### Homework 6

The purpose of this assignment is to grapple with how dynamic binding arises and how to formally specify semantics. Concretely, we will extend JAKARTASCRIPT with recursive functions and implement two interpreters. The first will be a big-step interpreter that is an extension of Homework 4 but implements dynamic binding "by accident". The second will be a small-step interpreter that exposes evaluation order and implements static binding by substitution.

Try to make your code as concise and clear as possible. Challenge yourself to find the most crisp, concise way of expressing the intended computation. This may mean using ways of expressing computation currently unfamiliar to you.

For submission instructions and the due date, please see the README.md file.

# Problem 1 JAKARTASCRIPT Interpreter: Recursive Functions, and Dynamic Binding (12 Points)

We now have the formal tools to specify exactly how a JAKARTASCRIPT program should behave. Unless otherwise specified, we will continue to try to match JavaScript semantics as implemented by Node.js. Thus, it is still useful to write little test JavaScript programs and run them through Node.js to see how the tests should behave. Finding bugs in the JAKARTASCRIPT specification with respect to JavaScript is certainly deserving of extra credit.

In this homework, we extend JakartaScript with recursive functions. This language is very similar to the LETREC language in Section 3.4 of Friedman and Wand.

For this problem, we try to implement functions as an extension of Homework 4 in the most straightforward way. What we will discover is that we have made a historical mistake and ended up with a form of dynamic binding.

The syntax of JakartaScript for this homework is given in Figure 1. Note that the grammar specifies the abstract syntax using notation borrowed from the concrete syntax. The new constructs are highlighted. We have added function expressions **function** p(x)e and function calls  $e_1(e_2)$ .

In a function expression, the function name p can either be a variable name or empty. The variable name is used for recursion. For simplicity, all functions are one argument functions. Since functions are first-class values (i.e., they are treated like any other expression), we can get multi-argument functions via currying.

We have also added a "marker" typeerror to the expression language. This marker is not part of the source language. It is used in our definition of the evaluation relations. We discuss this in more detail further below.

As before, the concrete syntax accepted by the parser is slightly less flexible than the abstract syntax in order to match the syntactic structure of JavaScript. For function expressions, the body is surrounded by curly braces (i.e.,  $\{\ \}$ ) and consists of an optional statement  $st_{opt}$  for **const** bindings followed by a **return** with an expression e.

An abstract syntax tree representation is provided for you in ast.scala. We also provide a parser and main driver for testing. The correspondence between the concrete syntax and the abstract syntax representation is shown in Figure 3.

```
n \in Num
                                                                                               numbers (double)
              s \in Str
                                                                                                              strings
              x \in Var
                                                                                                           variables
 b \in Bool ::= true \mid false
                                                                                                          Booleans
    v \in Val ::= \mathbf{undefined} \mid n \mid b \mid s \mid \mathbf{function} \ p \ (x) \ e \mid \mathsf{typeerror}
                                                                                                              values
  e \in Expr ::= x \mid v \mid uop \ e \mid e_1 \ bop \ e_2 \mid e_1 \ ? \ e_2 : e_3 \mid
                                                                                                       expressions
                    const x = e_1; e_2 \mid \mathbf{console.log}(e) \mid e_1 \mid e_2 \mid
uop \in Uop := - \mid !
                                                                                                 unary operators
bop \in Bop ::= + |-| * | / | === | ! == | < | > | <= | >= | & & | | | | |,
                                                                                                binary operators
            p := x \mid \epsilon
                                                                                                  function names
```

Figure 1: Abstract syntax

```
e ::= \cdots \mid \overrightarrow{\textbf{const}} \ \overrightarrow{x} \ e_1; e_2 \mid \overrightarrow{\textbf{typeerror}} \mid (e) expressions  \overrightarrow{\textbf{function}} \ p(x) \ e \mid \textbf{function} \ p(x) \ \{st_{opt} \ \textbf{return} \ e\} st ::= \textbf{const} \ x = e_1; \mid e; \mid \ ; \mid \{st\} \mid st_1 \ st_2 statements  st_{opt} ::= st \mid \epsilon
```

Figure 2: Concrete syntax

A big-step operational semantics of JakartaScript is given in Figure 4. The rules may seem daunting at first, but reading them will become easier with practice. This homework is such an opportunity to practice.

A formal semantics enables us to describe the semantics of a programming language clearly and concisely. The initial barrier is getting used to the meta-language of judgment forms and inference rules. However, once you cross that barrier, you will see that we are telling you exactly how to implement the interpreter—it will almost feel like cheating!

In Figure 4, we define the relation  $env \vdash e \Downarrow v$ . This relation corresponds directly to the eval function that we are asking you to implement (not a coincidence). It similarly has three parts:

```
def eval(env: Env, e: Expr): Val
```

It takes as input a value environment env and an expression e, and returns a value.

In Homework 4, all expressions could be evaluated to something (because of type conversion). With functions, we encounter one of the very few run-time errors in JavaScript: trying to call something that is not a function. In JavaScript and in JakartaScript, calling a non-function raises a run-time error. In the formal semantics, we model this with evaluating to the "marker" typeerror.

Such a run-time error is known as a *dynamic type error*. Languages are called dynamically

```
sealed abstract class Expr extends Positional abstract class Val extends Expr ... case class Function(p: Option[String], x: String, e1: Expr) extends Val Function(p, x, e) function p(x)e case class Call(e1: Expr, e2: Expr) extends Expr Call(e_1, e_2) e_1(e_2)
```

Figure 3: Representing in Scala the abstract syntax of JAKARTASCRIPT. After each case class or case object, we show the correspondence between the representation and the concrete syntax.

typed when they allow all syntactically valid programs to run and check for type errors during execution.

In our Scala implementation, we will not clutter our Expr type with a typeerror marker. Instead, we will use a Scala exception DynamicTypeError:

```
case class DynamicTypeError(e: Expr)
```

to signal this case. In other words, when your interpreter discovers a dynamic type error, you should throw this exception using the following Scala code:

```
throw DynamicTypeError(e)
```

The argument e should be the input expression to eval where the type error was detected. One advantage of using a Scala exception for typeerror is that the marker does not need to be propagated explicitly as in the inference rules in Figure 5. In particular, your interpreter will implement the EVALTYPEERROR rules explicitly, but the EVALPROPAGATE rules are implemented implicitly with Scala's exception propagation mechanism.

Note in rule EVALEEQUAL, we disallow equality and disequality checks (i.e., === and !==) on function values. If either argument to a equality or disequality check is a function value, then we consider this a dynamic type error. This choice is a slight departure from JavaScript.

(a) First, write some JakartaScript programs and execute them as JavaScript programs. This step will inform how you will implement your interpreter and will serve as tests for your interpreter.

Think of one test case that behaves differently under dynamic binding versus static binding (and does not crash with a dynamic type error). Once you have finished the implementation of both interpreters, run your interpreters on your test case using sbt to confirm that the two interpreters indeed behave differently.

(b) Then implement:

```
def eval(env: Env, e: Expr): Val
```

that evaluates a JakartaScript expression e in a value environment env to a value according to the evaluation judgment  $env \vdash e \Downarrow v$ . The following helper functions for converting values to numbers, Booleans, and strings are provided:

```
def toNumber(v: Val): Double
def toBoolean(v: Val): Boolean
def toStr(v: Val): String
```

We suggest the following step-by-step process:

- 1. Bring your Homework 4 implementation into Homework 6 and make sure your previous test cases work as expected.
- 2. Extend your implementation with non-recursive functions. On function calls, you need to extend the environment for the formal parameter but not for the function itself. Do not worry yet about dynamic type errors.
- 3. Add support for checking for dynamic type errors. This requires changes in the cases for function calls and (dis)equality operators.
- 4. Check that your interpreter, unfortunately, implements dynamic binding instead of static binding.
- 5. Modify your implementation to support recursive functions.

# Problem 2 JAKARTASCRIPT Interpreter: Substitution and Evaluation Order (28 Points)

In this problem, we will do two things. First, we will remove environments and instead use a language semantics based on substitution. This change will "fix" the scoping issue, and we will end up with static binding.

As an aside, substitution is not the only way to "fix" the scoping issue. Another way is to represent a function value as a closure, which is a pair of the function with the environment when it is defined. Substitution is a fairly simple way to get static binding, but in practice, it is rarely used because it is not the most efficient implementation.

The second thing that we do is move to implementing a small-step interpreter. A small-step interpreter makes explicit the evaluation order of expressions. These two changes are orthogonal, that is, one could implement a big-step interpreter using substitution or a small-step interpreter using environments.

#### (a) Implement

```
def subst(e: Expr, x: String, v: Val): Expr
```

that substitutes value v for all free occurrences of variable x in expression e. The function subst requires that v does not have free variables, so you do not need to worry about variable capturing. Still, be careful that you treat the variable binding constructs correctly (i.e., ConstDecl and Function) so that you only substitute free occurrences of x in e by v. In particular, calling subst on expression

```
x, (const x = 4; x)
```

with value 3 and variable x should return

Figure 4: Big-step operational semantics of JAKARTASCRIPT (with dynamic binding)

$$\frac{env \vdash e_1 \Downarrow \mathbf{function} \ p(x) \ e \quad bop \in \{===,!==\}}{env \vdash e_1 \ bop \ e_2 \Downarrow \mathbf{typeerror}} \quad \text{EvalTypeErrorEqual}_1$$
 
$$\frac{env \vdash e_2 \Downarrow \mathbf{function} \ p(x) \ e \quad bop \in \{===,!==\}}{env \vdash e_1 \ bop \ e_2 \Downarrow \mathbf{typeerror}} \quad \text{EvalTypeErrorEqual}_2$$
 
$$\frac{env \vdash e_1 \Downarrow v_1 \quad v_1 \neq \mathbf{function} \ p(x) \ e}{env \vdash e_1 \ (e_2) \ \Downarrow \mathbf{typeerror}} \quad \text{EvalTypeErrorCall}$$
 
$$\frac{env \vdash e_1 \Downarrow \mathbf{typeerror}}{env \vdash e_1 ? \ e_2 : e_3 \Downarrow \mathbf{typeerror}} \quad \text{EvalPropIf} \quad \frac{env \vdash e_1 \Downarrow \mathbf{typeerror}}{env \vdash e_1 \ bop \ e_2 \Downarrow \mathbf{typeerror}} \quad \text{EvalPropBop}_1$$
 
$$\frac{env \vdash e \Downarrow \mathbf{typeerror}}{env \vdash uop \ e \Downarrow \mathbf{typeerror}} \quad \text{EvalPropUop} \quad \frac{env \vdash e_2 \Downarrow \mathbf{typeerror}}{env \vdash e_1 \ bop \ e_2 \Downarrow \mathbf{typeerror}} \quad \text{EvalPropBop}_2$$
 
$$\frac{env \vdash e_1 \Downarrow v_1 \quad v_1 \neq \mathbf{typeerror}}{env \vdash e_1 \& \& e_2 \Downarrow \mathbf{typeerror}} \quad \text{EvalPropAnd}$$
 
$$\frac{env \vdash e_1 \Downarrow v_1 \quad v_1 \neq \mathbf{typeerror}}{env \vdash e_1 \vdash e_2 \Downarrow \mathbf{typeerror}} \quad \text{EvalPropOr}$$
 
$$\frac{env \vdash e_1 \Downarrow v_1 \quad v_1 \neq \mathbf{typeerror}}{env \vdash e_1 \vdash e_2 \Downarrow \mathbf{typeerror}} \quad \text{EvalPropOr}$$
 
$$\frac{env \vdash e_1 \Downarrow v_1 \quad v_1 \neq \mathbf{typeerror}}{env \vdash e_1 \vdash e_2 \Downarrow \mathbf{typeerror}} \quad \text{EvalPropOr}$$
 
$$\frac{env \vdash e_1 \Downarrow v_1 \quad v_1 \neq \mathbf{typeerror}}{env \vdash e_1 \vdash e_2 \Downarrow \mathbf{typeerror}} \quad \text{EvalPropOr}$$
 
$$\frac{env \vdash e_1 \Downarrow \mathbf{typeerror}}{env \vdash e_1 \vdash e_2 \Downarrow \mathbf{typeerror}} \quad \text{EvalPropCall}_1$$

Figure 5: Big-step operational semantics of JAKARTASCRIPT: dynamic type error rules

```
3, (const x = 4; x)
not
3, (const x = 4; 3)
```

The function subst is a helper for the step function, but you might want to implement all of the cases of step that do not require subst first.

#### (b) Implement

```
def step(e: Expr): Expr
```

that performs one-step of evaluation by rewriting the input expression e into a "one-step reduced" expression. This one-step reduction should be implemented according to the small-step semantics  $e \to e'$  defined in Figures 6, 7, and 8. We write e[v/x] for substituting value v for all free occurrences of the variable x in expression e (i.e., this corresponds to subst (e, x, v) in the Scala implementation).

It is informative to compare the small-step semantics used in this problem and the big-step semantics from the previous one. In particular, for all programs where dynamic binding is not an issue, your interpreters in this problem and the previous should behave the same. We have provided the functions evaluate and iterateStep that evaluate "top-level" expressions to a value using your interpreter implementations. Executing your interpreter

$$\frac{e_1 \rightarrow e_1'}{e_1 \, bop \, e_2 \rightarrow e_1' \, bop \, e_2} \, \operatorname{SearchBop}_1 \quad \frac{bop \notin \{\text{, , &&, | |, ===, !==}\}}{v_1 \, bop \, e_2 \rightarrow v_1 \, bop \, e_2'} \, \operatorname{SearchBop}_2$$

$$\frac{e_2 \rightarrow e_2'}{v_1 \neq \operatorname{function} p \, (x) \, e \quad bop \in \{===, !==\}}{v_1 \, bop \, e_2'} \, \operatorname{SearchEqual}$$

$$\frac{e \rightarrow e'}{v_1 \, bop \, e_2 \rightarrow v_1 \, bop \, e_2'} \, \operatorname{SearchPrint}$$

$$\frac{e \rightarrow e'}{v_1 \, bop \, e_2 \rightarrow v_2'} \, \operatorname{SearchPrint}$$

$$\frac{e_1 \rightarrow e_1'}{e_1 \, ? \, e_2 \, : \, e_3 \rightarrow e_1' \, ? \, e_2 \, : \, e_3} \, \operatorname{SearchFrint}$$

$$\frac{e_1 \rightarrow e_1'}{\operatorname{const} x = e_d; e_b \rightarrow \operatorname{const} x = e_d'; e_b} \, \operatorname{SearchConst}$$

$$\frac{e_1 \rightarrow e_1'}{e_1 \, (e_2) \rightarrow e_1' \, (e_2)} \, \operatorname{SearchCall}_1 \quad \frac{e_2 \rightarrow e_2'}{v_1 \, (e_2) \rightarrow v_1 \, (e_2')} \, \operatorname{SearchCall}_2$$

Figure 6: Small-step operational semantics of JakartaScript: search rules

on a JavaScript input file using the run command of sbt will evaluate the input program with both of these functions.

$$\frac{v'=-toNum(v)}{-v \to v'} \text{ DoUMINUS} \qquad \frac{v'=!toBool(v)}{! \ v \to v'} \text{ DoNot} \qquad \frac{v_1, \ e_2 \to e_2}{v_1, \ e_2 \to e_2} \text{ DoSeq}$$
 
$$\frac{v=toNum(v_1)+toNum(v_2)}{v_1+v_2 \to v} \quad v_1, v_2 \notin Str \quad \text{DoPlusNum}$$
 
$$\frac{v=s_1+toStr(v_2)}{s_1+v_2 \to v} \text{ DoPlusStr}_1 \qquad \frac{v=toStr(v_1)+s_2}{v_1+s_2 \to v} \text{ DoPlusStr}_2$$
 
$$\frac{v=toNum(v_1) \ bop \ toNum(v_2)}{v_1 \ bop \ v_2 \to v} \qquad \frac{toPlusStr}{v_2 \ bop \ e_2 \times e_3 \to e_2} \qquad \frac{v_1 \notin Str}{v_2 \ bop \ e_3 \times e_3 \to e_3} \qquad \frac{v_1 \notin Str}{v_2 \ bop \ e_3 \times e_3 \to e_3} \qquad \frac{v_2 \notin Str}{v_2 \notin Str} \qquad \frac{v_2 \notin Str}{$$

Figure 7: Small-step operational semantics of JakartaScript: do rules

Figure 8: Small-step operational semantics of JakartaScript: dynamic type error rules