How closures could have been invented

Let us consider some type and a function f from that type to that type, a so-called endofunction. Argument and function value being of the same type means that we can now consider the equation

(0) $\quad x: f.x = x \quad ,$

solutions of which are called fixpoints of f.

Without any further knowledge of the type, there is little more we can say about the solutions of (0) than that each value of the type solves (0) if f is the identity function (and vice versa).

In order to be able to say somewhat more we impose some structure on the type by introducing a relation $\sqsubseteq$ (pronounced "under") on it, which we postulate to be a partial order, i.e. to be

(1) reflexive, i.e. $\quad x \sqsubseteq x$

(2) antisymmetric, i.e. $x \sqsubseteq y \wedge y \sqsubseteq x \Rightarrow x = y$

(3) transitive, i.e. $x \sqsubseteq y \wedge y \sqsubseteq z \Rightarrow x \sqsubseteq z$,

the above three formal explanations holding for all x, y, z of the type in question.

Aside From a logical point of view, the existence of a partial order on the type is a very mild requirement, since $=$ (read "equals") is a partial order; we leave to the reader to verify that $=$ is indeed reflexive, antisymmetric, and transitive. Equality is a very special relation in that it is also symmetric i.e. $x = y \equiv y = x$, a constraint that does not hold for the general partial order $\sqsubseteq$: $x \sqsubseteq y$ may differ from $y \sqsubseteq x$. (End of Aside.)

We can use our partial order to strengthen equation (0) and express that we are only interested in fixpoints of f that are "above" some lower bound h ; in formula, we consider the solutions of

(4) x: $f.x = x \wedge h \sqsubseteq x$.

Did the solutions of (0) depend only on f, those of (4) depend in general on h as well.

Equation (4) can still have many solutions, but very often the relation $\sqsubseteq$ provides the means for singling out a special one, which –if it exists– is known a <u>the lowest solution</u> of (4). It is singled out by stating that we are interested in that solution x of (4) that in addition satisfies

(5) $f.y = y \wedge h \sqsubseteq y \Rightarrow x \sqsubseteq y$ for any y.

In words: the lowest solution (of (4)) is required to lie under any solution y (of (4)).

<u>Aside</u> For any equation, there is at most 1 lowest solution, as is shown by the following argument. Let p and q both be a lowest solution of the equation $x\colon B.x$, i.e. we have

(6) $B.p$

(7) $B.y \Rightarrow p \sqsubseteq y$ for all y

(8) $B.q$

(9) $B.y \Rightarrow q \sqsubseteq y$ for all y .

We now deduce the equality of p and q by observing

$$
\begin{array}{ll}
 & p = q \\
\Leftarrow & \{ (2) \text{, i.e. } \sqsubseteq \text{ is antisymmetric} \} \\
 & p \sqsubseteq q \ \wedge\ q \sqsubseteq p \\
\Leftarrow & \{ (7) \text{ with } y := q \text{ ; } (9) \text{ with } y := p \} \\
 & B.q \ \wedge\ B.p \\
\equiv & \{ (8) \text{ ; } (6) \} \\
 & true \quad .
\end{array}
$$

If a lowest solution exists, it is unique.
(End of Aside.)

The lowest solution of (4) depends in general on f and h. The simplest expression that does so is $f.h$ and that raises the question: what properties are required of f such that (for any h), $f.h$ is the lowest solution of (4)? Formally: which properties of f enable us to establish

(10) $\quad f.(f.h) = f.h \ \wedge\ h \sqsubseteq f.h$

(11) $\quad f.y = y \wedge h \sqsubseteq y \ \Rightarrow\ f.h \sqsubseteq y \qquad ?$

Relation (10) follows if f is

(12) raising, i.e. $x \sqsubseteq f.x$, and

(13) idempotent, i.e. $f.(f.x) = f.x$

(the formal explanations holding for all x)

In order to establish (11) we observe, starting with the consequent

$$
\begin{array}{ll}
 & f.h \sqsubseteq y \\
\Leftarrow & \{\text{substituting equals for equals}\} \\
 & f.y = y \;\wedge\; f.h \sqsubseteq f.y \\
\Leftarrow & \{f \text{ is monotonic, see (14)}\} \\
 & f.y = y \;\wedge\; h \sqsubseteq y \quad ,
\end{array}
$$

i.e. for the demonstration of (11) we need that f is

(14) monotonic, i.e. $x \sqsubseteq y \Rightarrow f.x \sqsubseteq f.y$.

(Another name for "monotonic (with respect to $\sqsubseteq$" is "$\sqsubseteq$-preserving".)

A function that is raising, idempotent and monotonic is called a closure, and we have established the

Theorem If f is a closure, $f.h$ is the lowest solution of

$$x:\; f.x = x \;\wedge\; h \sqsubseteq x \quad .$$

And that concludes my story of how closures could have been discovered.

Austin, 1 September 1998

prof. dr Edsger W. Dijkstra
Department of Computer Sciences
The University of Texas at Austin
Austin, TX 78712-1188
USA