is displayed on the screen. The program

\( \text{screen}! \ “\text{Hi there.”} \)

sends the text “Hi there.” to the screen. Output is buffered so it will be available when \( \text{screen} \) is
ready to receive it. Texts can be joined and sent together.

\[ \text{screen! "Answer = \text{quote} x; \text{nl} } \]

where \text{quote} is a predefined function that converts to a text, and \text{nl} is the new line character, or next line character, or return character. Function \text{quote} can be omitted (see \text{Quote and Unquote}).

When \text{screen} or \text{printer} receives the \text{delete} (backspace) character, the previous character is deleted; when the \text{tab} character is received, some amount of space is substituted; when the \text{nl} character is received, further characters start on a new line.

Email input and output are made convenient for nonprogrammers by a mail program, but at the level of ProTem programming, emailing looks like this:

\[ \text{mail! "To: hehner@cs.utoronto.ca}
\]

\[ \text{Hey Rick, have a look at the \text{exec} program: \text{quote};
\]

\[ \text{definition \text{"exec"}} \]

The keyboard is a program that runs concurrently with other programs; you do not need to initiate it; it is already running. It monitors what key combinations are pressed, and for what duration, and outputs a string of characters. The shift-A combination is a single character “A”. Likewise the control-Q combination is a single character. The click button is just a key like any other; \text{click} is a character, and \text{doubleclick} is a character.

Input has two forms: without echo, and with echo. The first form, without echo, is

\[ \text{channelname ? data} \]

Text from the keyboard (including the click button) can be received from channel \text{keys}. Five characters of input are received from channel \text{keys} by saying

\[ \text{keys? 5*char} \]

What follows \text{?} is called the pattern (or grammar). If input is not yet available, it is awaited. The input read is the earliest input on that channel that has not yet been read. The \text{tab}, \text{delete}, and \text{nl} characters may be part of the input; they are just characters like any others. The input is not echoed on the screen. The shortest input that fits the pattern is read. The program

\[ \text{keys? text; \text{nl}} \]

reads text up to and including the first \text{nl} character, but

\[ \text{keys? text} \]

just inputs the empty text.

Patterns \text{digit} and \text{numpat} are predefined as follows.

\[ \text{new digit:= "0", "1", "2", "3", "4", "5", "6", "7", "8", "9"}. \]

\[ \text{new numpat:= ("+", "–", ""); digit; *digit; (".", digit; *digit), "").} \]

To receive a text that can be interpreted as a number, possibly preceded or followed by spaces, possibly preceded by a sign, ending in a new line character, input

\[ \text{keys? \text{"\text{\textquotesingle}} \text{\textquotesingle}; \text{numpat; \text{\textquotesingle}} \text{\textquotesingle}; \text{nl}} \]

If \text{nl} were omitted, only leading spaces and an optional sign and the first digit would be read; subsequent digits, a point, and more digits would not be read.

When input is received, it is referred to by the channel name followed by \text{?}. After the previous example input, we might have the assignment

\[ x:= \text{unquote (\text{keys}?) + unquote (\text{keys}?)} \]

where \text{unquote} is a predefined function that converts from a text to the type required. Function \text{unquote} can be omitted (see \text{Quote and Unquote}), so we may write

\[ x:= \text{keys} + \text{keys} \]
Both occurrences of \texttt{keys?} refer to the same input, and \texttt{keys?} continues to refer to the same input until the next time new input is read from channel \texttt{keys}.

If \( c \) is the name of an input channel, then \texttt{c??} is a binary expression with value \( \top \) if there is written but unread data on the channel, and \( \bot \) if there is not. For example,

\begin{verbatim}
if keys?? [keys? char. screen! keys?] else [screen! "Are you still there?"]
\end{verbatim}

(See \texttt{if-Program} and \texttt{Sequential Composition}.) Input on a channel that does not currently have any written-but-not-yet-read data waits until data is written to the channel by a concurrent program.

If channel \( c \) is defined to input text, the program

\begin{verbatim}
c? “y”, “n”
\end{verbatim}

inputs one character, either “y” or “n”, from channel \( c \). If the first available character on channel \( c \) is “a”, or more generally, if the input on the channel does not fit the pattern, what happens is not defined. Here are three options.

- The program cannot be executed, so execution ends.
- An error message is sent to channel \texttt{screen} to say that the input is unacceptable, and execution ends.
- An error message is sent to channel \texttt{screen} to say that the input is unacceptable, and the sender is given another opportunity to send an input that fits the pattern.

What happens depends on the implementation, on the channel, and on the number of attempts. Perhaps the last option is appropriate for channel \texttt{keys}, and the first is appropriate for a secure channel. Of course, you can read a character \( c? \texttt{char} \) and program what happens if it's neither “y” nor “n”.

An input program consisting of an input channel name, a question mark, and a pattern, does not echo the input on the screen; the input is invisible. This is useful, for example, when reading a password (see \texttt{Read Password}). To input and echo together, character by character, add an exclamation mark and the output channel name for the echo.

\begin{verbatim}
channelName ? data ! channelName
\end{verbatim}

For example,

\begin{verbatim}
keys? text; nl !screen
\end{verbatim}

This programs inputs, from channel \texttt{keys}, text to and including the first \texttt{nl} character, and outputs the same on channel \texttt{screen}. Each character is echoed as it is input. A space \( \rightarrow \), tab \( \leftarrow \), new line \( \Rightarrow \), or delete \( \Rightarrow \) character is displayed visibly as a graphic symbol.

In an input with echo, the pattern between the input question mark and the echo exclamation mark can be omitted. This results in a special pattern called the correcting pattern. For example,

\begin{verbatim}
keys?!screen
\end{verbatim}

The correcting pattern reads a line of text to and including the first \texttt{nl} character, but this text is corrected according to \texttt{delete} (backspace) characters. The echo displays space, tab, and new line characters as spaces, tabs, and new lines, and displays delete characters as deletions of previous characters. If, during correction, there are more \texttt{delete} characters than other characters, the extra \texttt{delete} characters are ignored. After this input, the value of \texttt{keys?} is the corrected text, excluding the final \texttt{nl}; the \texttt{nl} character is consumed, but not included in the value read.

An output channel name can be omitted, in which case the output channel is \texttt{screen}. For example,

\begin{verbatim}
! “Hello World”
\end{verbatim}

prints \texttt{Hello World} on channel \texttt{screen}. And

\begin{verbatim}
! 2+2
\end{verbatim}

prints \texttt{4} on channel \texttt{screen}. Thus ProTem can be used as a calculator. This program is short for
screen! quote (2+2)

An input channel name can be omitted, in which case the input channel is keys. For example,

? char

reads one character from keys, with no echo. The most recent text read on channel keys can be referred to as just ?. And ?? is a binary expression saying whether there is written but unread data on channel keys. If the input channel is omitted, and the name keys has been redefined, the input channel is the predefined channel keys. If the output channel is omitted, and the name screen has been redefined, the output channel is the predefined channel screen.

The most common input program

?!

is short for keys? ! screen. It reads one line from keys, correcting it according to delete characters, up to the first nl, which is not included in the value of ?, and echoes character by character to screen. The output program

!?

is short for screen! ? keys. It prints on screen the most recently read input on keys.

The expression unquote (keys?) can be written unquote (?) or keys? or ?. But it cannot be written unquote keys? because that is parsed as (unquote keys)? . And it cannot be written unquote ? because the compiler will complain that unquote is not a channel. Similarly for quote (keys??) . When keys? is shortened to ? and used as an index to a list, or as an argument to a function or plan, it must be in evaluation brackets (?). When keys?? is shortened to ?? and used as an argument to a function or plan, it must be in evaluation brackets (??).

In summary, output and input are, respectively,

channelname ! data
channelname ? pattern ! channelname

After reading input, the input most recently read is referred to as channelname ?

The binary expression

channelname ??

says whether there is written but unread data on the channel. If the channelname is keys or screen it can be omitted. If input echoing is not wanted, omit both ! and the echo channelname. In an input program with echo, omitting the pattern results in the correcting pattern.

Sequential Composition

Sequential composition is denoted by a period (point, dot).

program . program

It is an infix connective; in other words, the period comes between and joins programs.

Concurrent Composition

Concurrent composition has the form

program || program

The concurrent composition of programs P, Q, and R is P||Q||R. A variable defined before the concurrent composition remains a variable in at most one component of the concurrent composition; in all the other components of the concurrent composition, it becomes a constant.


In the concurrent composition [a:= 4. A] || [b:= 8. B], variable a can be reassigned in one of the
concurrent programs, but not in both. Likewise variable \( b \) can be reassigned in one of the concurrent programs, but not in both. At the start of \( A \), variable \( a \) has value 4, constant \( b \) has value 2, and data \( c \) has value 6. At the start of \( B \), constant \( a \) has value 1, variable \( b \) has value 8, and data \( c \) has value 9. If \( A \) does not reassign \( a \), and \( B \) does not reassign \( b \), then at the start of \( C \), variable \( a \) has value 4, variable \( b \) has value 8, and data \( c \) has value 12.

\[
\text{new } a : \text{nat} := 1 \parallel a := 2 \quad \text{illegal}
\]
is not allowed because the use of \( a \) does not sequentially follow its definition. Concurrent programs cannot affect each other through assignments of variables. For co-operation, programs can communicate with each other on channels (see Channel Definition).

We have previously seen that program brackets \( [] \) are scope limiters. In the previous paragraph, we see their other use: precedence. They are needed for a concurrent composition of sequential compositions

\[
[A. B] \parallel [C. D]
\]
because concurrent composition has precedence over sequential composition.

In summary: a name defined in one part of a concurrent composition cannot be defined or used in the other parts, but it can be used in a sequentially following program; a variable defined before a concurrent composition remains a variable in at most one part of the composition, and is a constant in the other parts.

**if-Program**

An if-program has the form

\[
\text{if data } [ \text{ program } ]
\]
The if-program

\[
\text{if } b [ \text{ P } ]
\]
is executed as follows: binary expression \( b \) is evaluated; if its value is \( \top \), then program \( [P] \) is executed; if its value is \( \bot \), then program \( [P] \) is not executed. An if-else-program has the form

\[
\text{if data } [ \text{ program } ] \text{ else } [ \text{ program } ]
\]
The if-else-program

\[
\text{if } b [ \text{ P } ] \text{ else } [ \text{ Q } ]
\]
is executed as follows: binary expression \( b \) is evaluated; if its value is \( \top \), then program \( [P] \) is executed and program \( [Q] \) is not executed; if its value is \( \bot \), then program \( [P] \) is not executed and program \( [Q] \) is executed. The program if \( b [P] \) is equivalent to if \( b [P] \) else \( [\text{ok}] \) where \( \text{ok} \) is a predefined program whose execution does nothing and takes no time.

**case-Program**

A case-program has the form

\[
\text{case data } [ \text{ program } ]
\]
in which the program is a sequence of cases. For example, in a case-program with three cases,

\[
\text{case } n [ P ] [ Q ] [ R ]
\]
the programs, called cases, are numbered in order. Case 0 is \( [P] \), case 1 is \( [Q] \), and case 2 is \( [R] \). It is executed as follows: natural expression \( n \) is evaluated; then one of the programs is executed. In the example, if \( n \) has value 1, then just \( [Q] \) is executed. If \( n \) is equal to or greater than the number of cases, an error message is printed and execution stops. The example case-program is equivalent to

\[
\text{if } n=0 [P] \text{ else } [\text{if } n=1 [Q] \text{ else } [\text{if } n=2 [R] \text{ else } [!]\text{"Error: case index too large.". stop}]]}
\]
Here is another example, also with three cases:
\[
\text{case } 2-n \\
\{ \text{[ok]} \} \text{ case 0: all is well, nothing to do} \\
\{ \text{["warning: don't do it again"]} \} \text{ case 1: warning} \\
\{ \text{["you're fired"]} \} \text{ case 2: firing}
\]

As the example illustrates, it is helpful to put a comment with each case to say the case index.

A case-else-program has the form
\[
\text{case } \text{data } [ \text{ program } ] \text{ else } [ \text{ program } ] 
\]

in which the first program is a sequence of cases. A case-else-program with three cases,
\[
\text{case } n \{ P \} \{ Q \} \{ R \} \text{ else } [ S ]
\]
is the same as the case-program, but if \( n \geq 3 \), then the program \( [ S ] \) after else is executed. The example case-else-program is equivalent to
\[
\text{if } n=0 \{ P \} \text{ else } \{ \text{if } n=1 \{ Q \} \text{ else } \{ \text{if } n=2 \{ R \} \text{ else } [ S ] \} \}
\]

for-Program

A for-program has the form
\[
\text{for } \text{simplename } := \text{ data } [ \text{ program } ]
\]

The simplename becomes a constantname in the program. Here is a nest of for-programs that computes the transitive closure of array \( A \):
\[
\text{for } j:= 0;..n \{ \text{for } i:= 0;..n \{ \text{if } A i j \land A j k \{ A := (i;k) \to \top \} \} \}
\]

The if-program \( \text{if } A i j \land A j k \{ A := (i;k) \to \top \} \) can be restated as
\[
A := (i;k) \to (A i k \lor (A i j \land A j k)) \land A
\]

if you prefer. The name being defined by for is called a for-index. It is known only within the for-program, and it is known there as a constant, and so it is not assignable. In the example, each of the for-indexes \( j, i \), and \( k \) takes values \( 0, 1, 2 \), and so on up to but excluding \( n \).

For a second example, here is the sieve of Eratosthenes.

\[
\text{new } n:= 1000. \text{ new prime: } n*ybin:= 2\times1; (n-2)*1.
\text{for } i:= 2;..\text{ceil } (\text{sqrt } n) \{ \text{if } \text{prime}_i \{ \text{for } j:= i;..\text{ceil } (n/i) \{ \text{prime} := \text{prime } \lor i \times j \} \}\}
\]

A for-index is “by initial value”, so
\[
\text{for } i:= x; x \{ x := i+1 \}
\]

increases \( x \) by 1 , not 2 .

This next example prints the natural numbers forever.
\[
\text{for } n:= 0;..\infty \{ ! n; " " \}
\]

After the := we can have any string expression; the for-index stands for each item in the string, in sequence. We can also have any bunch expression; the for-index stands for each element of the bunch, concurrently. As an example (note the use of .. rather than ;.. as earlier),
\[
\text{for } i:= 0;..\#L \{ L := i \to 0 \land L \}
\]

makes the items of \( L \) be 0 , concurrently. We could also write either of these:
\[
\text{for } i:= \square L \{ L := i \to 0 \land L \}
\text{L := [\#L*0]}
\]

The domain of the for-index can also be a bunch of strings, or a string of bunches, and so on, so that sequential and concurrent execution can be nested within each other. (Note: distribution and factoring laws are not applied; the structure of the expression is the structure of execution.)
A `for`-index begins its scope after `⟦` and ends its scope at the corresponding `⟧`. Consequently, the `for`-index can be any simple name, even one that has already been defined in the scope that encloses the `for`-program. The domain of the `for`-index cannot use the `for`-index.

**Program Definition**

Program definition has the form

```
new newname [ program ]
```

The `newname` becomes a programname within the program. Program definition gives a program a name, but does not execute the program. For example,

```
new switchends [L:= 0 → L 9 | 9 → L 0 | L]
```

Execution of this definition defines the program name `switchends`, but does not execute program `⟦L:= 0 → L 9 | 9 → L 0 | L⟧`. After execution of this definition, the name `switchends` can be used to call (cause execution of) the program it names. Program definitions can be recursive. Predefined program names include `await`, `exec`, `ok`, `stop`, `wait`.

A fancy name can be used as a specification. For example,

```
new « xʹ > x » [x:= x+1]
```

The specification `« xʹ > x »` is implemented (refined, implied) by the program `⟦x:= x+1⟧`. A prover is invoked by the `ctl v` command (see `Verify`). If the specification is written within the language that the prover understands, the prover attempts to prove that the specification is implemented (refined, implied) by the program. If the program makes use of a specification, the inner specification is used in the outer proof. For example,

```
new « xʹ = 0 » [if x≠0 [x:= x–1. « xʹ = 0 »]]
```

In the program `⟦if x≠0 [x:= x–1. « xʹ = 0 »]⟧`, the specification `« xʹ = 0 »` means exactly what it says, rather than the program that it names. Thus the use of specifications makes complicated fixed-point semantics unnecessary. If the prover fails to understand the specification, or fails to prove the refinement, it informs the programmer, and treats the specification as just a name. (See the paper `Specified Blocks`.)

**Named Program**

A named program has the form

```
simplename [ program ]
```

The `simplename` becomes a programname within the program that it names. Although the name is situated before a `⟦`, it begins its scope after `⟦` and ends its scope at the closing `⟧`. Consequently, the name can be any simple name, even one that has already been defined in the scope that encloses the named program. The name is attached to the program (like a program definition), and the program is executed (unlike a program definition). One purpose of this naming is to make loops. Here is a two-dimensional search for `x` in an `n×m` array `A` of integers (that is, `A: [n*[m*int]]`).

```
new i: nat:= 0.
tryThisI [if i=n [! x; " does not occur."]]
else [new j: nat:= 0.
tryThisJ [if j=m [i:= i+1. tryThisJ]]
else [if A i j = x [! x; " occurs at "; i; " "; j]
else [j:= j+1. tryThisJ]]]]
```

The next example is a fast remainder program, assigning natural variable `r` to the remainder when natural `a` is divided by positive natural `d`, using only addition and subtraction.
**Programs and Names**

A named program is equivalent to the program (including the program brackets \([\ ]\)) with occurrences of the name replaced by the named program. So the fast remainder example means the same as

\[
r := a.\\
\text{outerloop \[if \ r \geq d \ \text{new} \ dd: \ \text{nat} := d.\\
\text{innerloop} \[r := r - dd. \ dd := dd + dd.\\
\text{if} \ r \lt dd \ [\text{outerloop}] \ \text{else} \ [\text{innerloop}]]]\\
\]

The next example illustrates that named programs provide general recursion, not just tail recursion. It computes the Fibonacci numbers \(x := \text{fib} \ n\) and \(y := \text{fib} \ (n+1)\) in \(\log n\) time.

\[
\text{Fib} \ [\text{if} \ n = 0 \ [x := 0. \ y := 1] \ \text{else} \ [\text{if} \ \text{odd} \ n \ [n := (n-1)/2. \ \text{Fib.} \ n := x. \ x := x^2 + y^2. \ y := 2nx + y^2] \ \text{else} \ [n := n/2 - 1. \ \text{Fib.} \ n := x. \ x := 2x^2y + y^2. \ y := n^2 + y^2 + x]]
\]

As in a program definition, a fancy name can be used as a specification. For example,

\[
\langle x' > x \rangle [x := x+1]
\]

The specification \(\langle x' > x \rangle\) is implemented (refined, implied) by the program \([x := x+1]\). A prover is invoked by the \texttt{ctl v} command (see \texttt{Verify}). If the specification is written within the language that the prover understands, the prover attempts to prove that the specification is implemented (refined, implied) by the program. If the program makes use of a specification, the inner specification is used in the outer proof. For example,

\[
\langle x' = 0 \rangle [\text{if} \ n \neq 0 \ [x := x-1. \ \langle x' = 0 \rangle]]
\]

Inside the program brackets, the specification \(\langle x' = 0 \rangle\) means exactly what it says, rather than the program that it names. Thus the use of specifications makes complicated fixed-point semantics unnecessary. If the prover fails to understand the specification, or fails to prove the refinement, it informs the programmer, and treats the specification as just a name. (See \texttt{Specified Blocks}.)

Suppose a name is defined within a loop. For example, the name \(a\) in

\[
\text{infiniteLoop} \ [\text{new} \ a := \text{“a”}. \ !a. \ \text{infiniteLoop}]\]

Executing this loop prints an infinite sequence of the letter “a”. Replacing the call with the named program, it is equivalent to

\[
[\text{new} \ a := \text{“a”}. \ !a. \ \text{infiniteLoop} \ [\text{new} \ a := \text{“a”}. \ !a. \ \text{infiniteLoop}]]
\]

In a general recursion, each call opens a new scope, and each new definition hides but does not destroy the previous definition. But when the recursive call is the last action performed in the named program (a tail recursion), as in this example, the old scope and its definitions cannot be used again, so the new scope replaces the old one; the scopes and variables do not pile up.

Let \textit{name} be a new name (not defined in the local scope), and let \textit{program} be a program, possibly using the name \textit{name}. Then the following three lines are equivalent to each other.

\[r := a.\\\text{outerloop} \ [\text{if} \ r \geq d \ [\text{new} \ dd: \ \text{nat} := d.\\\text{innerloop} \ [r := r - dd. \ dd := dd + dd.\\\text{if} \ r < dd \ [\text{outerloop}] \ \text{else} \ [\text{innerloop}]]]]\]
name [program]
[new name [program], name]
new name [program], name, old name

Therefore named programs are unnecessary. For example
loop [if n>0 [n:= n–1. loop]]
is equivalent to
[new loop [if n>0 [n:= n–1. loop]]. loop]

A named program is a succinct way of recursively defining a program, and using the program once. However, if you need to use (call) a program more than once, you must make a program definition.

Plan

A plan is a program with a parameter. There are five forms of plan. The first is
plan simplename := data [ program ]
The simplename is being defined as a constant parameter within the program. It can be any simple name, even one that has already been defined in the current scope. Its type (after the := ) cannot make use of the parameter. The scope of the parameter is from [ to ]. For example,
plan y:= real [x:= x*y]]
A plan can be argumented by adjacency in the same way that lists are indexed and functions are argumented. The argument provides a value for the parameter. For example,
plan y:= real [x:= x*y]] 3
is the same as
x:= x*3
Commonly, a plan is named
new P [plan y:= real [x:= x*y]]
and then argumented P 3 , but a plan is not required to have a name.

A program with n+1 parameters is a program with 1 parameter whose body is a program with n parameters. For example, here is a program with two parameters.
plan x:= int [plan y:= int [z:= x+y]]
Each argument reduces the number of parameters.
plan x:= int [plan y:= int [z:= x+y]] 3 4
is equivalent to
plan y:= int [z:= 3+y]] 4
which is equivalent to
z:= 3+4

A plan can be named (in a program definition or named program); a plan can be argumented; and a plan can be the body of a plan. A plan that is not fully argumented cannot be executed. A plan that has been fully argumented can be used wherever any program can be used. For examples,
plan x:= int [plan y:= int [z:= x+y]]. z:= 2  `this is not allowed
plan x:= int [plan y:= int [z:= x+y]] 3. z:= 2  `this is not allowed
plan x:= int [plan y:= int [z:= x+y]] 3 4. z:= 2  `this is allowed

Here is a named program to find the maximum value in nonempty list L in \( log (#L) \) time. (L is a variable, and its initial value is destroyed in the process.) We define \( \text{findmax} \ i \ j \) to find the maximum in the segment of \( L \) from (including) index \( i \) to (excluding) index \( j \), reporting the result as \( L \i \), and then apply it to \( 0 \ (#L) \), which makes \( L \ 0 \) the maximum value of the initial list.
findmax \[\text{plan \ } i := \Box L \ [\text{plan \ } j := \Box L+1 \]
\[\text{if } j - i \geq 2 \ [\text{findmax (i \ div \ (i+j) \ 2) \ || \ findmax (\ div \ (i+j) \ 2) \ j.} \]
L := i \rightarrow L i \lor L (\ div \ (i+j) \ 2) \lor L] \] 0 (#L)

In the previous paragraphs, the parameter is a constant (note the := ); it is not assignable. It is “by initial value”, so
\[\text{plan \ } i := \text{int \ } [x := i. \ y := i] \ (x+1) \]
assigns both \(x\) and \(y\) to a value one greater than \(x\)'s initial value.

The second form of plan
\[\text{plan \ simplename : data \ [ program \ ]} \]
(note the : ) defines a variable parameter. For example,
\[\text{plan \ } x : \text{int \ } [x := 3] \]
A plan with a variable parameter applies to a variable argument. But it cannot be applied to a variable appearing in the plan. This restriction is required for reasoning about the plan. This example plan can be applied to any variable, even one named \(x\), because that \(x\) (the argument) is nonlocal, and is not the local variable \(x\) (the parameter) appearing in the plan. But the plan
\[\text{plan \ } x : \text{int \ } [x := 3. \ y := 4] \]
cannot be applied to variable \(y\).

Here is a plan named \(\text{norm}\) to reduce rational \(\text{num/denom}\) to lowest terms. If \(a = 8\) and \(b = 12\), then after \(\text{norm a b}\), we have \(a = 2\) and \(b = 3\).
\[\text{new \ norm \ [plan \ num: \text{nat+1 \ [plan \ denom: \text{nat+1 \ [normalize \ num/denom} \ [new \ gcd \ \langle a: \ (\text{nat+1}) \rightarrow \langle b: \ (\text{nat+1}) \rightarrow \text{greatest common divisor of} \ a \ and \ b \]
a = b \Rightarrow a \Rightarrow a < b \Rightarrow \gcd a \ (b-a) \Rightarrow \gcd (a-b) b```].}
\[\text{new \ g := gcd num denom. \ num := num/g. \ denom := denom/g ]} \]

The main use for variable parameters is probably to affect many files in the same way; for example, a plan to sort files.

The next form of plan
\[\text{plan \ simplename ! data \ [ program \ ]} \]
creates a plan with an output channel parameter. For example.
\[\text{plan \ c! text \ [c! \ “abc”]} \]
This plan can be applied to any channel that receives text. A plan with a channel parameter cannot be applied to a channel appearing in the plan. This example plan can be applied to any output channel, even one named \(c\), because that \(c\) (the argument) is nonlocal, and is not the local channel \(c\) (the parameter) appearing in the plan. But
\[\text{plan \ c! text \ [c! \ “abc”. \ d! \ “def”]} \]
cannot be applied to channel \(d\). The channel name \(\text{screen}\) cannot be omitted when used as an argument for an output channel parameter.

The next form of plan
\[\text{plan \ simplename ? data \ [ program \ ]} \]
creates a plan with an input channel parameter. For example.
\[\text{plan \ c? text \ [c? \ 3*char. \ screen! c?]} \]
This plan can be applied to any input channel that delivers text. But
\[\text{plan \ c? text \ [c? \ text; \ nl. \ d? \ text; \ nl]} \]
cannot be applied to channel \(d\). The channel name \(\text{keys}\) cannot be omitted when used as an argument for an input channel parameter.
The following program \textit{pps} has three channel parameters. On the first, \textit{a}, it reads the coefficients of a rational power series; on the second, \textit{b}, it reads the coefficients of another rational power series; on the last, \textit{c}, it writes the coefficients of the product power series.

\begin{verbatim}
  new pps [plan a? rat [plan b? rat [plan c! rat
  [ a? rat || b? rat c! a?x|b?.
  new a0:= a? || new b0:= b? || new d? rat! 0. `see Channel Definition
  pps a b d
  || [a? rat || b? rat c! a0x|b+a?x|b0.
  loop [a? rat || b? rat || d? rat c! a0x|b+d+a?x|b0. loop]]]]]]
\end{verbatim}

The final form of plan

\begin{verbatim}
  plan simplename [\ program ]
\end{verbatim}

creates a plan with a dictionary parameter. This plan can be applied to any dictionary name that does not appear in the plan.

\textbf{Dictionary Definition}

Dictionaries are the way you organize your programs and data. There are two forms of dictionary definition. The first form is

\begin{verbatim}
  new newname \ 
\end{verbatim}

The newname becomes a dictionary name. To create a new dictionary named \textit{abc}, write

\begin{verbatim}
  new abc\ 
\end{verbatim}

(It does not matter whether there are spaces between the name and the backslash.)

Now you can define names within this dictionary. For example,

\begin{verbatim}
  new abc\x:= 2
\end{verbatim}

defines \textit{x} in dictionary \textit{abc} to be the constant 2. (It does not matter whether there are spaces before or after the backslash.) This constant can then be used as \textit{abc\x}. A name being defined in a dictionary must not already be defined in that dictionary in the current scope. If the current scope is a local scope, the name becomes undefined at the end of the scope, as usual. Each name in a dictionary is defined, using the keyword \textit{new} and a compound name, to be one of the following: a variable name, a constant name, a data name, a program name, a channel name, a unit name, or a dictionary name. To define new dictionary \textit{def} within dictionary \textit{abc} write

\begin{verbatim}
  new abc\def
\end{verbatim}

When a name in a dictionary is defined to be a dictionary, this dictionary also can contain names, some of which can be defined as dictionaries, and so on. So a dictionary can be a tree structure. Suppose there is a dictionary named \textit{ProTem} within which there is a dictionary named \textit{grammars} within which there is a text named \textit{LL1}. Its name is \textit{ProTem\grammars\LL1}.

A name within a dictionary can be undefined by \textit{old} in the same scope where it was defined (see \textbf{Name Removal}). For example,

\begin{verbatim}
  old abc\x
\end{verbatim}

ends the definition of \textit{x} in dictionary \textit{abc}. In the scope where \textit{abc} was defined,

\begin{verbatim}
  old abc
\end{verbatim}

ends the definition of dictionary \textit{abc}. When a name becomes undefined, what it named remains in existence, anonymously, as long as something refers to it. When a dictionary becomes undefined, so do all the names within it. Here is an example.
new stack\.
new stack\: *nat:= nil.
new stack\:push [plan x:= nat [stack\:= stack\; x]].
new stack\:pop [stack\:= stack\_(_0;..↔stack\_1)].
new stack\:top (stack\_(_↔stack\_1)).
old stack\s

Dictionary stack now has three visible names in it: push, pop, and top. Variable s still exists, but it is hidden.

The second form of dictionary definition is
new newname \ dictionaryname

We can create a dictionary named parseStack populated with the same definitions as stack.
new parseStack\stack

It is equivalent to writing
new parseStack\.
new parseStack\: *nat:= nil.
new parseStack\:push [plan x:= nat [parseStack\:= parseStack\; x]].
new parseStack\:pop [parseStack\:= parseStack\_(_0;..↔parseStack\_1)].
new parseStack\:top (parseStack\_(_↔parseStack\_1)).
old parseStack\s

Here is a recursive dictionary definition that is probably useless.
new d\.
new d\c:= 2.
new d\d

Within d there is a dictionary named d with the same two definitions as in the outer d. Now d\c, d\d\c, d\d\d\c, d\d\d\d\c, and so on, are all the constant 2.

There is a dictionary in the persistent scope named predefined (see Predefined Names).

Measuring Unit Definition

There are three predefined units of measurement. They are g, representing mass in grams, m, representing distance in meters, and s, representing time in seconds. A unit of measurement has all the properties of an unknown positive finite real number constant. So, for example, we write 10\times m/s for the speed 10 meters per second. And we can define
new km:= 1000\times m
to make km be a kilometer, and
new h:= 60\times 60\times s
to make h be an hour. So 1\times m/s = 3.6\times km/h evaluates to \top. To assign a variable to a quantity with units attached, the variable's type must have compatible units attached. For example,
new speed: real\times m/s:= 3.6\times km/h
assigns speed to 1\times m/s (which could be written m/s\times 1 or m\times 1/s or 1/s\times m or just m/s).

You can define a new unit of measurement, unrelated to the existing units. Measuring unit definition has the form
new newname #1

The newname becomes a unitname. For example,
new sheet #1
defines a new unit of measurement called the sheet. Now you can define the related units
new quire := 25 × sheet.
new ream := 20 × quire

You can define a variable using the new units.
new order := 3 × ream

This assigns order to 1500 × sheet. Another example is a monetary unit, such as
new dollar #1.
new cent := dollar%

When the value 5×m/s is converted to text by quote, the result is “5 m/s” without the × sign and without evaluating the unknown real value m/s. And unquote “5 m/s” = 5×m/s. Similarly for all units of measurement. One more example: quote (2×3×km/h) = “1.6667 m/s”.

Forward Definition

Forward definition has the form
new newname
The newname becomes either a dataname or a programname. For example
new mutual
is a notice that a definition of mutual will follow later in the same scope. In a data definition or program definition, the scope of the name being defined starts with the definition. A forward definition facilitates mutual recursion by starting the scope of a data name or program name even before its definition. For example, leaving gaps for missing parts, in
new f := 3. [new f. new g ( | ) f g . new f ( | ) f g . ]
the inner f and g are each defined in terms of both of them. Without the forward definition of f (following [ ], g would be defined in terms of the earlier constant definition new f := 3.

Channel Definition

Channel definition has the form
new newname ? data ! data
The newname becomes a channelname. The definition
new c ? nat! 0
defines c to be a new local channel that transmits values of type nat, with 0 as initial output and input. Channel c can be used for output and input. Now c? refers to the most recent input on the channel, and c! is a binary expression saying whether there is written but unread data on channel c. Before there has been any output or input, c? refers to the initial output and input supplied in the channel definition, and c! is ⊥. The type of the channel cannot use the name of the channel being defined.

To use a channel for communication between concurrent programs, define the channel before the start of the concurrent programs. Then one of the concurrent programs can write to the channel, and all the concurrent programs can read from the channel; they read the same inputs independently. If a program sequentially follows a concurrent composition, on each channel, the sequentially following program begins reading after all the inputs read by all concurrent programs. For example,
new c ? nat! 0. AllB. D
On channel c, program D begins reading after the last input read by A or B.

new c ? nat! 0. x := c? `assigns x to 0
new c ? nat! 0. c! 7. x := c? `assigns x to 0
new c ? nat! 0. c! 7. c ? nat. x := c? `assigns x to 7
new c? nat! 0. c? nat. c! 7. x:= c?  `deadlock, stuck at c? nat
new c? nat! 0. c? nat. x:= c? // c! 7 `assigns x to 7
new c? nat! 0. c! 7. c? nat. x:= c?  `assigns x to 7
new c? nat! 0. c! 7. c? nat. x:= c?. y:= c?] `assigns x and y to 7

Name Removal

Names defined with the keyword **new** can be undeclared with the keyword **old**. Name removal has the form

old oldname

Ironically, by saying **old** x, the name x becomes available for reuse as a new name. Even though a name becomes undeclared, what it named will remain as long as there is an indirect way to refer to it. For example, in predefined dictionary **rand** there are three names next, Int, and Real. They might be defined as:

```protem
new rand\var: 0..maxint:= 123456789. `will be hidden
new rand\next [rand\var:= mod (rand\var \times 5^13) maxint].
new rand\Int (from: int \rightarrow (to: int \rightarrow floor (from + (to-from) rand\var/maxint))).
new rand\Real (from: real \rightarrow (to: real \rightarrow from + (to-from) rand\var/maxint))).
```

old rand\var

Variable rand\var is now hidden; its name is removed, but rand\next, rand\Int, and rand\Real still use it. Note that rand\Int 0 10 has the same value each time it is used until rand\next is called. So

```
rand\Int 0 10 + rand\Int 0 10 = 2 \times rand\Int 0 10
```

has value $\top$. Similarly for rand\Real.

Synonym Definition

Synonym definition has the form

new newname oldname

The newname becomes a synonym for the oldname. One use is to shorten all names that are deep within several dictionaries. For example, if dictionary a contains dictionary b, which contains dictionary c, which contains variable x, then

```
new x a\b\c\d\x
```

shortens the name a\b\c\d\x to just x. The definition

```
new d a\b\c\d
```

shortens all names within a\b\c\d, for example, from a\b\c\d\x to d\x.

Another use is to rename something. To rename a to b, write

```
new b a. old a
```

The following sequence swaps the names p and q.

```
new t p. old p. new p q. old q. new q t. old t
```

Format

Although not part of the ProTem language, here are some suggested formatting (indentation) rules. The choice of alternative depends on the length of component data and programs.

```
A. B   for x:= A [B]
or A. or for x:= A [B]
```

---------------------------------------------

```protem
A. B
```

---------------------------------------------

```protem
or for x:= A [B]
```
More indentation would show the structure better, but it would crowd programs onto the right side of the page. Consistent indentation improves readability, and is useful redundancy for error checking. (Python uses indentation as part of the syntax; so in Python, indentation is not redundant and cannot be used for error checking.)

**Commands**

There are 12 commands in ProTem. They are not presented in the grammar, and they cannot be part of a stored program. A command may be given at any time; it does not have to respect the grammatical structure of a program. Each command is the control character combined with a letter. The commands are:

- **ctl x**: execute with type checking for debugging
- **ctl f**: execute without type checking for faster execution
- **ctl p**: pause or resume execution
- **ctl a**: abort execution
- **ctl s**: stop current session and start new session
- **ctl u**: undo current session
- **ctl e**: enter or exit editor for saved program or data
- **ctl d**: display source or object code for saved persistent program or data
- **ctl n**: display names defined in current scope or persistent scope or in dictionary
- **ctl m**: attach or modify or retrieve memo to defined name
- **ctl c**: generate context comments
- **ctl v**: verify program according to a specification

**Execute**

Use **ctl x** to execute the program keyed in since the latest prompt \( \bowtie \). Execution includes checking that values assigned are in the variable's type. Type checking is useful for debugging and verification, but is redundant in a correct program.

**Faster**

Use **ctl f** to execute the program keyed in since the latest prompt \( \bowtie \). Execution omits checking that values assigned are in the variable's type, and that makes execution faster.
Pause

If there is a program being executed, the \texttt{[ctl p]} command pauses its execution. If there is a paused program, \texttt{[ctl p]} resumes its execution.

Abort

Use \texttt{[ctl a]} to abort execution of the currently executing or paused program.

Session

Sessions are defined for security and error recovery. When the computer is turned on, a session begins. When some idle time passes (how much time is a parameter of the system and may be set to infinity), a session ends and a new one begins. When the computer is turned off, a session ends. The \texttt{[ctl s]} command causes the current session to end and a new session to begin. If there is a currently executing or paused program, ending the session asks if you want to abort it.

Sessions do not define the lifetime of definitions. A definition in the persistent scope, outside all program bracket pairs \texttt{⟦⟧}, lasts from the execution of the definition (\texttt{new}) to the execution of the corresponding name removal (\texttt{old}). This may be less than a session, or more than a session. Turning off the computer should not cut the power instantly, but should first cause the values of any variables in the persistent scope to be saved in nonvolatile memory.

At the start of each session, a programmer must login. This connects the programmer to their persistent scope. Then channels \texttt{keys}, \texttt{screen}, \texttt{microphone}, \texttt{speaker}, and \texttt{printer} are initialized and connected to the session. The data definition \texttt{session}, which is dependent on channel \texttt{keys}, is a text consisting of all keystrokes since the start of the current session. (This is quite practical: an hour of hard work produces only 10 kbytes of keystrokes.) This text can be saved as a record of work done, or for error recovery (see \texttt{Undo}).

Undo

The command \texttt{[ctl u]} undoes a session (except for inputs and outputs and \texttt{session}). Implementing it requires capturing the state at the start of a session. On many computers, returning to the prior state may be cheap; nonvolatile memory (that does not require power) contains the state as it was at the start of the current session, and volatile memory (that requires power) contains the current state. After undo, you can capture the current value of \texttt{session}, let us call it \texttt{recovery},

\begin{verbatim}
    new recovery: text:= session
\end{verbatim}

then reassign \texttt{recovery}, and then execute the result by writing \texttt{exec recovery}. This gives us perfectly flexible error recovery for the modest cost of a keystroke file.

Edit

The edit command \texttt{[ctl e]} is used to modify an existing persistent program or data definition. It invokes a dialogue using \texttt{keys} and \texttt{screen} to determine which definition. It then invokes an editor. In the editor, \texttt{[ctl e]} exits the editor, and asks if you want to throw away the old definition, and save and compile the new definition. If the new definition has an error, you receive an error message, an error comment is inserted into the saved source, and the compiled object code, when executed, prints “unable to execute [definition name]”. If you want to create a definition using the editor, first create the definition, for example, \texttt{new p \{ok\}}, and then invoke the editor to modify it. If you want to
delete a persistent definition, use `old`.

**Display**

The command `ctl d` starts a dialogue using `keys` and `screen` to determine the name (simple or compound) of the persistent program or data whose source or object code you want to view.

**Names**

The command `ctl n` begins a dialogue using `keys` and `screen` to determine whether you want the names defined in the current scope, the persistent scope, or in a dictionary (which may be a subdictionary). In a (sub)dictionary you will see only the first-level names, not the names in its subdictionaries.

**Memo**

Each definition can optionally have a memo attached to it. The memo might explain the purpose or use of the definition. It is there to be read by a human, not for execution. A memo is similar to a comment that you would make at the point of definition, but differs in that you can retrieve it anytime. The command `ctl m` starts a dialogue using `keys` and `screen` to determine which name (simple or compound), whether you want to attach a new memo, modify an existing memo, or retrieve an existing memo. For example, you may say that you want to attach the memo

```
This variable accumulates the sum of the products.
```

to name `x`. Asking for the memo attached to predefined name `e` prints

```
e := 2.718281828459045 (approximately) constant  The base of the natural logarithms.
```

**Context**

The command `ctl c` starts a dialogue using `keys` and `screen` to determine the program, bracketed by `⟦⟧`, for which context comments are wanted. The comments are then generated. These comments say which nonlocal names are used, and in what way they are used. Here is the format.

```
`input: on these channels`
`output: on these channels`
`use: the values of these variables, constants, data names, units, and function names`
`assign: these variables`
`call: these program names and plan names`
`refer: to these dictionaries`
```
If there already are comments in this format, they are replaced. For examples of context comments, see [Example Programs](#). Additionally, a programmer may want to include comments like

```
`spec: specification`
`pre: precondition`
`post: postcondition`
`inv: invariant`
```
but these are not generated by `ctl c`.

**Verify**

The command `ctl v` starts a dialogue using `keys` and `screen` to determine the program, bracketed by `⟦⟧` and named by a fancy name, for which verification is wanted. The verification is then attempted. (See [Program Definition](#) and [Named Program](#).)
Predefined Names

The predefined names are defined in dictionary `predefined` in the persistent scope (see Scope). Predefined names can be redefined in a local scope. For example, one of the predefined names is the imaginary number `i` (a square root of `-1`). You may also want to define a local variable `i`. If you do, you can still refer to the predefined `i` as `predefined\$i` (unless you have covered the name `predefined` with a local definition). If predefined name `i` is covered by a definition in a scope outside the local scope where you are working, you can get back the simple name `i` as the imaginary number in these three ways:

- `new i:= predefined\$i` `\`constant definition`  
- `new i predefined\$i` `\`synonym definition`  
- `new i:= 0\&1` `\`constant definition`  

The command `ctl n` lists the names in the `predefined` dictionary. The command `ctl m` gets a description of a predefined name. The collection of predefined names may change over time.

Each predefined name is one of:

- `variable` at present, there are no predefined variables
- `constant` evaluated; not assignable
- `data` unevaluated; evaluation upon use; not assignable
- `program` unexecuted; execution upon use
- `channel` reinitialized at the start of each session
- `unit` unrelated to other predefined units
- `dictionary` at present, there is one predefined dictionary `rand`

Some definitions use `§` (solutions, those) which is not part of ProTem; it is defined in *a Practical Theory of Programming*. Here are the predefined names.

`abs: com→real data` Absolute value. `abs x = sqrt (re x^2 + im x^2)`.

`all = com, char, bin, fall, *all, [all] data` All ProTem data values.

`arc: com → §(r: real → 0 ≤ r < 2×pi) data` The angle or arc of a complex number.

`arccos: (§(r: real → -1 ≤ r ≤ +1)) → §(r: real → 0 ≤ r ≤ pi/2) data` A trigonometric function.

`arcsin: (§(r: real → -1 ≤ r ≤ +1)) → §(r: real → 0 ≤ r ≤ pi/2) data` A trigonometric function.

`arctan: real → §(r: real → 0 < r < pi/2) data` A trigonometric function.

`await program` A plan with one constant parameter of type `real\*s`. If the argument represents the present or a future time, its execution does nothing but takes time until the instant given by the argument. If the argument represents the present or a past time, its execution does nothing and takes no time. See `time` and `wait` and `s`.

`back: (*nat) → *nat data` If `i` is an item, `back (s; i) = s`.

`bin:= T, \bot constant` The binary values.

`bold: text→text data` Same text but in bold font.

`capital: char constant` The English capital letters. See `small` and `letter`.

`ceil: real→int data` `r ≤ ceil r < r+1`  

`char data` The characters.

`charnat: char→nat data` A one-to-one function with inverse `natchar`. The encoding might be extended ASCII or unicode. Character combinations, for example shift-option-a, also have numeric encodings.

`click: char constant` The click character.

`com = real&real data` The complex numbers.

`cos: real → §(r: real → -1 ≤ r ≤ +1) data` The trigonometric cosine function.
cosh: com→com data  The hyperbolic cosine function.
ctlx: char data  The combination of the control and x characters.
cursor: nat; nat data  A data name whose value is the current cursor position.
definition: text→text data  If the argument is the name of a persistent definition, then the result is the textual definition. For example, if combo is defined as new combo:= “10-20-30”, then
definition “combo” = “combo:= “10-20-30”” . Otherwise the result is “undefined”.
define: char constant  The delete or backspace character.
div: real → ((§(r: real → r>0)) → int data  div a d is the integer quotient when a is divided by d.
(0 ≤ mod a d < d) ∧ (a = div a d × d + mod a d)
doubleclick: char constant  The doubleclick character.
drain program  A plan with one input channel parameter. Reads and discards any unread input.
e:= 2.718281828459045 (approximately) constant  The base of the natural logarithms.
end: char constant  The end-of-file character. It is greater than all other characters.
even: int→bin data  A function that says whether its argument is even.
exec program  A plan with one text parameter. If the argument represents a ProTem program, the execution is that of the represented program. It “unquotes” its argument. If applied to “x:= x+1”, the “x” refers to whatever x refers to at the location where exec “x:= x+1” occurs. If the argument does not represent a ProTem program, execution displays an error message. To evaluate data represented as text, for example “2+2”, use unquote or textreal.
exp: com→com data  The exponential function. exp x = e^x.
false:= ⊥ constant  A binary value.
find: all→all→nat data  If i is an item in string S, then find i S is the index of its first occurrence; if not, then find i S = ⊥.
fit: int→text→text data  If i≥0 then fit i t is a text of length i obtained from t either by chopping off excess characters from the right end or by extending t with spaces on the right end. If i≤0 then fit i t is a text of length –i obtained from t either by chopping off excess characters from the left end or by extending t with spaces on the left end.
floor: real→int data  floor r ≤ r < 1 + floor r
g unit  A mass of one gram.
i:= 0&1 constant  An imaginary number: a square root of –1.
im: com→real data  The imaginary part of a complex number.
infinity:= ∞ constant  An infinite number, greater than all other numbers.
int = nat, –nat data  The integers.
itl: text→text data  Same text but in italic font.
keys? text! “” channel  To the program that monitors key presses, it is an output channel; to all other programs, it is an input channel.
lb: (§(r: real → r>0)) → real data  The binary (base 2) logarithm.
let:= small, capital constant  The English small and capital letters.
ln: §(r: real → r>0) → real data  The natural or Napierian (base e) logarithm.
log: §(r: real → r>0) → real data  The common (base 10) logarithm.
m unit  A distance of one meter.
mail channel  A text input and output channel for email.
machine: all→all→nat data  If pattern occurs within subject, then machine pattern subject is the index of its first occurrence. If not, then machine pattern subject = ⊥ subject.
maxint: int constant  The maximum representable integer (machine dependent).
maxreal: real constant  The maximum representable real number (machine dependent).
microphone? *sound! silence channel  To the microphone, it is an output channel; to all other programs, it is an input channel.
minint: int constant  The minimum representable integer (machine dependent).
mod: real → ($\lfloor r: real \to r>0\rfloor$) → real  data  mod a d is the remainder when a is divided by d .

(0 ≤ mod a d < d) ∧ (a = div a d × d + mod a d)

movie = *picture  data A string of pictures.

nat = 0,..,∞ data The natural numbers.

natchar: nat→char  data A one-to-one function with inverse charnat . The encoding might be extended ASCII or unicode. Character combinations, for example shift-option-a, also have numeric encodings.

nil  constant The empty string.

nl: char constant The new line character or next line character or return character.

null  constant The empty bunch.

numpat: text  constant A text pattern for numbers, useful for reading a number from a text channel.

odd: int→bin  data A function that says whether its argument is odd.

ok  program A program whose execution does nothing and takes no time.

ord = real, char, bin, {all, *ord}, [ord]  data The ordered type, for which ∧ v < > ≤ ≥ are defined.

pi:= 3.141592653589793 (approximately)  constant The ratio of a circle's circumference to its diameter.

picture = [x^y[0,..,z]]  data where x is the number of pixels in the horizontal dimension, y is the number in the vertical dimension, and z is the number of pixel values.

plain: text→text  data Same text but not italic, not bold, and not underlined.

point  data A function that applies to a list and gives its deep domain (a bunch of strings of indexes).

It is a signal to the implementation that the strings in it will be used only as indexes to the list. It can therefore be implemented as a memory address (pointer).

pre: char→char  constant The character predecessor function.

printer?  text! “”  channel To the printer, it is an input channel; to all other programs, it is an output channel.

ProTem: text  constant This document.

quote: all→text  data produces a text representation of its argument. The argument is evaluated before quoting. See unquote and realtext .

rand \ dictionary containing three definitions.

next  program Assigns a hidden variable to the next value in a random sequence.

Int: int→int→int  data A function that is dependent on a hidden variable, and is reasonably uniform over the interval from (including) the first argument to (excluding) the second argument.

Real: real→real→real  data A function that is dependent on a hidden variable, and is reasonably uniform over the interval between the arguments.

rat = int/(nat+1)  data The rational numbers.

re: com→real  data The real part of a complex number.

real  data The real numbers including ∞ and −∞.

realtext: nat→nat→(nat+1)→real→text  data Format a real number. realtext d e w r  is a text representing real r with the final digit rounded. d is the number of digits after the decimal point; if d=0 the point is omitted. e is the number of digits in the exponent; if e>0 the decimal point will be placed after the first significant digit; if e=0 the ^^ is omitted and the decimal point will be placed as necessary. w is the total width; if w is greater than necessary, leading blanks are added; if w is less than necessary, the text contains blobs.

realtext 4 1 10 pi = “3.1416^^0”  realtext 2 0 6 (−pi) = “–3.14”

realtext 0 0 3 5 = “5”  realtext 0 0 3 (−5) = “−5”  realtext 0 0 2 123 = “••”

See textreal and quote .

round: real→int  data r–0.5 ≤ round r < r+0.5

s unit A time of one second.

screen?  text! “”  channel To the screen, it is an input channel; to all other programs, it is an output
channel.

session: text data The join of all texts from channel keys since the start of a session.

sign: real → (–1, 0, 1) data The sign of a real number.

silence: sound data The silent sound.

sin: real → §{r: real → –1 ≤ r ≤ +1} data The trigonometric sine function.

sinh: com→com data The hyperbolic sine function.

small: char constant The English small letters. See capital and letter .

sort: («ord)→«ord data Sorts in nondecreasing order.

sound data The sounds.

speaker? *sound! silence channel To the speaker, it is an input channel; to all other programs, it is an output channel.

sqrt: com→com data The principal square root. 4^(1/2) = 2, –2 but sqrt 4 = 2 .

stop program Its execution does nothing and takes forever so that no computation can follow.

sub: all→nat→nat→all→all data sub s n m t is a string formed from string s by replacing the substring from index n to index m with string t . The substring being replaced s_(n..m) does not have to be the same length as the string t replacing it. If n=m this is insertion. If t=nil this is deletion. sub s n m t = s_(0..;n); t; s_(m;.. «s)

subst: all→all→all→all data subst s x y is a string formed from string s by replacing all occurrences of item x with item y .

suc: char→char constant The character successor function.

tab: char constant The tab character.

tan: real→real data The trigonometric tangent function.

tanh: com→com data The hyperbolic tangent function.

text = *char data The character strings.

textreal: text→real data If the argument represents a real number, possibly preceded by space, tab, and new line characters, possibly followed by space, tab, and new line characters, the result is the represented number. Otherwise the result is –∞ . See realtext , unquote .

texttime: text→(realxs) data If the argument represents a time, possibly preceded by space, tab, and new line characters, possibly followed by space, tab, and new line characters, the result is the represented time in seconds since 2000 January 1 at 0:00 UTC (the midnight that begins 2000 January 1 at longitude 0). Times before then are negative. For example, texttime “1947 September 16 at 14:24:32.5 UTC–5” = –68675727.5s . Otherwise the result is –∞xs . See texttime .

time: realxs data A data name whose value depends on a clock. It gives the current time in seconds since 2000 January 1 at 0:00 UTC (the midnight that begins 2000 January 1 at longitude 0).

timetext: (realxs)→(–12,..12)→text data Given the time in seconds since 2000 January 1 at 0:00 UTC (the midnight that begins 2000 January 1 at longitude 0), and a time zone, the result is a readable text. Times before then are negative. See texttime . For example, timetext (–68675727.5s) (–5) = “1947 September 16 at 14:24:32.5 UTC–5”

trim: text→text data A text formed from the argument by removing all leading and trailing space, tab, and new line characters.

true:= ⊤ constant A binary value.

underline: text→text data Same text but underlined. Any part of the text that is already underlined is underlined again.

unquote: text→all data If the argument represents a ProTem data expression, the result is the value of the represented data. The argument is evaluated after un quoting. If the argument does not represent a ProTem data expression, the result is not defined. See quote , textreal .

wait program A plan with one constant parameter of type realxs . If the argument is nonnegative, its execution does nothing and takes the time given by the argument. If the argument is nonpositive, its execution does nothing and takes no time. See await and time and s .
**Miscellaneous**

The ProTem equivalent of enumerated type is shown here.

```protem
new color:= “red”, “green”, “blue”.
new brush: color:= “red”
```

The ProTem equivalent of the record type (structure type) is as follows.

```protem
new person:= “name” → text | “age” → nat.
new p: person:= “name” → “Josh” | “age” → 16
```

The fields of `p` can be selected by data that evaluates to text, for example

```
p “name” is the text “Josh”. The value of `p` can be changed using a function arrow and selective union.
p := “name” → “Amanda” | “age” → 2
```

We can even have a whole file (string) of records and join new records onto its end.

```protem
new file: *person:= nil.
file:= file; p
```

The efficiency of pointers is obtained through the use of the predefined function `point`. When applied to a list argument, it yields the deep domain of the list. For example,

```
point [10; [11; 12]; 13] = 0, 1;(0, 1), 2 = 0, 1;0, 1;1, 2
```

The use of `point` is a signal to the implementation that its strings of natural numbers will be used only as indexes into the list (and the implementation will check that this is so). For example, we can define a linked list `G` as follows.

```protem
new G: [* (“name” → text | “next” → point G)]:= [“name” → end | “next” → 0].
new first: point G:= 0
```

We can add a constant to `first` or subtract a constant from it, for example

```
first:= first+1
```

and similarly for the “next” field of each record of `G`. We can join it, for example

```
first:= first; 0
```

But we cannot multiply it, or use it in any way other than as indexes into `G`, for example

```
first:= G@first “next”.
G:= first → (“name” → “Aaron” | “next” → first) | G
```

With this limited use, the implementation of these indexes can be memory addresses. This way we obtain all the performance benefits of pointers without destroying the logic of our language.

The previous example, with linked list `G`, does not show the full generality of `point`. Here is a tree-structured example.

```protem
new t: tree (nil, tree; [all]; tree):= nil.
new p: point t:= nil
```

To move `p` down to the left in the tree we reassign it this way:

```
p:= p; 0
```

To move it down to the right, reassign it this way:

```
p:= p; 2
```

Thus `p` is a string of indexes indicating a subtree `t@p` of `t`. We can replace this subtree with tree `s` using the assignment

```
t:= p → s | t
```

We can express the information at the node indicated by `p` as

```
t@p 1 or t@(p; 1)
```

and we can replace the information at this node with the integer 6 using the assignment
We can feed the output from the left side of | "catches" the exception and "throws" it to the right side, where it is "handled".

Then if i fails to be an integer in the range 0..7 in the expression (weekday | all→“domain error”) i
the left side of | "catches" the exception and "throws" it to the right side, where it is "handled".

Input choice, as in CSP, can be obtained as follows.

In the persistent scope, ProTem functions as an operating system, where programs are executed as soon as they are entered. Unix directories are dictionaries. Unix pipes is obtained by channel parameters. For example, suppose trimmer is a plan to trim off leading and following blanks and tabs from lines of text, and italicize is a plan to italicize texts.

new trimmer [plan in? text [plan out! text
[loop [in? text; nl. end.
   if in?=end [out! end] else [out! trim (in?); nl. loop]]]]]

new italicize [plan in? text [plan out! text
[loop [in? text; nl. end.
   if in?=end [out! end] else [out! italic (in?); nl. loop]]]]]

We can feed the output from trimmer to the input of italicize by defining a channel for the
purpose. If the original input comes from keys, and the final output goes to screen, then

    new pipe? text! "". trimmer keys pipe. italicize pipe screen. old pipe

Even better:

    new pipe? text! "". trimmer keys pipe \ italicize pipe screen. old pipe

If italicize wants input before it is available from trimmer, italicize waits. If trimmer sends output before italicize wants it, it is buffered.

Unix mail is ProTem's mail channel. If you are the creator of the definition of something in the persistent scope, and you want to send it to someone for them to make changes, then

    mail! “To: someone@address.domain”; definition “something”

(see predefined definition). When someone sends back the changed definition, receive it, delete your old definition, and then redefine it (see predefined exec) by

    mail? text; end. old something. exec (“new ”; mail?)

An implementation may provide plans for a variety of languages. For example, it may provide a plan named Python, with one text parameter, whose execution is that of the Python fragment represented by the argument. It may provide asm, whose execution is that of the assembly-language program represented by the argument.

ProTem considers object orientation to be a programming style, rather than a programming-language style, or collection of language features. Object-oriented programming (as a style of programming) can be done in ProTem. Data structures, and the functions and procedures that access and update them, can be defined together in one dictionary. If many “objects” of the same type are wanted, new dictionaries just like old ones are easily defined using \ (see parseStack in Dictionary Definition).

To execute a program stored on someone else's computer, just invoke that remote program using its full address (computername and programname). For efficiency, it might be best to compile that remote program for your own computer and run it locally. Any nonlocal names (variables, channels, and so on) refer to entities on the computer where the program is compiled.

I would prefer to use the symbols ↑ and ↓ for maximum and minimum as in unified algebra, rather than ∨ and ∧. But ↑ and ↓ are not the tradition for binary disjunction and conjunction, nor for set union and intersection, and there are no good keyboard substitutes for them. As it is, the ordering symbols ∨ ∧ < > ≤ ≥ are nicely related.

In ProTem, a text is written with quotation marks. An attractive alternative to quotation marks is underlining. For example, the text “abc” would instead be written abc. It is a good-looking syntax, but awkward to key in, so quotation marks could be the keyboard substitute. (The empty text would be nil, string indexing would not be underscore, and quote and unquote would be renamed textify and untextify.) I have used underlining in ProTem just for quotation marks within a text; for example, “Just say “no”.”. This would become Just say “no”, with no exception needed for quotation marks within a text. In ProTem, if you want an underlined quotation mark within a text, you have to underline it again. And so on. Thus every character can occur within a text. In the alternative notation, no special rule is needed. Underlining presents a theoretical (but not practical) limitation: we cannot write a self-reproducing text (try). We can write a self-reproducing text with quotation marks, repeating the left and right quotation marks as characters within a text. For example, “Just say ““no”“.”. Using this convention, here is a self-reproducing text:

    ""_ (0;0;(0..32);31;31;(1;..31))""_ (0;0;(0..32);31;31;(1;..31))

Perform the indexing to see what you get.
In the Presentation Grammar, the case-program has the form
\[
\text{case data } \llbracket \text{ program } \rrbracket
\]
with the program form
\[
\text{program } \llbracket \text{ program } \rrbracket \llbracket \text{ program } \rrbracket
\]
to make more cases. Although it is correct, it is not a clear presentation of
\[
\text{case data } \llbracket \text{ program } \rrbracket \llbracket \text{ program } \rrbracket \llbracket \text{ program } \rrbracket
\]
and so on. A clear presentation requires another nonterminal, as in the LL(1) Grammar and LR(0)
Grammar. My principle is that the Presentation Grammar has only two nonterminals (program and
data), but the case-program stretches that principle rather thin. Furthermore, the program form
\[
\text{program } \llbracket \text{ program } \rrbracket \llbracket \text{ program } \rrbracket
\]
is the only program or data form whose use is confined to a single context, and I am unhappy to have
any program or data form that cannot be used more generally.

Sounds and pictures are data structures. This part of ProTem is not yet designed. Perhaps a picture
is an element of \([x^*y^*(0..z)])\) where \(x\) is the number of pixels in the horizontal direction, \(y\) is
the number of pixels in the vertical direction, and \(z\) is the number of pixel values. A picture could
therefore be expressed in the same way as any other two-dimensional array, and one could refer to
the pixel in column \(3\) and row \(4\) of picture \(p\) as \(p\, 3\, 4\). Perhaps a movie is a string of pictures.
The operations on movies would be those of strings, such as substring and join. To help in the
creation of movies, one of the pixel values should be transparent, and one of the operations on
pictures should be overlaying. Predefined silence is a sound, and predefined sound is all sounds.
Sounds are input on channel microphone; pictures are input on channel camera. A constant can
be defined as a sound or picture. A variable can be assigned to a sound or picture. Sounds and
pictures can be included in a data structure, and manipulated using the operators on that data
structure. Sounds can be output on channel speaker. Channel screen must be modified so pictures
can be output on channel screen.

I am perfectly well aware that “data” is a Latin word; it is the plural gerund from the verb “dare”
which means “to give”, so the data are the givens. The singular is “datum”. But this is an English
document, and I have decided to use the word “data” for both singular and plural. I have also
decided to say “indexes” rather than “indices”.

**Intentionally Omitted Features**

Each of the following omitted features would be a small syntactic convenience, and would be easy to
add to the language. But they would make the language larger, and that is a cost. And they move
away from the form needed for verification. So they are not included in ProTem.

assertion
\[
\text{assert } x \leq y \quad \text{means} \quad \text{if } -(x \leq y) \llbracket \! \! \! \! \! \text{ ! “assert failure”. stop} \rrbracket
\]
name grouping
\[
\text{new } x, y : \text{int} := 0 \quad \text{means} \quad \text{new } x : \text{int} := 0 \parallel \text{new } y : \text{int} := 0
\]
\[
\text{old } x, y \quad \text{means} \quad \text{old } x \parallel \text{old } y
\]
\[
x, y := 0 \quad \text{means} \quad x := 0 \parallel y := 0
\]
\[
(a, b : \text{nat} \rightarrow a+b) \quad \text{means} \quad (a : \text{nat} \rightarrow (b : \text{nat} \rightarrow a+b))
\]
\[
\text{plan } a, b : \text{nat} \llbracket x := a+b \rrbracket \quad \text{means} \quad \text{plan } a := \text{nat} [\text{plan } b := \text{nat} [x := a+b]]
\]
item assignment
\[
S_{3} := 5 \quad \text{means} \quad S := S \llbracket 3 \rightarrow 5 \rrbracket
\]
\[
L_{3} := 5 \quad \text{means} \quad L := 3 \rightarrow 5 \parallel L
\]
\[
A_{3\, 4} := 5 \quad \text{means} \quad A := (3;4) \rightarrow 5 \parallel A
\]
loop constructs

\[
\text{while } n > 0 \ [n := n - 1] \quad \text{means} \quad \text{while } [\text{if } n > 0 \ [n := n - 1. \ \text{while}]]
\]
\[
\text{repeat } [n := n - 1] \ n = 0 \quad \text{means} \quad \text{repeat } [n := n - 1. \ \text{if } -(n = 0) \ [\text{repeat}]]
\]
\[
\text{loop } [n := n - 1. \ \text{if } n = 0 \ [\text{exit 1}. \ \text{loop} \ [m := m - 1. \ \text{if } -(m = 0) \ [\text{loop}]]]]
\]
\[
\text{means} \quad \text{loop } [n := n - 1. \ \text{if } -(n = 0) \ [m := m - 1. \ \text{if } -(m = 0) \ [\text{loop}]]]
\]

return from named program

\[
\text{name } [\text{if } n = 0 \ [\text{return name}. \ \text{more}]] \quad \text{means} \quad \text{name } [\text{if } -(n = 0) \ [\text{more}]]
\]

The assignment \[ L := 3 \rightarrow 5 \] \( L \) should be compiled the same as \( L := 5 \) would be if it were included in ProTem; the list \( L \) should not be copied. The same for string item assignment. In the loop
\[
\text{while } [\text{if } n > 0 \ [n := n - 1. \ \text{while}]]
\]

the last-action (tail recursive) call should be compiled as a branch instruction, with no stack activity, the same as a \text{while}-loop would be if it were included. The same for other loop constructs. Omitting item assignment and looping constructs should not cost execution time or space.

Plans with parameters and arguments of type program

\[
\text{plan } \text{simplename } [\text{program } ] \quad \text{plan, parameter is program}
\]
\[
\text{program program plan, program argument}
\]

were considered and rejected due to syntactic and semantic ambiguities.

As a counterpart to the Unix cd command, I considered

\[
\text{open } \text{dictionaryname}
\]
\[
\text{close } \text{dictionaryname}
\]

to allow names in that dictionary to be referred to without stating the dictionary. For example, if we have dictionary \( abc \), and within it names \( x \) and \( y \), we refer to these names as \( abc\backslash x \) and \( abc\backslash y \). By saying \text{open} \( abc \) we can then refer to them as just \( x \) and \( y \). But the interaction between \text{open} and scope is complex, we can already refer to names within \text{predefined}, and we can shorten names by synonym definition, so I left out \text{open} and \text{close}.

The syntax

\[
\text{new \abc\def}
\]
defines a new dictionary \( \abc \) populated with the definitions from existing dictionary \( \def \). Perhaps it should instead add the definitions of existing dictionary \( \def \) to existing dictionary \( \abc \). At present, there is no way to do that. If I were to make this change, there would still be a way to define a new dictionary populated with definitions from an existing dictionary:

\[
\text{new \abc\}. \ \text{new \abc\def}
\]

But I would have to decide how to resolve duplicate definitions. I am guessing that we always want to populate a new dictionary, not further populate an existing one.

There is no frame construct in ProTem, but Ctrl-c serves a similar purpose. In some languages there is a module or object construct for the purpose of grouping together related definitions. In ProTem, dictionaries serve that purpose.

In order to allow sharing of variables in the persistent scope, I created a personal identity type and a scheme of permits to say who can use and change what. The scheme was more flexible than Unix chmod, and I was quite pleased with it. But with sharing, all mathematical reasoning is invalid; \( x=x \) might evaluate to \( \bot \). And with sharing, locks are required so that assignments do not interfere with uses, and assignments do not interfere with other assignments. Sharing problems are solved by the use of channels: anyone can email anything to anyone else; any program can send anything to any other program on a channel. So, even though it hurts to leave out something I worked hard on, identities and shared variables are intentionally omitted.
On input, do we ever want the correcting pattern without an echo? For a password, we usually want to output some character, like •, for each key press, and program the corrections, rather than using the correcting pattern (see Read Password). If there is a reason anyone wants the correcting pattern without an echo, we can add it, but then we will have to use a symbol, perhaps _, for the correcting pattern. Input with correcting pattern and no echo would be ?_ and input with correcting pattern and echo would be ?_! . I believe input with correcting pattern but without echo is rarely wanted, perhaps never wanted, and so it is not worth lengthening the input forms that are frequently wanted.

Two binary operators ∆ (nand) and ∇ (nor) are missing. They are not wanted very often, there are no good keyboard substitutes for them, and they are easily synthesized:

\[ x \Delta y = \neg(x \land y) \]
\[ x \nabla y = \neg(x \lor y) \]

The empty program, denoted by nothing, whose execution does nothing and takes no time, is easy to add to ProTem. With it, we can almost always consider the period ( . ) to be a program terminator, rather than an infix connective. But there is one context where it causes confusion.

\[ A. || B \]
looks like program A (terminated with a period) is in parallel with program B (terminated with a period). But, following A, the empty program is in parallel with B, as in

\[ A. [ || B ] \]

So I have not included the syntactically empty program in ProTem, providing instead the predefined program ok whose execution does nothing and takes no time (as in aPToP).

**Implementation Philosophy**

An implementation checks whether the text presented to it represents a program, and issues an error message if it does not. Using \[ \text{ctl } x \], that check includes determining whether every variable assignment is to a value that is included in the type of the variable. That determination is most helpful if it can be made before execution; but if not, it is still helpful if it can be made during an execution attempt. When we are satisfied that all assignments are always to values included in their types, \[ \text{ctl } \neg \] does not check types, resulting in faster execution.

Some expressions do not need to be evaluated. For examples,

\[ !\langle r:\text{ rat} \to 5 \rangle (1/0) \quad \text{`should print 5} \]
\[ !1/0 = 1/0 \quad \text{`should print } \top \]
\[ ![0; 1] 2 = ![0; 1] 2 \quad \text{`should print } \top \]

There are also expressions that cannot or should not be evaluated further. For example,

\[ !-3 \quad \text{`prints } -3 \]

ProTem does not evaluate the application of the negation operator – to the operand 3 (see Number Representation); it just prints the operator and operand. Similarly

\[ !1/0 \quad \text{`should print 1/0} \]
\[ ![0; 1] 2 \quad \text{`should print } [0; 1] 2 \]

Due to the difficulty of implementation, it is permissible for an implementation to behave differently.

No programming language has ever been, nor will ever be, implemented entirely. Every programming language is infinite; every implementation is finite. There is always a program too big for the implementation. There is a multitude of size limitations: the parse stack might overflow, the dictionary (symbol table) might be too small, the forward branch fixup list might be exceeded, and so on. It would be ugly to define a programming language by listing all the size limitations of programs. And it would be counter-productive because it would exclude implementations that can accommodate larger programs.
Whenever a program exceeds a size limitation, the implementation should not say “Error: limitation exceeded.”, because the program is not in error. The implementation should say “Apology: this implementation is too limited to accommodate your program.”. An “error” message tells a programmer to correct the error; there is no other option. An “apology” message gives the programmer 3 options: change the program to live within the limitation; change the implementation options to increase the limit that was exceeded; take the program to a different implementation.

Natural numbers and integers are usually limited to those that are representable in a specific number of bits, for example, 32 bits. This is a size limitation, just the same as other size limitations. It is more complicated and uglier to define arithmetic within finite limitations than to define the naturals and the integers. And it is counter-productive to do so, because it excludes an implementation with 64-bit arithmetic. As with other implementation limitations, numeric overflow should not get an “error” message; it should get an “apology” message.

Floating-point numbers and arithmetic should never be offered as a language feature. The programmer wants rational or real numbers and arithmetic, but may be willing to accept the floating-point approximation for the sake of efficiency. Floating-point, with a specific number of bits, is an implementation limitation. Any alternative to floating-point that increases the accuracy without taking too much time or space should be welcome.

ProTem is a rich programming system, offering many kinds of data and operators on data, and many ways to structure a computation. Some features may be difficult to implement. And some features may be of little use to most programmers. It may be a wise decision not to implement some features. For example, an implementer might decide that in a variable definition, the type must be one of

\[
\text{int} \ \text{rat} \ \text{bin} \ \text{text} \ [n]\text{type}
\]

where \( n \) is a natural number and \( \text{type} \) is any of these types just listed. An implementer may decide not to implement concurrent execution. No-one can complain that the complete language is not implemented, since it is impossible to completely implement any language. But ProTem is defined to allow all type expressions that make sense, and to allow concurrency, so the next implementation can accommodate programs that previous implementations could not accommodate.

**Example Programs**

**Portation Simulation**

```
\textbf{new simport} ` a program to simulate \textbf{portation}
\[
\begin{array}{c}
\ `\text{input: keys} \\
\ `\text{output: screen} \\
\ `\text{use: ceil m nat nil nl point real s sqrt time} \\
\ `\text{call: await stop} \\
\ `\text{refer: rand} \\
\end{array}
\]

` Distance between control boxes is always 1 m.
` Merges do not overlap, so there is at most 1 corresponding box on the merging portway.
` Each divergence has a left branch and a right branch: there is no straight.
` Leading to a divergence, boxes record only one square speed.

` start of definitions

\textbf{new km:=} 1000\times m. \textbf{new h:=} 60\times 60\times s. ` kilometer and hour
```
new maxaccel := 1.5 m/s^2. ` maximum deceleration = -maxaccel
new speedlimit := 60 km/h. ` speed limit is 60 km/h everywhere
new cushion := 1 s. ` reaction time for all porters
new impatience := 10 s. ` acceleration factor
new maxdistance := ceil(speedlimit^2 / (2 * maxaccel)). ` max search distance ahead
new numporters := 120. new numboxes := 7480.
new visualDelayTime := 0.5 s. ` for human viewing
new porter. ` so porter can be referenced before it is defined

new box: [numboxes * ("ahead left", "ahead right", "behind left", "behind right") → point box
| "beside" → point box
| "above" → (point porter, numporters)
| ("horizontal", "vertical") → nat }] ` box position on screen
:= [numboxes * ("ahead left", "ahead right", "behind left", "behind right") → 0
| "beside" → 0
| "above" → numporters ` indicates no porter above
| ("horizontal", "vertical") → 0 ].

new porter: [numporters * ("below" → point box ` what is beneath
| "arrival time" → (realxs) ` arrival time at this box
| "speed" → (realsm/s))] ` current speed
:= [numporters * ("below" → 0
| "arrival time" → (0 s)
| "speed" → (0 m/s)]].

new draw [plan b := nat [plan c := "grey", "blue", "red" [UNFINISHED]]]. `end of draw
` draws a box at screen position (box b "horizontal") (box b "vertical") of color c.
` "grey" means no porter present, "blue" means porter present, "red" means crash
UNFINISHED because graphical output has not yet been designed

` end of definitions, start of initialization

for b := 0;..numboxes
!! "What box is ahead-left of box "; b; "? ". ?!
box := (b; "ahead left") → ? | (?; "behind left") → b | box.
! "What box is ahead-right of box "; b; "? ". ?!
box := (b; "ahead right") → ? | (?; "behind right") → b | box.
! "What box is beside box "; b; "? ". ?!
box := (b; "beside") → ? | box.
! "What are the horizontal and vertical coordinates of box "; b; "? ". ?!
box := (b; "horizontal") → ? | box.
?!. box := (b; "vertical") → ? | box.
! draw b "grey"[\]. ` default color; may be changed below

for p := 0;..numporters
!! "Porter "; p; " is over what box? ". ?!
porter := (p; "below") → ? | porter. box := (?; "above") → p | box.
! draw (?) "blue"[\].
end of initialization, start of simulation

infiniteLoop
[ new iterationStartTime:= time. ` the time of the start of each iteration of the infiniteLoop

new p: point porter:= 0. ` p:= the porter that arrived at its current position first
new t: realsx:= ∞xs. ` t is a time, initially an infinite time
for q:= 0..numporters [if porter q “arrival time” < t [t:= porter q “arrival time”. p:= q]].
old t.

new b:= porter p “below”. ` the box below porter p
new bb:= box b “beside”. ` the box beside b; if none then bb=b
new boxesToDo: *[point box; natxms]:= nil.
  ` queue of boxes to be explored; their distances ahead of porter p
  ` queue is sorted by increasing distance ahead
  ` difference between any two distances in the queue is at most 1

` initialize boxesToDo
if bb = b [boxesToDo:= nil]
else [if box bb “above” = numporters [boxesToDo:= nil]]
  else [if porter (box bb “above”) “speed” < porter p “speed” [boxesToDo:= nil]]
    else [boxesToDo:= [bb; 0×m]]].
boxesToDo:= boxesToDo; [box b “ahead left”; 1×m].
if box b “ahead left” ≠ box b “ahead right” [boxesToDo:= boxesToDo; [box b “ahead right”; 1×m]].
old b. old bb.

new accel: realxmsls:= maxaccel. ` acceleration for porter p

` using boxesToDo calculate accel for porter p

nextBox [new b:= (boxesToDo_0) 0. ` the box we are looking at
new d:= (boxesToDo_0) 1. ` its distance ahead of porter p
boxesToDo:= boxesToDo _(1..↔boxesToDo).
if d≤maxdistance
  [new desiredspeed ` according to porter pa
  @(pa: (point porter, numporters) →
    pa=numporters ⊨ speedlimit
  =1 (  sqrt (  porter pa “speed”^2 + 2×maxaccel×d
          + (maxaccel×cushion)^2 )
        − maxaccel×cushion) ∧ speedlimit)].
accel:= ( ( ( desiredspeed (box b “above”)
        ∧ desiredspeed (porter (box b “beside”) “above”))
    − porter p “speed”)
  × impatience)
  ∨ − maxaccel) ∧ maxaccel.
if box b “above” = numporters = porter (box b “beside”) “above”
  [ ` add boxes ahead to queue and continue
    boxesToDo:= boxesToDo; [box b “ahead left”; d+1×m].
  if box b “ahead left” ≠ box b “ahead right”
\[\text{boxesToDo} := \text{boxesToDo}; \ [\text{box } b \ 	ext{“ahead right”}; \ d+1 \times m]\].
nextBox
 else \[\text{if } \leftrightarrow \text{boxesToDo} > 0 \ [\text{nextBox}]\].

old boxesToDo.

` using accel, move porter \(p\) ahead one box
new \(b\): point \(box := \text{porter } p \ 	ext{“below”}\).
\(\text{box}_{\text{next}} := b \text{ box} (\text{rand} \times 0 \ 2 = 0 \implies \text{“ahead left”} \equiv \text{“ahead right”})\).
\(\text{if } b \text{ “porter”} < \text{numporters} \ [\text{draw } b \ 	ext{“red”}. \ \text{stop}]. \ \text{` crash}\)
\(\text{porter} := (p; \ 	ext{“below”}) \rightarrow \ b \text{ box } \). \(\text{draw } b \ 	ext{“blue”}.\)

old \(b\).

new \(\text{speed} := \sqrt{(\text{porter } p \ 	ext{“speed”}^2 + 2 \times \text{accel} \times m)} \wedge \text{speedlimit} \).
\(\text{ Porter } := (p; \ 	ext{“arrival time”}) \rightarrow \ (\text{porter } p \ 	ext{“arrival time”} + 2 \times m/(\text{porter } p \ 	ext{“speed”} + \text{speed}))\)
\(\text{ Porter } := (p; \ “speed”\) \rightarrow \text{speed}
\(\text{porter} .\)

old \(\text{speed}\). old \(\text{accel}\). old \(p\). ` these olds are not really necessary

\(\text{await (iterationStartTime} + \text{visualDelayTime}).\)
infiniteLoop ```

```

Quote Notation Lengths

` program to compare quote notation lengths with numerator/denominator lengths

` output: screen
` use: bin div even nat odd

new shl \(\langle n: \text{nat} \rightarrow \langle m: \text{nat} \rightarrow \text{shl } n \text{ left } m \text{ places}; \ n \times 2^m \rangle \rangle\).
\text{ value } r: \text{nat} := n \ [\text{for } i:= 0,..m \ [r:= r \times 2]]\)\).

new shr \(\langle n: \text{nat} \rightarrow \langle m: \text{nat} \rightarrow \text{shl } n \text{ right } m \text{ places}; \ \text{floor } (n \times 2^m \rangle \) \text{ or } \text{div } n (2^m)\).
\text{ value } r: \text{nat} := n \ [\text{for } i:= 0,..m \ [r:= \text{div } r \ 2]]\)\).

new gcd \(\langle a: (n+1) \rightarrow \langle b: (n+1) \rightarrow \text{gcd } a \text{ and } b\).
\(a = b \equiv a = 1 \land a < b \equiv \text{gcd } a \ (b-a) = \equiv \text{gcd } (a-b) b\rangle\).

new norm \[\text{plan } \text{num} := \text{nat}+1 \ \text{plan } \text{denom} := \text{nat}+1 \ \text{` normalize } \text{num/denom}
[\text{new } g := \text{gcd } \text{num } \text{denom}. \ \text{num} := \text{num}/g. \ \text{denom} := \text{denom}/g]]\).

new count: \text{nat} := 0. ` number of examples
new qlen: \text{nat} := 0. ` total length of quote representations
new rlen: \text{nat} := 0. ` total length of numerator/denominator representations

\text{for } \text{length} := 1,..15
\[\text{for } \text{string} := 0,..(\text{shl } 1 \text{ length}) \ ` each string of that length
\[\text{for } \text{quote} := 0,..\text{length} \ ` each quote position (at least one bit to left of quote)
\[\text{if } \text{even} (\text{shl } \text{string} (\text{length}-1)) \neq \text{even} (\text{shl } \text{string} (\text{quote}-1)) \ ` \text{roll-normalized}
\[\text{if } \ ` \text{repeat-normalized}
value repeatnorm: bin := ⊤

[new len: nat := div (length−quote) 2, \ the length of the possibly repeating part\]
trythislen [if len>0 1 ≤ len ≤ (length−quote)/2
  [new extract \(\langle i: nat \rightarrow \langle l: nat \rightarrow \langle \text{shrink } i \text{ – } \text{shift } (\text{shrink } (i+l) \text{ l} \rangle \rangle \rangle\).
new ex := extract quote len.
if \(\text{the negative part is a repetition (twice or more) of ex}\)
value b: bin := ⊤
[new i: nat := quote+len, i+len ≤ length
iloop [new ey := extract i len.
  if ex=ey \[i := i+len, \text{islength}\]
    if i+len ≤ length [iloop]
  else \[b := ⊥\]]
else \[b := ⊥\]]
][repeatnorm := ⊥\]
else \[len := len−1, \text{trythislen}\]]]]
[for point := 0;..length+1 \each point position (right end, interior, left end)\]
[if \(\text{the rightmost bit is 1 or it is to the left of quote or point}\)
odd string ν (quote=0) ν (point=0)
][\text{convert to numerator/denominator}\]
new num: nat := shl string (length−quote) − string − shr (shr string quote) length.
if num<0 \[num := −num\].
new denom: nat := shr (shr 1 (length−quote) − 1) point.
\text{update statistics}\n\text{count} := count+1, qlen := qlen+length.
rlen := rlen+1. \text{for the sign}\nloop \[num := div num 2, rlen := rlen+1.\]
if num>0 [loop]].
loop \[denom := div denom 2, rlen := rlen+1.\]
if denom>0 [loop]]]]]]]]]

! “In ”; count; “ examples, quote average length = ”;
qlen/count; “, num/denom average length = ”; rlen/count.

old shl. old shr. old gcd. old norm. old count. old qlen. old rlen

Minimum Redundancy Codes

new MRC \` \text{a program to compute minimum redundancy prefix codes (Huffman codes)\}
[\text{input: keys}\]
`output: screen
`use: back find nat nil nl point real text textreal

new forest: *[real; tree (\text{[text], [tree; tree]]):= nil. \text{The data structure is a string of lists.}\]
\Each list is a pair of real and tree; the real is the frequency of the tree.
\Each tree is binary with texts at the leaves.

inputstart
[! “Enter a frequency, then a colon, then a message, then a new line, and repeat. ”;
  “To end, just enter a new line.”; nl.
readloop
if ↔? = 0 ` Just new line was pressed.
[if ↔? = 0 `We have not had any input yet. We need at least one.
[! “Insufficient input. Try again.”. inputstart]

new c:= find ":" (?).
if c = ↔? [! “Bad format: no colon. Try again.”. readloop]
new freq:= textreal (?_(0;..c)).
if freq=–∞ [! “Bad frequency format. Try again.”. readloop]
new message:= ?_(c+1;..↔?).
` find where the new data goes in forest and put it there.
new i: nat:= 0. ` A new tree must be moved to position here or later.
loop [if ↔? forest ≥ 2
[ ` combine the first two trees into a new tree t
new t:= [(forest_0)0 + (forest_1)0; [(forest_0)1 ; (forest_1)1]].
` remove those first two trees from the forest
forest:= forest_(2;..↔?forest).
` put tree t into its place in the forest
innerloop [if here = ↔? forest v (t 0 < (forest_here)0)` we have found where it goes
[forest:= forest_(0;..here); t; forest_(here;..↔?forest). loop]
else [here:= here+1. innerloop]]].
` forest is now a nonempty string of pairs, each pair consisting of a frequency and a tree, each
` tree is a single leaf, each leaf is a list-text. They are in non-decreasing frequency order.
` For example: [3; [“c”]]; [4; [“a”]]; [9; [“f”]]; [12; [“b”]]; [15; [“e”]]; [20; [“d”]]

new here: nat:= 0. ` A new tree must be moved to position here or later.
loop [if ↔? forest ≥ 2
[ ` combine the first two trees into a new tree t
new t:= [(forest_0)0 + (forest_1)0; [(forest_0)1 ; (forest_1)1]].
` remove those first two trees from the forest
forest:= forest_(2;..↔?forest).
` put tree t into its place in the forest
innerloop [if here = ↔? forest v (t 0 < (forest_here)0)` we have found where it goes
[forest:= forest_(0;..here); t; forest_(here;..↔?forest). loop]
else [here:= here+1. innerloop]]].
` forest is now a single pair consisting of the total of all frequencies and a code tree.
new t:= forest_1. ` the code tree
` Walk the tree, depth-first, printing leaves and their codes
new p: point t:= nil. ` a path within t starting at the root
new pt: text:= “”. ` same path as p but as a text for printing
loop [if ~(t p): text ` we are at a leaf
[! “code”; pt; “; message”; ~(t p); nl]

Read Password

` program to read a password, allowing corrections, displaying blobs
` input: keys
` output: screen
` use: char delete nl text
new password: text:= “”.
pswd [!“Please enter password followed by return: ”.
read [?] char.
if ?=nl [if password="" ["Empty password. Try again."]; nl. pswd]
else [!nl]]
else [if ?=delete [if password≠""] [password:= password_(0;..↔password–1). !delete]]
else [password:= password; ?. !“•”].

Grammars

LL(1) Grammar

In this grammar, for each nonterminal, every production except possibly the last begins with a different terminal. Director sets are needed to create the parser, but they are not needed for parsing, and that is a special case of LL(1) that deserves its own name; perhaps LL(1/2). To parse a program, the parse stack begins with only the program nonterminal on it, and ends empty with no more input. A name control program is responsible for classifying names. For efficiency, the productions (except possibly the last) for each nonterminal should be placed in order of frequency. The following nonterminals can be eliminated by replacing them with their one production: program sequent data data6 data5 data4 data3 data1. This leaves the grammar with 33–8 = 25 nonterminals.

program sequent moresequents
moresequents . program empty
sequent phrase parallelphrases
parallelphrases || sequent empty
phrase new simplename afternewname
old simplename compounder
[ program ]
if data [ program ] elsepart
case data [ program ] morecases elsepart
for simplename := data [ program ]
plan simplename parameterkind [ program ] arguments
! data
? inputafterq
simplename aftersimplename

afternewname : data := data
( data ) := data
? data ! data
[ program ] \afterbackslash
#1
simplename compounder
empty
afterbackslash simplename afternewname
\ simplename compounder
empty

compounder \ simplename compounder
empty

elsepart else [ program ]
empty

morecases [ program ] morecases
empty

parameterkind := data
: data
! data
? data
\

aftersimplename [ program ]
compounder aftername

aftername := data
! data
? inputafterq
arguments

inputafterq ! echo
data afterpattern

afterpattern ! echo
empty

echo simplename compounder
empty

arguments number arguments
∞ arguments
text arguments
⊤ arguments
⊥ arguments
value simplename : data := data [ program ] arguments
{ data } arguments
[ data ] arguments
( data ) arguments
⟨ simplename : data1 → data ⟩ arguments
simplename specificand arguments
empty

data data6 moredata
moredata \implies \data = \data 
empty

data6 \data5 \moredata6

\moredata6 \data5 \moredata6
\neq \data5 \moredata6
\lt \data5 \moredata6
\gt \data5 \moredata6
\leq \data5 \moredata6
\geq \data5 \moredata6
:\data5 \moredata6
:: \data5 \moredata6
\in \data5 \moredata6
empty

data5 \data4 \moredata5

\moredata5 \data4 \moredata5
\ldots \data4 \moredata5
| \data4 \moredata5
\triangleleft \data4 \moredata5
\triangleright \data4 \moredata5
empty

data4 \data3 \moredata4

\moredata4 \data3 \moredata4
\ldots \data3 \moredata4
\ldots \data3 \moredata4
\ldots \data3 \moredata4
\ldots \data3 \moredata4
\texttt{ printer } \data3 \moredata4
empty

data3 \data2 \moredata3

\moredata3 \data2 \moredata3
/ \data2 \moredata3
\land \data2 \moredata3
\lor \data2 \moredata3
empty

data2 \# \data2
\ldots \data2
\ldots \data2
+ \data2
\Box \data2
\forall \data2
\not \exists \data2
* \data2
\[ \varepsilon \text{ data2} \]
\[ \$ \text{ data2} \]
\[ \leftrightarrow \text{ data2} \]
\[ \text{data1} \text{ moredata2} \]

moredata2
\[ * \text{ data2} \text{ moredata2} \]
\[ \rightarrow \text{ data2} \text{ moredata2} \]
\[ ^\text{data2} \text{ moredata2} \]
\[ ^\text{data2} \text{ moredata2} \]
\[ \text{empty} \]

data1
\[ ? \text{ moredata1} \]
\[ ?? \text{ moredata1} \]
\[ \text{data0} \text{ moredata1} \]

moredata1
\[ \% \text{ moredata1} \]
\[ ? \text{ moredata1} \]
\[ ?? \text{ moredata1} \]
\[ _\text{data0} \text{ moredata1} \]
\[ @ \text{data0} \text{ moredata1} \]
\[ \& \text{data0} \text{ moredata1} \]
\[ \text{arguments} \]

data0
\[ \text{number} \]
\[ \infty \]
\[ \text{text} \]
\[ \top \]
\[ \bot \]
\[ \text{value} \text{ simplename} : \text{data} := \text{data} \llbracket \text{program} \rrbracket \]
\[ \{ \text{data} \} \]
\[ [ \text{data} ] \]
\[ ( \text{data} ) \]
\[ \langle \text{simplename} : \text{data1} \rightarrow \text{data} \rangle \]
\[ \text{simplename specificand} \]

specificand
\[ \langle \text{data} \rangle \]
\[ \text{comounder} \]

**LR(0) Grammar**

The following grammar has no reduce-reduce choices and no shift-reduce choices. It has shift-shift choices. Such a grammar is commonly called LR(0), but it should not be, because a shift action pushes an input symbol onto the parse stack, and therefore a shift action depends on the input symbol. It is a special case of LR(1) that deserves its own name, but not LR(0); perhaps LR(1/2).

To parse a program, the parse stack begins empty, and ends with only the program nonterminal on it and no more input. A name control program is responsible for classifying names.

program
\[ \text{sequent} \]
\[ \text{program . sequent} \]
sequent phrase
sequent || phrase

phrase

new name : data := data
new name := data
new name ( data )
new name [ program ]
new name ? data ! data
new name #1
new name \n
new name \ \ name
new name name
new name
old name
name := data
name ! data
! data
name ? data
? data
name ? data ! name
? data ! name
name ? data !
? data!
name ? ! name
? ! name
name ? !
? !
simplename [ program ]
if data [ program ]
if data [ program ] else [ program ]
case data cases
case data cases else [ program ]
for simplename := data [ program ]
[ program ]
plan

cases
cases [ program ]

plan
plan simplename := data [ program ]
plan simplename : data [ program ]
plan simplename ? data [ program ]
plan simplename ! data [ program ]
plan simplename \ [ program ]
plan data0
name

data
data6 ⊨ data ⫤ data
data6
data6

- data6 = data5
- data6 ≠ data5
- data6 < data5
- data6 > data5
- data6 ≤ data5
- data6 ≥ data5
- data6 : data5
- data6 :: data5
- data6 ∈ data5
- data5

data5

- data5 , data4
- data5 .. data4
- data5 ! data4
- data5 < data4
- data5 > data4
- data5 < data4
- data4

data4

- data4 ; data3
- data4 .. data3
- data4 ;; data3
- data4 * data3
- data4 + data3
- data4 – data3
- data3

data3

- data3 × data2
- data3 / data2
- data3 ∧ data2
- data3 ∨ data2
- data2

data2

- + data2
- - data2
- & data2
- $ data2
- ↔ data2
- # data2
- ~ data2
- □ data2
- $ data2
- * data2
- data1 * data2
- data1 → data2
- data1 ∧ data2
- data1 ^^ data2
- data1

data1

- data1 data0
- data1 _ data0
- data1 @ data0
data1 %
data1 & data0
name ?
name ??
data0

data0
number
∞
text
⊥
[ data ]
{ data }
( data )
⟨ simplename : data1 → data ⟩
value simplename : data := data [] program ]
?
??
name
simplename ⟨ data ⟩

name
simplename
name \ simplename

Acknowledgements

The first public mention of ProTem was
E.C.R.Hehner, T.S.Norvell: “ProTem: a Programming System”, University of Toronto,
Computer Systems Research Group, technical report CSRG213, 1988 September

Theo Norvell wrote an MSc thesis in 1988 titled “Expressions, Types, and Data Structures in
ProTem”. Jim Horning suggested error recovery using the session file. Hugh Redelmeier acted as
design consultant and critic in 1990. Brian Parkinson found a bug in the implementation in 1990,
and he wrote an MSc thesis in 1991 titled “Automated Theorem Proving in the ProTem
Programming Language”. The design of ProTem has been improved since then, and the old
implementation is now out-of-date. A new implementation is partly written.