Python: The Even Fuller Monty

Version 5.90.0.4

August 29, 2013

```
E ::= []
      |(fetch E) | (set! E e) | (set! val E)
      (alloc E)
       (module E e)
      \langle E, mval \rangle
      | ref := E | x := E
      |E[e := e] | v[E := e] | v[v := E]
      | E e
       if Eee
       let x = E in e
      | list(E,[e ...]) | list(val,Es)
        tuple \langle E, (e \dots) \rangle | tuple \langle val, Es \rangle
        set\langle E, (e \ldots) \rangle \mid set\langle val, Es \rangle
       E[e] | val[E]
       (builtin-prim op Es)
       (raise E)
      (return E)
      (tryexcept E x e e) | (tryfinally E e)
      | (loop E e) | (frame E)
      | E (e ...) | val Es
      | E (e ...)*e | val Es*e
       val (val ...)*E
      | (construct-module E) | (in-module E \epsilon)
Es ::= (val ... E e ...)
```

Figure 1: Evaluation contexts

Appendix 1: The Rest of λ_{π}

The figures in this section show the rest of the λ_{π} semantics. We proceed to briefly present each of them in turn.

1 Contexts and Control Flow

Figures 1, 2, 3, and 4 show the different *contexts* we use to capture left-to-right, eager order of operations in λ_{π} . *E* is the usual *evaluation context* that enforces left-to-right, eager evaluation of expressions. *T* is a context for the first expression of a tryexcept block, used to catch instances of raise. Similarly, *H* defines contexts for loops, detecting continue and break, and *R* defines contexts for return statements inside functions. Each interacts with a few expression forms to handle non-local control flow.

```
T ::= []
      |(fetch T)|(set! T e)|(set! val T)
      (alloc T)
      |\langle T, mval \rangle
      | ref := T | x := T
       T[e := e] | v[T := e] | v[v := T]
      | T e
      | if Tee
      | let x = T in e
      | list(T,e) | list(val,Ts)
       tuple \langle T, e \rangle | tuple \langle val, Ts \rangle
       set\langle T, e \rangle \mid set\langle val, Ts \rangle
      | T[e] | val[T]
      (builtin-prim op Ts)
      (raise T)
      |(1 \text{ op } T e)|(1 \text{ frame } T)
      | T (e ...) | val Ts
      | T (e ...)*e | val Ts*e
      | val (val ...)*T
      (construct-module T)
Ts ::= (val ... T e ...)
```

Figure 2: Contexts for try-except

```
H ::= []
      |(fetch H)|(set! H e)|(set! val H)
      (alloc H)
      |\langle H, mval \rangle
      | ref := H | x := H
      | H[e := e] | v[H := e] | v[v := H]
      | H e
      | if Hee
      | let x = H in e
      | list(H,e) | list(val,Hs)
      | tuple \langle H, e \rangle | tuple \langle val, Hs \rangle
      | set\langle H, e \rangle | set\langle val, Hs \rangle
      | H[e] | val[H]
      | (builtin-prim op Hs)
      (raise H)
      (tryexcept H x e e)
      | H (e ...) | val Hs
      | H (e ...)*e | val Hs*e
      | val (val ...)*H
      | (construct-module E)
Hs ::= (val ... H e ...)
```

Figure 3: Contexts for loops

```
R ::= []
      | (fetch R) | (set! R e) | (set! val R)
      | (alloc R)
      \langle R, mval \rangle
      | ref := R | x := R
       |R[e := e] | v[R := e] | v[v := R]
      | R e
      | if Ree
      | let x = R in e
       | list\langle R, e \rangle | list\langle val, Rs \rangle
      | tuple \langle R, e \rangle | tuple \langle val, Rs \rangle
      | set\langle R, e \rangle | set\langle val, Rs \rangle
      | R[e] | val[R]
      | (builtin-prim op Rs)
      (raise R)
      (loop R e)
      | (tryexcept R \ x \ e \ e)
      | R (e ...) | val Rs
      | R (e ...)*e | val Rs*e
      | val (val ...)*R
      | (construct-module E)
Rs ::= (val ... R e ...)
```

Figure 4: Contexts for return statements

Figure 5 shows how these contexts interact with expressions. For example, in first few rules, we see how we handle break and continue in while statements. When while takes a step, it yields a loop form that serves as the marker for where internal break and continue statements should collapse to. It is for this reason that *H* does *not* descend into nested loop forms; it would be incorrect for a break in a nested loop to break the outer loop.

One interesting feature of while and tryexcept in Python is that they have distinguished "else" clauses. For while loops, these else clauses run when the condition is False, but *not* when the loop is broken out of. For tryexcept, the else clause is only visited if *no* exception was thrown while evaluating the body. This is reflected in E-TryDone and the else branch of the if statement produced by E-While.

We handle one feature of Python's exception raising imperfectly. If a programmer uses raise without providing an explicit value to throw, *the exception bound in the most recent active catch block* is thrown instead. We have a limited solution that involves raising a special designated "reraise" value, but this fails to capture some subtle behavior of nested catch blocks. We believe a more sophisticated desugaring that uses a global stack to keep track of entrances and exits to catch blocks will work, but have yet to verify it. We still pass a number of tests that use raise with no argument.

2 Mutation

There are *three* separate mutability operators in λ_{π} , (set! *e e*), which mutates the value stored in a reference value, *e* := *e*, which mutates variables, and (set-field *e e e*), which updates and adds fields to objects.

Figure 6 shows the several operators that allocate and manipulate references in different ways. We briefly categorize the purpose for each type of mutation here:

- We use (set! *e e*), (fetch *e*) and (alloc *e*) to handle the update and creation of objects via the δ function, which reads but does not modify the store. Thus, even the lowly + operation needs to have its result re-allocated, since programmers only see references to numbers, not object values themselves. We leave the pieces of object values immutable and use this general strategy for updating them, rather than defining separate mutability for each type (e.g., lists).
- We use *e* := *e* for assignment to both local and global variables. We discuss global variables more in the next section. Local variables are handled at binding time by allocating references and substituting the new references wherever the variable appears. Local variable accesses and assignments thus work over references directly, since the variables have been substituted away by the time the actual assignment or access is reached. Note also that E-AssignLocal can override potential store entries.

(while $e_1 \ e_2 \ e_3$) \implies if e_1 (loop e_2 (while $e_1 \ e_2 \ e_3$)) e_3	[E-While]
(loop H [continue] e) \Longrightarrow e	[E-LoopContinue]
(loop H [break] e) \implies vnone	[E-LoopBreak]
(loop val e) \implies e	[E-LoopNext]
(tryexcept val x $e_{\scriptscriptstyle catch}$ $e_{\scriptscriptstyle else}$) \implies $e_{\scriptscriptstyle else}$	[E-TryDone]
$(tryexcept T[(raise val)] x e_{catch} e_{else}]$ $\implies let x = val in e_{catch}$) [E-TryCatch]
(frame $R[(return val)]) \Longrightarrow val$	[E-Return]
(frame val) \implies val	[E-FramePop]
$(tryfinally R[(return val)] e) \implies e (return val)$	[E-FinallyReturn]
(tryfinally H [break] e) $\implies e$ break	[E-FinallyBreak]
$(tryfinally T[(raise val)] e) \implies e (raise val)$	[E-FinallyRaise]
$(tryfinally H[continue] e) \implies e \text{ continue}$	[E-FinallyContinue]
$(E[if val e_1 e_2] \varepsilon \Sigma)$ $\longrightarrow (E[e_1] \varepsilon \Sigma)$ where (truthy? val Σ)	[E-lfTrue]
$(E[if val e_1 e_2] \varepsilon \Sigma)$ $\longrightarrow (E[e_2] \varepsilon \Sigma)$ where (not (truthy? val Σ))	[E-IfFalse]
val e \Longrightarrow e	[E-Seq]

Figure 5: Control flow reductions

$(E[@ref_{fun} (val)] \epsilon \Sigma)$	[E-App]
$\longrightarrow (E[(frame subst[[(x \dots), (ref_{arg} \dots), e]])] \epsilon \Sigma_1)$	
where $\langle any_c, \lambda(x \dots) \text{ (no-var). } e , any_{dict} \rangle = \Sigma(ref_{fun}),$ (equal? (length (val)) (length (x))), (Σ_1 (ref_{arg}1)) = extend-store/list[Σ , (val)]	
$(E[let x = v+undef in e] \varepsilon \Sigma)$	[E-LetLocal]
$ \longrightarrow (E[[x/ref]e] \varepsilon \Sigma_1) $ where $(\Sigma_1 ref) = \text{alloc}(\Sigma, v+undef)$	
$(E[let x = v+undef in e] \varepsilon \Sigma)$	[E-LetGlobal]
$\longrightarrow (E[e] \text{ extend-env}[\varepsilon, x, ref] \Sigma_1)$ where $(\Sigma_1 ref) = \text{alloc}(\Sigma, v+undef)$	
$(E[ref] \ \varepsilon \ \Sigma) \longrightarrow (E[val] \ \varepsilon \ \Sigma)$ where $\Sigma = ((ref_1 \ v + undef_1) \ \dots \ (ref \ val) \ (ref_n \ v + undef_n) \ \dots$	[E-GetVar]
$(E[ref] \in \Sigma) \longrightarrow (E[val] \in \Sigma)$	[E-GetVar]
where $\Sigma = ((ref_1 \ v + undef_1) \ \dots \ (ref \ val) \ (ref_n \ v + undef_n) \ \dots$.)
$(E[ref := val] \varepsilon \Sigma) \longrightarrow (E[val] \varepsilon \Sigma[ref := val])$	[E-AssignLocal]
$(E[x := val] \varepsilon \Sigma) \longrightarrow (E[val] \varepsilon \Sigma[ref:=val])$ where $\varepsilon = ((x_2 \ ref_2) \dots (x \ ref) (x_3 \ ref_3) \dots)$	[E-AssignGlobal]
$(E[(alloc val)] \varepsilon \Sigma) \longrightarrow (E[@ref_{new}] \varepsilon \Sigma_1)$ where $(\Sigma_1 ref_{new}) = alloc(\Sigma, val)$	[E-Alloc]
$(E[(fetch @ref)] \epsilon \Sigma) \longrightarrow (E[\Sigma(ref)] \epsilon \Sigma)$	[E-Fetch]
$\begin{array}{ll} (E[(set! @\textit{ref} val)] \ \varepsilon \ \Sigma) \longrightarrow (E[val] \ \varepsilon \ \Sigma_1) \\ \text{where } \Sigma_1 = \ \Sigma[\textit{ref}:=val] \end{array}$	[E-Set!]
$ \begin{array}{l} (E[@ref_{obj} \ [@ref_{str} \ := val_1]] \ \varepsilon \ \Sigma) \longrightarrow (E[val_1] \ \varepsilon \ \Sigma_2) \\ \text{where } \langle any_{cls1}, mval, \{string: ref, \ldots\} \rangle = \Sigma(ref_{obj}), \\ (\Sigma_1 \ ref_{new}) = alloc(\Sigma, val_1), \\ \langle any_{cls2}, string_1, any_{dict} \rangle = \Sigma(ref_{str}), \\ \Sigma_2 = \Sigma_1[ref_{obj}:=\langle any_{cls1}, mval, \{string_1: ref_{new}, string: ref, (not (member \ string_1 \ (string \ \ldots))) \end{array} $	[E-SetFieldAdd]
$(\boldsymbol{E}[@ref_{obj} [@ref_{str} := val_1]] \boldsymbol{\varepsilon} \boldsymbol{\Sigma})$	[E-SetFieldUpdate]
$\longrightarrow (E[val_1] \ \varepsilon \ \Sigma[ref_1:=val_1])$	
where $\langle any_{cls1}, mval, \{string_2: ref_2, \dots, string_1: ref_1, string_3: ref_3, \langle any_{cls2}, string_1, any_{dict} \rangle = \Sigma(ref_{str})$	\ldots } = $\Sigma(ref_{obj})$,
$(E[(delete ref)] \epsilon ((ref_1 v_1) \dots (ref v) (ref_n v_n) \dots))$	[E-DeleteLocal]
$\longrightarrow (E[v] \ \varepsilon \ ((ref_1 \ v_1) \ \dots \ (ref \ \And) \ (ref_n \ v_n) \ \dots))$	
$(E[(\text{delete } x)] ((x_1 \text{ ref}_1) \dots (x \text{ ref}) (x_n \text{ ref}_n) \dots) \Sigma)$	[E-DeleteGlobal]
$\longrightarrow (E[\Sigma(ref)] ((x_1 ref_1) \dots (x_n ref_n) \dots) \Sigma)$	

 $\begin{array}{ll} (E[\operatorname{eref}_{obj} \ [\operatorname{eref}_{str} \]] \ \varepsilon \ \Sigma) & [E-\operatorname{GetField-Class/Id}] \\ \longrightarrow (E[\operatorname{val}_{result}] \ \varepsilon \ \Sigma_{result}) & \\ & \text{where} \ \langle \operatorname{any}_{cls}, \operatorname{string}, \operatorname{any}_{dict} \rangle = \Sigma(\operatorname{ref}_{str}), & \\ & \langle x_{cls}, \operatorname{mval}, \{\operatorname{string}_1: \operatorname{ref}_2, \ldots\} \rangle = \Sigma(\operatorname{ref}_{obj}), & \\ & (\Sigma_{result} \ val_{result}) = \operatorname{class-lookup}[\![\operatorname{eref}_{obj}, \ \Sigma(\operatorname{env-lookup}[\![\varepsilon, \ x_{cls}]\!]), \ \operatorname{string}, \ \Sigma], & \\ & (\operatorname{not} \ (\operatorname{member} \ \operatorname{string} \ (\operatorname{string}_1 \ \ldots))) & \\ & &$

Figure 7: Accessing fields on a class defined by an identifier

• We use (set-field *e e e*) to update and add fields to objects' dictionaries. We leave the fields of objects' dictionaries as references and not values to allow ourselves the ability to share references between object fields and variables. We maintain a strict separation in our current semantics, with the exception of modules, and we expect that we'll continue to need it in the future to handle patterns of exec.

Finally, we show the E-Delete operators, which allow a Python program to revert the value in the store at a particular location back to \Re , or to remove a global binding.

3 Global Scope

While local variables are handled directly via substitution, we handle global scope with an explicit environment ε that follows the computation. We do this for two main reasons. First, because global scope in Python is truly dynamic in ways that local scope is not (exec can modify global scope), and we want to be open to those possibilities in the future. Second, and more implementation-specific, we use global scope to bootstrap some mutual dependencies in the object system, and allow ourselves a touch of dynamism in the semantics.

For example, when computing booleans, λ_{π} needs to yield numbers from the δ function that are real booleans (e.g., have the built-in %bool object as their class). However, we need booleans to set up if-tests while we are bootstrapping the creation of the boolean class itself! To handle this, we allow global identifiers to appear in the class position of objects. If we look for the class of an object, and the class position is an identifier, we look it up in the global environment. We only use identifiers with special %-prefixed names that aren't visible to Python in this way. It may be possible to remove this touch of dynamic scope from our semantics, but we haven't yet found the desugaring strategy that lets us do so. Figure 7 shows the reduction rule for field lookup in this case.

4 True, False, and None

The keywords True, False, and None are all singleton references in Python. In λ_{π} , we do not have a form for True, instead desugaring it to a variable reference bound in the environment. The same goes for None and False. We bind each to an allocation of an object:

```
let True = (alloc (%bool,1,{})) in
let False = (alloc (%bool,0,{})) in
let None = (alloc (%none, meta-none ,{})) in
eprog
```

and these bindings happen before anything else. This pattern ensures that all references to these identifiers in the desugared program are truly to the same objects. Note also that the boolean values are represented simply as number-like values, but with the built-in %bool class, so they can be added and subtracted like numbers, but perform method lookup on the %bool class. This reflects Python's semantics:

isinstance(True, int) # ==> True

5 Variable-arity Functions

We implement Python's variable-arity functions directly in our core semantics, with the reduction rules shown in figure 8. We show first the two arity-mismatch cases in the semantics, where either no vararg is supplied and the argument count is wrong, or where a vararg is supplied but the count is too low. If the count is higher than the number of parameters and a vararg is present, a new tuple is allocated with the extra arguments, and passed as the vararg. Finally, the form $e (e \dots) * e$ allows a variable-length collection of arguments to be *passed* to the function; this mimics apply in languages like Racket or JavaScript.

6 Modules

We model modules with the two rules in figure 9. E-ConstructModule starts the evaluation of the module, which is represented by a meta-code structure. A meta-code contains a list of global variables to bind for the module, a name for the module, and an expression that holds the module's body. To start evaluating the module, a new location is allocated for each global variable used in the module, initialized to & in the store, and a new environment is created mapping each of these new identifiers to the new locations.

Evaluation that proceeds inside the module, *replacing* the global environment ϵ with the newly-created environment. The old environment is stored with a new in-module form that is left in the current context. This step also sets up an expression that will create a new module object, whose fields hold the global variable references of the running module.

 $(E[@ref_{fun} (val ...)] \varepsilon \Sigma)$ [E-AppArity] \longrightarrow (*E*[(err (%str,"arity-mismatch",{})] $\epsilon \Sigma$) where $\langle any_c, \lambda(x \dots) \rangle$ (no-var).*e* , $any_{dict} \rangle = \Sigma(ref_{fun})$, (not (equal? (length (*val* ...)) (length (*x* ...)))) $(E[@ref_{fun} (val ...)] \varepsilon \Sigma)$ [E-AppVarArgsArity] \longrightarrow (*E*[(err (%str,"arity-mismatch-vargs",{}))] $\epsilon \Sigma$) where $\langle any_c, \lambda(x \dots) (y) . e , any_{dict} \rangle = \Sigma(ref_{fun}),$ (< (length (*val* ...)) (length (*x* ...))) $(E[@ref_{fun} (val ...)] \varepsilon \Sigma)$ [E-AppVarArgs1] $\longrightarrow (E[(frame subst[[(x \dots y_{varg}), (ref_{arg} \dots ref_{tupleptr}), e]])] \epsilon \Sigma_3)$ where $\langle any_c, \lambda(x \dots) (y_{varg}) \cdot e , any_{dict} \rangle = \Sigma(ref_{fun}),$ (>= (length (*val* ...)) (length (*x* ...))), $(val_{arg} \dots) = (take (val \dots) (length (x \dots))),$ $(val_{rest} \dots) = (drop (val \dots) (length (x \dots))),$ $val_{tuple} = \langle \text{%tuple}, (val_{rest} \ldots), \{\} \rangle$ $(\Sigma_1 \ (ref_{arg} \ \dots)) = \text{extend-store/list} [\Sigma, \ (val_{arg} \ \dots)],$ $(\Sigma_2 \ ref_{tuple}) = alloc(\Sigma_1, val_{tuple}),$ $(\Sigma_3 \text{ ref}_{tupleptr}) = \text{alloc}(\Sigma_2, \text{@ref}_{tuple})$ $(E[@ref_{fun} (val ...) * @ref_{var}] \varepsilon \Sigma)$ [E-AppVarArgs2] \longrightarrow (*E*[@ref_{fun} (val ... val_{extra} ...)] $\varepsilon \Sigma$) where $\langle any_c, (val_{extra} \dots), any_{dict} \rangle = \Sigma(ref_{var})$

Figure 8: Variable-arity functions

(*E*[(construct-module @*ref*_{mod})] $\varepsilon_{old} \Sigma$) [E-ConstructModule] \longrightarrow (*E*[(in-module $e_{body} \epsilon_{old}$) (alloc (\$module,(no-meta),{*string*_{arg}:*ref*_{new},...}))] $\boldsymbol{\varepsilon}_{new} \boldsymbol{\Sigma}_1$) where $\langle any_{cls}, (meta-code (x_{arg} \dots) x_{name} e_{body}), any_{dict} \rangle = \Sigma(ref_{mod}),$ $(\Sigma_1 \ \varepsilon_{new} \ (ref_{new} \ \ldots)) = vars -> fresh-env \llbracket \Sigma, \ (x_{arg} \ \ldots) \rrbracket,$ $(string_{arg} \dots) = (map symbol -> string (X_{arg} \dots))$ $(\textit{E}[(\text{in-module } \textit{v} \ \textit{\epsilon}_{old})] \ \textit{\epsilon}_{mod} \ \textit{\Sigma}) \longrightarrow (\textit{E}[\text{vnone}] \ \textit{\epsilon}_{old} \ \textit{\Sigma})$ [E-ModuleDone] vars->fresh-env[[Σ, ()]] $= (\Sigma () ())$ = $(\Sigma_1 ((x ref)) (ref))$ vars->fresh-env $\llbracket \Sigma$, (*x*) \rrbracket where $(\Sigma_1 \ (ref)) = (alloc \ \Sigma \Re)$ vars->fresh-env $\llbracket \Sigma$, $(x x_r \dots) \rrbracket = (\Sigma_2 ((x ref) (x_{rest} ref_{rest}) \dots) (ref ref_{rest} \dots))$ where $(\Sigma_1 \text{ ref}) = \text{alloc}(\Sigma, \Re)$, $(\Sigma_2 ((x_{rest} ref_{rest}) \dots) (ref_{rest} \dots)) = vars -> fresh-env \llbracket \Sigma_1, (x_r \dots) \rrbracket$

Figure 9: Simple modules in λ_{π}

When the evaluation of the module is complete (the in-module term sees a value), the old global environment is reinstated.

To desugar to these core forms, we desugar the files to be imported, and analyze their body to find all the global variables they use. The desugared expression and variables are combined with the filename to create the meta-code object. This is roughly an implementation of Python's compile, and it should be straightforward to extend it to implement exec, though for now we've focused on specifically the module case.