CS 61A   Lecture Notes   Second Half of Week 2

Topic: Data abstraction

Reading: Abelson & Sussman, Sections 2.1 and 2.2.1 (pages 79–106)

• Big ideas: data abstraction, abstraction barrier.

If we are dealing with some particular type of data, we want to talk about it in terms of its *meaning*, not in terms of how it happens to be represented in the computer.

Example: Here is a function that computes the total point score of a hand of playing cards. (This simplified function ignores the problem of cards whose rank-name isn't a number.)

```
;;;;;                          In file cs61a/lectures/2.1/total.scm
(define (total hand)
  (if (empty? hand)
      0
      (+ (butlast (last hand))
         (total (butlast hand)) )))


> (total '(3h 10c 4d))
17
```

This function calls `butlast` in two places. What do those two invocations mean? Compare it with a modified version:

```
;;;;;                          In file cs61a/lectures/2.1/total.scm
(define (total hand)
  (if (empty? hand)
      0
      (+ (rank (one-card hand))
         (total (remaining-cards hand)) )))

(define rank butlast)
(define suit last)

(define one-card last)
(define remaining-cards butlast)
```

This is more work to type in, but the result is much more readable. If for some reason we wanted to modify the program to add up the cards left to right instead of right to left, we'd have trouble editing the original version because we wouldn't know which `butlast` to change. In the new version it's easy to keep track of which function does what.

The auxiliary functions like `rank` are called *selectors* because they select one component of a multi-part datum.

Actually we're *violating* the data abstraction when we type in a hand of cards as `'(3h 10c 4d)` because that assumes we know how the cards are represented—namely, as words combining the rank number with a one-letter suit. If we want to be thorough about hiding the representation, we need *constructor* functions as well as the selectors:

```
;;;;;                        In file cs61a/lectures/2.1/total.scm
(define (make-card rank suit)
  (word rank (first suit)) )

(define make-hand se)


> (total (make-hand (make-card 3 'heart)
                    (make-card 10 'club)
                    (make-card 4 'diamond) ))
17
```

Once we're using data abstraction we can change the implementation of the data type without affecting the programs that *use* that data type. This means we can change how we represent a card, for example, without rewriting `total`:

```
;;;;;                        In file cs61a/lectures/2.1/total.scm
(define (make-card rank suit)
  (cond ((equal? suit 'heart) rank)
        ((equal? suit 'spade) (+ rank 13))
        ((equal? suit 'diamond) (+ rank 26))
        ((equal? suit 'club) (+ rank 39))
        (else (error "say what?")) ))

(define (rank card)
  (remainder card 13))

(define (suit card)
  (nth (quotient card 13) '(heart spade diamond club)))
```

We have changed the internal *representation* so that a card is now just a number between 1 and 52 (why? maybe we're programming in FORTRAN) but we haven't changed the *behavior* of the program at all. We still call `total` the same way.

Data abstraction is a really good idea because it helps keep you from getting confused when you're dealing with lots of data types, but don't get religious about it. For example, we have invented the *sentence* data type for this course. We have provided symmetric selectors `first` and `last`, and symmetric selectors `butfirst` and `butlast`. You can write programs using sentences without knowing how they're implemented. But it turns out that because of the way they *are* implemented, `first` and `butfirst` take $O(1)$ time, while `last` and `butlast` take $O(N)$ time. If you know that, your programs will be faster.

• Pairs.

To represent data types that have component parts (like the rank and suit of a card), you have to have some way to *aggregate* information. Many languages have the idea of an *array* that groups some number of elements. In Lisp the most basic aggregation unit is the *pair*—two things combined to form a bigger thing. If you want more than two parts you can hook a bunch of pairs together; we'll discuss this more next week.

The constructor for pairs is CONS; the selectors are CAR and CDR.

The book uses pairs to represent many different abstract data types: rational numbers (numerator and denominator), complex numbers (real and imaginary parts), points (x and y coordinates), intervals (low and high bounds), and line segments (two endpoints). Notice that in the case of line segments we think of the representation as *one pair* containing two points, not as three pairs containing four numbers. (That's what it means to respect a data abstraction.)

Note: What's the difference between these two:

```
(define (make-rat num den) (cons num den))
(define make-rat cons)
```

They are both equally good ways to implement a constructor for an abstract data type. The second way has a slight speed advantage (one fewer function call) but the first way has a debugging advantage because you can trace `make-rat` without tracing all invocations of `cons`.

• Data aggregation doesn't have to be primitive.

In most languages the data aggregation mechanism (the array or whatever) seems to be a necessary part of the core language, not something you could implement as a user of the language. But if we have first-class functions we can use a function to represent a pair:

```
;;;;;                           In file cs61a/lectures/2.1/cons.scm
(define (cons x y)
  (lambda (which)
    (cond ((equal? which 'car) x)
          ((equal? which 'cdr) y)
          (else (error "Bad message to CONS" message)) )))


(define (car pair)
  (pair 'car))

(define (cdr pair)
  (pair 'cdr))
```

This is like the version in the book except that they use 0 and 1 as the *messages* because they haven't introduced quoted words yet. This version makes it a little clearer what the argument named `which` means.

The point is that we can satisfy ourselves that this version of `cons`, `car`, and `cdr` works in the sense that if we construct a pair with this `cons` we can extract its two components with this `car` and `cdr`. If that's true, we don't need to have pairs built into the language! All we need is `lambda` and we can implement the rest ourselves. (It isn't really done this way, in real life, for efficiency reasons, but it's neat that it could be.)

• Big idea: abstract data type *sequence* (or *list*).

We want to represent an ordered sequence of things. (They can be any kind of things.) We *implement* sequences using pairs, with each `car` pointing to an element and each `cdr` pointing to the next pair.

What should the constructors and selectors be? The most obvious thing is to have a constructor `list` that takes any number of arguments and returns a list of those arguments, and a selector `nth` that takes a number and a list as arguments, returning the nth element of the list.

Scheme does provide those, but it often turns out to be more useful to select from a list differently, with a selector for the first element and a selector for all the rest of the elements (i.e., a smaller list). This helps us write recursive functions such as the mapping and filtering ones we saw for sentences earlier.

Since we are implementing lists using pairs, we ought to have specially-named constructors and selectors for lists, just like for rational numbers:

```
(define adjoin cons)
(define first car)
(define rest cdr)
```

Many Lisp systems do in fact provide `first` and `rest` as synonyms for `car` and `cdr`, but the fact is that this particular data abstraction is commonly violated; we just use the names `car`, `cdr`, and `cons` to talk about lists.

This abstract data type has a special status in the Scheme interpreter itself, because lists are read and printed using a special notation. If Scheme knew only about pairs, and not about lists, then when we construct the list `(1 2 3)` it would print as `(1 . (2 .(3 . ())))` instead.

• Lists vs. sentences.

We started out the semester using an abstract data type called *sentence* that looks a lot like a list. What's the difference, and why did we do it that way?

Our goal was to allow you to create aggregates of words without having to think about the structure of their internal representation (i.e., about pairs). We do this by deciding that the elements of a sentence must be words (not sublists), and enforcing that by giving you the constructor `sentence` that creates only sentences.

Example: One of the homework problems for this problem set asks you to reverse a list. You'll see that this is a little tricky using `cons`, `car`, and `cdr` as the problem asks, but it's easy for sentences:

```
(define (reverse sent)
  (if (empty? sent)
      '()
      (se (reverse (bf sent)) (first sent)) ))
```

To give you a better idea about what a sentence is, here's a version of the constructor function:

```
;;;;;                          In file cs61a/lectures/2.2/sentence.scm
(define (se a b)
  (cond ((word? a) (se (list a) b))
        ((word? b) (se a (list b)))
        (else (append a b)) ))

(define (word? x)
  (or (symbol? x) (number? x)) )
```

`Se` is a lot like `append`, except that the latter behaves oddly if given words as arguments. `Se` can accept words or sentences as arguments.

• Box and pointer diagrams.

Here are a few details that people sometimes get wrong about them:

1. An arrow can't point to half of a pair. If an arrowhead touches a pair, it's pointing to the entire pair, and it doesn't matter exactly where the arrowhead touches the rectangle. If you see something like

```
(define x (car y))
```

where `y` is a pair, the arrow for `x` should point to *the thing that the* `car` *of* `y` *points to*, not to the left half of the `y` rectangle.

2. The direction of arrows (up, down, left, right) is irrelevant. You can draw them however you want to make the arrangement of pairs neat. That's why it's crucial not to forget the arrowheads!

3. There must be a top-level arrow to show where the structure you're representing begins.

How do you draw a diagram for a complicated list? Take this example:

```
((a b) c (d (e f)))
```

You begin by asking yourself how many elements the list has. In this case it has three elements: first `(a b)`, then `c`, then the rest. Therefore you should draw a three-pair *backbone*: three pairs with the `cdr` of one pointing to the next one. (The final `cdr` is null.)

Only after you've drawn the backbone should you worry about making the `car`s of your three pairs point to the three elements of the top-level list.