Category Theory as Coherently Constructive Lattice Theory: An Illustration

Roland Backhouse and Marcel Bijsterveld

Department of Mathematics and Computing Science,

Eindhoven University of Technology,

P.O. Box 513,

5600 MB Eindhoven,

The Netherlands.

email: `rolandb@tue.win.nl`.

November 18, 1994

Abstract

Dijkstra and Scholten have formulated a theorem stating that all disjunctivity properties of a predicate transformer are preserved by the construction of least prefix points. An alternative proof of their theorem is presented based on two fundamental fixed point theorems, the abstraction theorem and the fusion theorem, and the fact that suprema in a lattice are defined by a Galois connection. The abstraction theorem seems to be new; the fusion theorem is known but its importance does not seem to be fully recognised.

The abstraction theorem, the fusion theorem, and Dijkstra and Scholten's theorem are then generalised to the context of category theory and shown to be valid. None of the theorems in this context seems to be known, although specific instances of Dijkstra and Scholten's theorem are known.

The main point of the paper is to discuss the process of drawing inspiration from lattice theory to formulate theorems in category theory (first advocated by Lambek in 1968). We advance the view that, in order to contribute to the development of programming methodology, category theory may be profitably regarded as "constructive" lattice theory in which the added value lies in establishing that the constructions are "coherent".

This paper was specially prepared for presentation at the meeting of IFIP Working Group 2.3 (Programming Methodology), June 1994. Knowledge of (elementary) lattice theory is assumed. Knowledge of category theory is not.

1 Introduction

In 1968, Lambek [11] began an article with the sentence

> Classical results on lattices provide a fruitful source of inspiration for discoveries about categories, in view of the fact that a partially ordered set may be regarded as a category in which there is at most one map between any two objects.

The novel contribution of this paper is to show how a *theorem* in lattice theory described by Dijkstra and Scholten [8, p. 159] as "beautiful" has inspired us to construct a similarly "beautiful" theorem in category theory. An additional contribution is to show how the emphasis placed in category theory on the notion of an adjunction has inspired us to construct a novel *proof* of Dijkstra and Scholten's theorem that, in our view, better deserves to be described as "beautiful" than the proof given by them.

The main goal of this paper is broader than these specific results. Our goal is to advance the view that, in order to contribute to the development of programming methodology, category theory can be profitably regarded as "constructive" lattice theory in which the added value lies in establishing that the constructions are "coherent".

Dijkstra and Scholten formulate the "beautiful theorem" in the context of the predicate calculus. In terms of lattice theory the theorem is stated as follows. Suppose that $\mathcal{A} = (A\, ,\sqsupseteq)$ and $\mathcal{B} = (B, \succeq)$ are complete lattices and suppose $\oplus \in (\mathcal{A}{\leftarrow}\mathcal{A}){\leftarrow}\mathcal{B}$ is a monotonic function. Denote application of $\oplus$ to x by $x\oplus$. Suppose that, for all x, $x\oplus$ is also monotonic. Then, by the well-known Knaster-Tarski fixed point theorem, $x\oplus$ has a least prefix point [1]. Denote the least prefix point of $x\oplus$ by $\mu(x\oplus)$ and consider the function $x \mapsto \mu(x\oplus)$, which we denote by †. The "beautiful theorem" is that † enjoys any type of supremum-preserving property that is enjoyed by the (uncurried binary) function $\oplus$. More precisely, letting $\mathsf{Sup}.f$ denote the supremum[2] of f , then

$$(1) \qquad \forall(f :: \quad \dagger.(\mathsf{Sup}.f) \;=\; \mathsf{Sup}.(\dagger{\bullet}f)) \;\Leftarrow\; \forall(f,g :: \quad (\mathsf{Sup}.f){\oplus}(\mathsf{Sup}.g) \;=\; \mathsf{Sup}.(f \mathbin{\dot{\oplus}} g)) \;,$$

where, by definition, $f \mathbin{\dot{\oplus}} g$ is the function mapping x to $f.x \oplus g.x$ and $\bullet$ denotes function composition.

We begin the paper by presenting a proof of this theorem based on the fact that the existence of all suprema "of shape $\mathcal{B}$" (i.e. suprema of functions f of type $\mathcal{A}{\leftarrow}\mathcal{B}$) is equivalent to the existence of a certain Galois connection. Two other ingredients to the proof are a rule we call the *μ-fusion rule* combining properties of least prefix points with properties of Galois-connected functions, and a rule we call the *μ-abstraction theorem* which states that the function $x \mapsto \mu(x\oplus)$ is itself the least prefix point of a certain

[1] A prefix point of monotonic endofunction f is a point x such that $x \sqsupseteq f.x$.

[2] By definition, the supremum of f is the unique solution of the equation

$$x :: \quad \forall(a :: \quad a \sqsupseteq x \equiv \forall(b :: \quad a \sqsupseteq f.b)) \;.$$

function. The three ingredients to the proof are discussed in separate subsections before being combined to form the proof of the lattice-theoretic “beautiful theorem”.

After a short introduction to category theory explaining and emphasising its “constructively coherent” nature we go on in section 8 to generalise the “beautiful theorem” to this context. In order to make this generalisation we use the correspondence between lattice-theoretic and category-theoretic notions shown in the table below:

Lattice theory concept	is an instance of the category theory concept
preorder	category
monotonic function	functor
(pointwise) ordering between functions	natural transformation between functors
supremum	colimit
least	initial
Galois connection	adjunction
prefix point	algebra

The expert in category theory may find this a frustrating process. Why bother to begin by proving the special case instead of going straight ahead and proving the general case? One response to this question is that it is pure overkill to prove the theorem in category theory if the application one is interested in only demands its use in the context of lattice theory. Those readers who are not interested in category theory may simply disregard the later sections of this paper; they are not required to suffer wading through long and complicated (at least in comparison to the corresponding lattice-theoretic) definitions of categorical concepts, and equally long and complicated (in comparison to the corresponding lattice-theoretic) proofs. The main reason for our presenting the proofs in this way, however, is that the effort expended in seeking a clean, simple, lattice-theoretic proof pays off amply in the subsequent task of formulating and establishing the generalisation in category theory.

The lemmas we establish on the way to proving Dijkstra and Scholten's “beautiful theorem”, in particular the μ-fusion and μ-abstraction rules may well be described as trivial. But their simplicity belies the breadth of their applicability and thus their importance. Several examples of the use of μ-fusion in the context of a study of mathematical induction are presented in [2]; many more examples are included in Voerman's forthcoming thesis [19]. Applications of μ-abstraction are considered in [5]. We are not aware of any publication where the corresponding theorems in category theory are stated. Particular instances of the categorical “beautiful theorem” are known to us but we are also not aware of any publication in which the theorem is stated at the same level of generality.

In fact, in the current paper we omit most proofs. Full proofs are included in [5] but the diligent reader should be able to fill in all proofs for themself provided careful note is taken of the order in which lemmas are presented. In the current paper we also make several severe assumptions in order to focus the presentation. In particular, we assume (co)completeness. This is not necessary but is done in order to avoid encumbering the statement of each lemma and theorem with details of which least fixed points (initial algebras) are assumed to exist and which can be inferred to exist. These simplifying assumptions are not made in the companion paper [5] , the theorems there being stated in full generality.

In the next section we review properties of Galois connections and then recall how suprema in a complete lattice are defined via such a connection. This is followed in section 3 by a discussion of the properties of least prefix points. The μ-fusion rule, which combines properties of Galois connections with properties of least prefix points, is presented and proved. The section is concluded by the proof of the μ-abstraction rule. These three elements (definition of suprema via a Galois connection, μ-abstraction and μ-fusion) are then combined together in section 4 to prove Dijkstra and Scholten's "beautiful theorem".

2 Galois Connections and Suprema

2.1 Pointwise Ordering of Functions

Suppose $\mathcal{A} = (A , \sqsupseteq)$ is a partially ordered set. The set A is the *carrier* and $\sqsupseteq$ is the —reflexive, transitive and anti-symmetric— ordering relation on elements of the carrier. "Partially ordered set" will be abbreviated from now on to "poset". (In fact the assumption that $\mathcal{A}$ is a preorder suffices and fits in better with the discussion of the generalisation to category theory. However, this would introduce additional complications that we prefer to postpone.)

Suppose $\mathcal{B} = (B, \succeq)$ is also a poset. Then, a function $f \in A{\leftarrow}B$ is *monotonic* if $\forall(x,y : \quad x \succeq y : \quad f.x \sqsupseteq f.y)$. We write $f \in \mathcal{A}{\leftarrow}\mathcal{B}$ whenever this is the case. Indeed, we shall restrict our attention throughout to monotonic functions even though in specific cases this is unnecessary. Two trivial, but vital, observations are that the identity function on a set is monotonic, and the composition of two monotonic functions is also monotonic.

Functions f and g, both of type $\mathcal{A}{\leftarrow}\mathcal{B}$, can be ordered pointwise. Specifically, we define the relation $\sqsupseteq_{\mathcal{A}\leftarrow\mathcal{B}}$ on the carrier set $\mathcal{A}{\leftarrow}\mathcal{B}$ by

$$f \sqsupseteq_{\mathcal{A}\leftarrow\mathcal{B}} g \quad \equiv \quad \forall(x :: \quad f.x \sqsupseteq g.x) \quad .$$

As is easily verified, $(\mathcal{A}{\leftarrow}\mathcal{B} , \sqsupseteq_{\mathcal{A}\leftarrow\mathcal{B}})$ is also a partially ordered set.

A very useful strategy —borrowed from category theory— in the study of posets is to "lift" statements about orderings between elements to statements about orderings between functions. We shall adopt this strategy frequently in the sequel. An example is the formulation of monotonicity of a function in terms purely of functions and function composition rather than function application. Specifically, denoting composition of functions by the infix operator $\bullet$, we have (for all $\mathcal{C}$):

$$f \text{ is monotonic} \quad \equiv \quad \forall(g,h : \quad g \sqsupseteq_{\mathcal{B}\leftarrow\mathcal{C}} h : \quad f \bullet g \ \sqsupseteq_{\mathcal{A}\leftarrow\mathcal{C}} \ f \bullet h) \quad .$$

Letting $f\bullet$ denote the function $g \mapsto f \bullet g$ we can express the above yet more strikingly:

$$f \text{ is monotonic } \equiv f\bullet \text{ is monotonic } .$$

Letting $\bullet f$ denote the function $g \mapsto g \bullet f$ we also have:

$$\bullet f \text{ is monotonic } .$$

Rather than subscripting the symbol $\sqsupseteq$ with the type of the functions involved to indicate a pointwise ordering of functions we will sometimes use the symbol $\dot{\sqsupseteq}$ instead (the point serving to remind the reader that the ordering is pointwise). This has the disadvantage that the same symbol is used for different ordering relations, sometimes in one rule, but which ordering is intended will never be ambiguous.

2.2 Galois Connections

The combination of two posets $(A\ ,\ \sqsupseteq)$ and $(B\ ,\ \succeq)$, and two functions, $F \in A{\leftarrow}B$ and $G \in B{\leftarrow}A$, forms a Galois connection if the following formula holds for all $x{\in}A$ and $y{\in}B$.

(2) $\quad x \sqsupseteq F.y \ \equiv\ G.x \succeq y \ .$

Function F will be called the *lower adjoint* and function G the *upper adjoint*. It is easily shown that both are necessarily monotonic, which fact we will tacitly exploit.

The concept of a Galois connection is supposedly "well known", see e.g. [6, 7], but even if it is not well known to the reader it is such a simple and elegant concept that no difficulty should be experienced in verifying any properties that we state without proof.

Instead of stating the definition of a Galois connection in terms of points it is useful to restate it in terms of functions. Specifically, the combination of two posets $(A\ ,\ \sqsupseteq)$ and $(B\ ,\ \succeq)$, and two functions, $F \in A{\leftarrow}B$ and $G \in B{\leftarrow}A$, forms a Galois connection if and only if for all functions h and k of appropriate type,

(3) $\quad h \mathrel{\dot{\sqsupseteq}} F \bullet k \ \equiv\ G \bullet h \mathrel{\dot{\succeq}} k \ .$

Another way of expressing the equivalence of (2) and (3) is

(4) $\quad (F,G) \text{ forms a Galois connection } \equiv (F\bullet, G\bullet) \text{ forms a Galois connection } .$

If two functions are inverses of each other then they are Galois connected. Suppose the inverse functions are F and G. Then we have, for all x in the domain of F, and y in the domain of G,

$$x = F.y \ \equiv\ G.x = y \ .$$

The two poset orderings needed to establish the connection are the trivial orderings whereby the only ordered elements are equal elements.

This observation has no significance whatsoever for a study of inverse functions: nothing can be gained in such a study by instantiating general theorems about Galois connections that is not predicted by much simpler, direct calculations using the fact that a composition of the one function followed by the other is an identity function. The main benefit that is gained from the observation is that it can suggest properties that one might investigate of Galois-connected functions. An important example is that inverse functions have "inverse" algebraic properties. The exponential function, for instance, has as its inverse the logarithmic function, and

$$\exp(-x) \;=\; \frac{1}{\exp x} \qquad \text{whereas} \qquad -\ln x \;=\; \ln(\frac{1}{x})$$

and

$$\exp(x+y) \;=\; \exp x \cdot \exp y \qquad \text{whereas} \qquad \ln x + \ln y \;=\; \ln(x \cdot y) \quad .$$

In general, if F and G are inverse functions then, for any functions h and k of appropriate type,

$$\forall(x :: \quad h.(F.x) \;=\; F.(k.x)) \;\equiv\; \forall(y :: \quad G.(h.y) \;=\; k.(G.y)) \quad .$$

More generally, and expressed at function level, if $(F_0\,,\,G_0)$ and $(F_1\,,\,G_1)$ are pairs of inverse functions, then for all functions h and k of appropriate type,

$$h \bullet F_0 \;=\; F_1 \bullet k \quad\equiv\quad G_1 \bullet h \;=\; k \bullet G_0 \quad .$$

The generalisation to Galois connections takes the following form. Suppose, for $i \;=\; 0\,,1$, the combination of two posets $\mathcal{A}_i \;=\; (A_i\,,\,\sqsupseteq_i)$ and $\mathcal{B}_i \;=\; (B_i\,,\,\succeq_i)$ and two functions $F_i \in \mathcal{A}_i {\leftarrow} \mathcal{B}_i$ and $G_i \in \mathcal{B}_i {\leftarrow} \mathcal{A}_i$ forms a Galois connection. Let $h \in \mathcal{A}_0 {\leftarrow} \mathcal{A}_1$ and $k \in \mathcal{B}_1 {\leftarrow} \mathcal{B}_0$ be arbitrary monotonic functions. Then

$$(5) \qquad h \bullet F_0 \;\dot{\sqsupseteq}\; F_1 \bullet k \quad\equiv\quad G_1 \bullet h \;\dot{\succeq}\; k \bullet G_0 \quad .$$

(As forewarned, subscripts have been omitted from the ordering relations since they can be inferred from the given type information.)

As a useful aide mémoire to property (5) we suggest the slogan "Galois-connected functions have pseudo-inverse algebraic properties".

Property (5) does not seem to be widely known but it is soon learnt and it is particularly useful since it captures in one rule several calculational properties of Galois-connected functions. As well as subsuming (3) it includes the following as special cases: First, by instantiating F_1 and G_1 to the identity function, and F_0 and G_0 to F and G, respectively, we obtain:

$$(F,G) \text{ forms a Galois connection } \Rightarrow (\bullet G, \bullet F) \text{ forms a Galois connection} \quad .$$

(In fact an equivalence can be proved.) Take care to note the switch in the order of F and G. The rule states that if F has *upper* adjoint G then $\bullet F$ has *lower* adjoint $\bullet G$. Second, by instantiating h and k to the identity function,

$$F_0 \;\dot{\sqsupseteq}\; F_1 \quad\equiv\quad G_1 \;\dot{\succeq}\; G_0 \quad .$$

Note this time the switch in the order of subscripts. Hence,

$$F_0 \;=\; F_1 \quad\equiv\quad G_1 \;=\; G_0 \quad .$$

Thus adjoints are uniquely defined.

2.3 Suprema

Defining Galois Connection

Suppose $\mathcal{A} = (A, \sqsupseteq)$ and $\mathcal{B} = (B, \succeq)$ are posets, and $f \in \mathcal{A}\leftarrow\mathcal{B}$ is a monotonic function. Then a *supremum* of f is a solution of the equation

(6) $\quad x :: \quad \forall(a :: \quad a \sqsupseteq x \;\equiv\; \forall(b :: \quad a \sqsupseteq f.b))$.

(Some readers may be more familiar with the definition of a supremum as a least upper bound. That is, x is a supremum of f if it is an upper bound:

$$\forall(b :: \quad x \sqsupseteq f.b) \;\; ,$$

and it is least among such upper bounds

$$\forall(a :: \quad a \sqsupseteq x \Leftarrow \forall(b :: \quad a \sqsupseteq f.b)) \;\; .$$

This is entirely equivalent to (6), as is easily verified.)

It is easily seen that all solutions of (6) are equal. Specifically, for all $a \in A$,

$$\begin{array}{ll}
 & a \sqsupseteq x_0 \\
\equiv & \quad \{ \quad x_0 \text{ solves (6)} \quad \} \\
 & \forall(b :: \quad a \sqsupseteq f.b) \\
\equiv & \quad \{ \quad x_1 \text{ solves (6)} \quad \} \\
 & a \sqsupseteq x_1 \;\; .
\end{array}$$

Hence, by indirect equality,

$$x_0 \text{ and } x_1 \text{ both solve (6)} \quad \Rightarrow \quad x_0 \;=\; x_1 \;\; .$$

Equation (6) need not, of course, have a solution. If it does, for a given f, we denote its solution by $\mathsf{Sup}.f$. By definition, then,

(7) $\quad \forall(a :: \quad a \sqsupseteq \mathsf{Sup}.f \equiv \forall(b :: \quad a \sqsupseteq f.b))$.

Suppose we fix the posets $\mathcal{A}$ and $\mathcal{B}$ and consider all functions of type $\mathcal{A}\leftarrow\mathcal{B}$. Suppose that there is a function Sup mapping all such functions to a supremum. Then we recognise (7) as a Galois connection. Specifically,

$$\begin{array}{ll}
 & \forall(b :: \quad a \sqsupseteq f.b) \\
\equiv & \quad \{ \quad \text{define the function } \mathsf{K} \in (\mathcal{A}\leftarrow\mathcal{B})\leftarrow\mathcal{A} \text{ by} \\
 & \quad\quad\;\; (\mathsf{K}.a).b \;=\; a \quad \} \\
 & \forall(b :: \quad (\mathsf{K}.a).b \sqsupseteq f.b) \\
\equiv & \quad \{ \quad \text{definition of } \dot{\sqsupseteq} \text{ (pointwise ordering on functions)} \quad \} \\
 & \mathsf{K}.a \;\dot{\sqsupseteq}\; f \;\; .
\end{array}$$

Thus, our supposition becomes that there is a function Sup that is the lower adjoint of the so-called "constant combinator" K of type $(\mathcal{A}\leftarrow\mathcal{B})\leftarrow\mathcal{A}$. That is, for all $a\in A$ and $f \in \mathcal{A}\leftarrow\mathcal{B}$,

$$(8) \qquad a \sqsupseteq \mathsf{Sup}.f \quad \equiv \quad \mathsf{K}.a \mathrel{\dot{\sqsupseteq}} f \quad .$$

The poset $\mathcal{B}$ is called the *shape* poset, and the existence of the Galois connection (8) (between the posets $\mathcal{A}$ and $(\mathcal{A}\leftarrow\mathcal{B} \ , \dot{\sqsupseteq})$) is put into words by saying that $\mathcal{A}$ is $\mathcal{B}$-*cocomplete*. If $\mathcal{A}$ is $\mathcal{B}$-cocomplete for all posets $\mathcal{B}$ then we say that $\mathcal{A}$ is *cocomplete*. (Dual to suprema, we may also define infima, leading to the dual notion of *completeness* of $\mathcal{A}$. Completeness and cocompleteness of posets are, however, equivalent and for this reason the "co" is redundant. We retain it, however, in order to emphasise the link with category theory .)

Examples

If $\mathcal{B}$ is the two-point set $\{0 \ , 1\}$ ordered by equality then the set of functions to $\mathcal{A}$ from $\mathcal{B}$ is in (1–1) correspondence with pairs of elements $(a_0 \ , a_1)$ (to be precise: $f\mapsto(f.0 \ , f.1)$ and $(a_0 \ , a_1)\mapsto f$ where $f.0 \ = \ a_0$ and $f.1 \ = \ a_1$). The function $\mathsf{K}.a$ corresponds to the pair $(a \ , a)$, and the ordering relation on two functions both to $\mathcal{A}$ from $\mathcal{B}$ corresponds to the elementwise ordering on the pairs to which the functions correspond. Writing $f.0 \sqcup f.1$ instead of $\mathsf{Sup}.f$, the Galois connection (8) thus corresponds to

$$a \sqsupseteq x{\sqcup}y \quad \equiv \quad (a \ , a) \mathrel{\dot{\sqsupseteq}} (x \ , y) \quad .$$

That is,

$$a \sqsupseteq x{\sqcup}y \quad \equiv \quad a \sqsupseteq x \wedge a \sqsupseteq y \quad .$$

This is the well-known Galois connection defining the supremum of a bag of two elements.

If $\mathcal{B}$ is the empty poset then there is exactly one function of type $\mathcal{A}\leftarrow\mathcal{B}$, namely the identity function (or the empty function, which is the same thing). The right side of (8) is vacuously true and, thus, for all $a\in A$ and $f \in \mathcal{A}\leftarrow\emptyset$,

$$a \sqsupseteq \mathsf{Sup}.f \quad .$$

In words, the poset $\mathcal{A}$ is $\emptyset$-cocomplete equivales $\mathcal{A}$ has a least element.

A third example is the case that $\mathcal{B}$ is $(\mathbb{N}, \leq)$ (the natural numbers ordered in the usual way). Functions in $\mathcal{A}\leftarrow\mathcal{B}$ are then in (1–1) correspondence with chains $a_0 \leq a_1 \leq a_2 \leq ..$. To say that $\mathcal{A}$ is $(\mathbb{N}, \leq)$-cocomplete is equivalent to all such chains having a supremum in $\mathcal{A}$.

2.4 Parameterised Suprema

Suppose $\mathcal{A} \ = \ (A, \sqsupseteq_{\mathcal{A}})$, $\mathcal{B} \ = \ (B, \sqsupseteq_{\mathcal{B}})$ and $\mathcal{C} \ = \ (C, \sqsupseteq_{\mathcal{C}})$ are posets, and suppose $\mathcal{A}$ is $\mathcal{B}$-cocomplete. Motivated by the strategy of lifting statements about points to statements about functions we are naturally led to the question whether the poset $\mathcal{A}{\leftarrow}\mathcal{C}$ is $\mathcal{B}$-cocomplete.

We can answer this question by answering the following question. Consider function $\oplus \in (\mathcal{A}{\leftarrow}\mathcal{C}){\leftarrow}\mathcal{B}$. For brevity we denote application of $\oplus$ to arguments b in $\mathcal{B}$ and c in $\mathcal{C}$ by $b{\oplus}c$. Assume that, for all $c{\in}C$, the function $\oplus c \in \mathcal{A}{\leftarrow}\mathcal{B}$ defined by $(\oplus c).b \;=\; b{\oplus}c$ has a supremum denoted by $\mathsf{Sup}.(\oplus c)$. Note that, since $\oplus c$ is a function of type $\mathcal{A}{\leftarrow}\mathcal{B}$, the existence of $\mathsf{Sup}.(\oplus c)$ is implied by $\mathcal{B}$-cocompleteness of $\mathcal{A}$. The question is whether this information is sufficient to guarantee that $\oplus$ has a supremum.

To express the question formally we need to refine our notation in order to avoid ambiguity. Where before we wrote K let us now write $\mathsf{K}_{\mathcal{A},\mathcal{B}}$. Thus $\mathsf{K}_{\mathcal{A},\mathcal{B}}$ is the function of type $(\mathcal{A}{\leftarrow}\mathcal{B}){\leftarrow}\mathcal{A}$ defined by

$$(\mathsf{K}_{\mathcal{A},\mathcal{B}}.a).b \;=\; a \;\;.$$

The assumption is thus that, for all $c{\in}C$, there is an element $\mathsf{Sup}.(\oplus c)$ in A satisfying

$$(9) \qquad \forall(a :: \quad a \sqsupseteq_{\mathcal{A}} \mathsf{Sup}.(\oplus c) \;\equiv\; \mathsf{K}_{\mathcal{A},\mathcal{B}}.a \;\sqsupseteq_{\mathcal{A}\leftarrow\mathcal{B}}\; \oplus c) \;\;.$$

The goal is to solve the equation

$$(10) \qquad x :: \quad x \in \mathcal{A}{\leftarrow}\mathcal{C} \;\;\wedge\;\; \forall(f :: \quad f \sqsupseteq_{\mathcal{A}\leftarrow\mathcal{C}} x \;\equiv\; \mathsf{K}_{\mathcal{A}\leftarrow\mathcal{C},\mathcal{B}}.f \;\sqsupseteq_{(\mathcal{A}\leftarrow\mathcal{C})\leftarrow\mathcal{B}}\; \oplus) \;\;.$$

The obvious candidate for x is the function $c{\mapsto}\mathsf{Sup}.(\oplus c)$. This function is monotonic (and thus in $\mathcal{A}{\leftarrow}\mathcal{C}$) since it is the composition of two monotonic functions, the function Sup (which is monotonic because the adjoints in a Galois connection are inevitably monotonic) and the function $\hat{\oplus} \in (\mathcal{A}{\leftarrow}\mathcal{B}){\leftarrow}\mathcal{C}$ defined by $(\hat{\oplus}.c).b \;=\; b{\oplus}c$. That it also satisfies (10) is a straightforward calculation. We present it nonetheless in order to show the benefit of exploiting (4).

$$\begin{array}{ll}
 & f \;\sqsupseteq_{\mathcal{A}\leftarrow\mathcal{C}}\; c{\mapsto}\mathsf{Sup}.(\oplus c) \\
\equiv & \qquad \{ \quad \text{see above for the definition of } \hat{\oplus}\text{ ; extensionality} \quad \} \\
 & f \;\sqsupseteq_{\mathcal{A}\leftarrow\mathcal{C}}\; \mathsf{Sup} \bullet \hat{\oplus} \\
\equiv & \qquad \{ \quad \mathsf{Sup} \text{ has upper adjoint } \mathsf{K}_{\mathcal{A},\mathcal{B}}\text{ , by assumption,} \\
 & \qquad\quad\;\; \text{so, by (4), } (\mathsf{Sup}\bullet) \text{ has upper adjoint } (\mathsf{K}_{\mathcal{A},\mathcal{B}}\bullet) \quad \} \\
 & \mathsf{K}_{\mathcal{A},\mathcal{B}} \bullet f \;\sqsupseteq_{(\mathcal{A}\leftarrow\mathcal{C})\leftarrow\mathcal{B}}\; \hat{\oplus} \\
\equiv & \qquad \{ \quad \text{We now try to express } \mathsf{K}_{\mathcal{A},\mathcal{B}} \bullet f \text{ in terms of } \mathsf{K}_{\mathcal{A}\leftarrow\mathcal{C},\mathcal{B}}.f \text{ :} \\
 & \qquad\qquad\;\; ((\mathsf{K}_{\mathcal{A},\mathcal{B}} \bullet f).c).b \\
 & \qquad\qquad = \qquad \{ \quad \text{definition} \quad \} \\
 & \qquad\qquad\;\; f.c \\
 & \qquad\qquad = \qquad \{ \quad \text{definition} \quad \} \\
 & \qquad\qquad\;\; ((\mathsf{K}_{\mathcal{A}\leftarrow\mathcal{C},\mathcal{B}}.f).b).c
\end{array}$$

Thus, defining the function Switch by

$((\mathsf{Switch}.g).c).b \;=\; (g.b).c$, we have

$\mathsf{K}_{\mathcal{A},\mathcal{B}} \bullet f \;=\; \mathsf{Switch}.(\mathsf{K}_{\mathcal{A}\leftarrow\mathcal{C},\mathcal{B}}.f)$ }

$$\mathsf{Switch}.(\mathsf{K}_{\mathcal{A}\leftarrow\mathcal{C},\mathcal{B}}.f) \;\sqsupseteq_{(\mathcal{A}\leftarrow\mathcal{C})\leftarrow\mathcal{B}}\; \hat{\oplus}$$

$\equiv$ { Switch is clearly an order isomorphism, $\hat{\oplus} \;=\; \mathsf{Switch}.\oplus$ }

$$\mathsf{K}_{\mathcal{A}\leftarrow\mathcal{C},\mathcal{B}}.f \;\sqsupseteq_{(\mathcal{A}\leftarrow\mathcal{C})\leftarrow\mathcal{B}}\; \oplus \quad .$$

Thus, we have indeed verified that $c{\mapsto}\mathsf{Sup}.(\oplus c) \;=\; \mathsf{Sup}.\oplus$.

The element $\mathsf{Sup}.(\oplus c)$ is called a *parameterised supremum* and the theorem just proved we call the *abstraction theorem* for suprema. In words it says that the result of abstracting from the parameter in a parameterised supremum is itself a supremum.

Since the existence of a supremum of $(\oplus c)$ for each c is implied by $\mathcal{B}$-cocompleteness of $\mathcal{A}$, a direct corollary is that if $\mathcal{A}$ is $\mathcal{B}$-cocomplete then $\mathcal{A}{\leftarrow}\mathcal{C}$ is also $\mathcal{B}$-cocomplete (independently of $\mathcal{C}$). In particular if $\mathcal{A}$ is (co)complete then $\mathcal{A}{\leftarrow}\mathcal{C}$ is (co)complete.

An example of a parameterised supremum is the binary supremum operator: let $\mathcal{B}$ be $(\{0{,}1\}\;,\;=)$ as in the definition of $\sqcup$ and let $\mathcal{C}$ be $(\mathcal{A}{\times}\mathcal{A}\;,\;\sqsupseteq{\times}\sqsupseteq)$. Define $\oplus \in (\mathcal{A}{\leftarrow}\mathcal{C}){\leftarrow}\mathcal{B}$ by $0{\oplus}(a_0\;,\;a_1) \;=\; a_0$ and $1{\oplus}(a_0\;,\;a_1) \;=\; a_1$. Then, by definition,

$$a_0{\sqcup}a_1 \;=\; (c{\mapsto}\mathsf{Sup}.(\oplus c)).(a_0\;,\;a_1) \quad .$$

Applying the abstraction theorem,

$$(\sqcup) \;=\; \mathsf{Sup}.\oplus \quad .$$

Category theoreticians will recognise in the above a special case of theorem 1 on page 111 of Mac Lane's classic text [14] concerning parameterised (co)limits. Neither the fact that $(\sqcup) \;=\; \mathsf{Sup}.\oplus$ nor the general abstraction theorem for suprema seem to be significant. The corresponding limit theorem in category theory is, however. The proof we have just given, in particular its use of (4), can be easily adapted to an attractive alternative to Mac Lane's proof, as we shall demonstrate in section 6.4.

3 Prefix Points

In this section we assume that $\mathcal{A} \;=\; (A{,}\sqsupseteq)$ and $\mathcal{B} \;=\; (B,\,\succeq)$ are cocomplete lattices, i.e. suprema of all shapes exist in both sets. As previously mentioned, cocompleteness and completeness (the existence of infima of all shapes) coincide for posets, so the assumption is that $\mathcal{A}$ and $\mathcal{B}$ are complete lattices.

3.1 Basic Properties

A *prefix point* of function $f \in \mathcal{A} \leftarrow \mathcal{A}$ is an element $x \in A$ such that $x \sqsupseteq f.x$. A *least prefix point* is a prefix point that is least among such elements. By the Knaster-Tarski theorem least prefix points always exist in complete lattices. Moreover, least elements are unique in posets. Thus, for given complete lattice $\mathcal{A}$ there is a function $\mu_{\mathcal{A}}$ of type $\mathcal{A} \leftarrow (\mathcal{A} \leftarrow \mathcal{A})$ such that

$$\mu_{\mathcal{A}} f \sqsupseteq f.(\mu_{\mathcal{A}} f) \ \wedge \ \forall (x : \quad x \sqsupseteq f.x : \quad x \sqsupseteq \mu_{\mathcal{A}} f) \ .$$

We will invariably omit the subscript $\mathcal{A}$ writing simply μf.

A *fixed point* of function $f \in \mathcal{A} \leftarrow \mathcal{A}$ is an element $x \in A$ such that $x = f.x$. A least prefix point of f is also a fixed point of f, i.e.

$$\mu f \ = \ f.\mu f \ \ ,$$

and thus is a least *fixed* point of f.

3.2 The Fusion Theorem

Given functions $f \in \mathcal{A} \leftarrow \mathcal{B}$, $g \in \mathcal{B} \leftarrow \mathcal{B}$ and $h \in \mathcal{A} \leftarrow \mathcal{A}$ it is easily shown that f maps prefix points of g to prefix points of h provided that $f \bullet g \mathrel{\dot{\sqsupseteq}} h \bullet f$. Denote the set of prefix points of g by $\mathsf{Pre}.g$. Then, formally, we have:

(11) $\quad \forall (x : \quad x \in \mathsf{Pre}.g : \quad f.x \in \mathsf{Pre}.h) \ \Leftarrow \ f \bullet g \mathrel{\dot{\sqsupseteq}} h \bullet f \ .$

In particular, $f.\mu g$ is a prefix point of h. Thus, since μh is the least prefix point of h,

(12) $\quad f.\mu g \ \sqsupseteq \ \mu h \quad \Leftarrow \quad f \bullet g \mathrel{\dot{\sqsupseteq}} h \bullet f \ .$

Instantiating f to the identity function (assuming that $\mathcal{A}$ and $\mathcal{B}$ coincide) it follows that μ is a monotonic function:

$$\mu g \sqsupseteq \mu h \ \Leftarrow \ g \mathrel{\dot{\sqsupseteq}} h \ .$$

By the same argument Pre is an antimonotonic function: for all functions $g,h \in \mathcal{A} \leftarrow \mathcal{A}$:

$$\mathsf{Pre}.g \subseteq \mathsf{Pre}.h \ \Leftarrow \ g \mathrel{\dot{\sqsupseteq}} h \ .$$

Properties (11) and (12) are *not* by themselves so interesting. They become more interesting, however, when we assume that f is the upper adjoint in a Galois connection. Suppose that this is so and let $f^\flat$ denote its lower adjoint. (So $f^\flat \in \mathcal{B} \leftarrow \mathcal{A}$.) Then we recognise in the premise of (12) one side of the "pseudo-invertability" of the algebraic properties of Galois-connected functions —see (5)— , and we can calculate as follows:

$$
\begin{array}{ll}
 & \mu g \;\succeq\; f^\flat.\mu h \\
\equiv & \quad \{ \quad (f^\flat\ ,\ f) \text{ is a Galois connection} \quad \} \\
 & f.\mu g \;\sqsupseteq\; \mu h \\
\Leftarrow & \quad \{ \quad (12) \quad \} \\
 & f\bullet g \;\dot{\sqsupseteq}\; h\bullet f \\
\equiv & \quad \{ \quad \text{Galois-connected functions have pseudo-inverse algebraic} \\
 & \quad\quad\quad \text{properties: (5) with } F_0,F_1,G_0,G_1,h,k \;:=\; f^\flat,f^\flat,f,f,g,h \quad \} \\
 & g\bullet f^\flat \;\dot{\succeq}\; f^\flat\bullet h \quad .
\end{array}
$$

We thus conclude:

(13) $\quad \mu g \;\succeq\; f^\flat.\mu h \;\Leftarrow\; g\bullet f^\flat \;\dot{\succeq}\; f^\flat\bullet h \quad .$

This theorem we call the *basic fusion theorem*. (The superscript $\flat$ in the statement of the theorem is intended to remind you of the requirement that the function $f^\flat$ be a lower adjoint in a Galois connection.)

A final step in this investigation is to enquire when the inclusion in the left side of (13) can be strengthened to an equality. By making the substitutions $f,g,h \;:=\; f^\flat,h,g$ in (11) we obtain immediately:

(14) $\quad \forall(x :\;\; x{\in}\mathsf{Pre}.h :\;\; f^\flat.x \in \mathsf{Pre}.g) \;\Leftarrow\; g\bullet f^\flat \;\dot{\preceq}\; f^\flat\bullet h \quad .$

Combining (11) and (14) with the aid of pseudo-inversality, it is evident that, if $g\bullet f^\flat$ and $f^\flat\bullet h$ are *equal*, the pair of functions $f^\flat$ and f forms a Galois connection between the set of prefix points of g and the set of prefix points of h. In particular, since lower adjoints map least elements to least elements, we have derived the *fusion theorem*:

(15) $\quad \mu g \;=\; f^\flat.\mu h \;\Leftarrow\; g\bullet f^\flat \;=\; f^\flat\bullet h \quad .$

3.3 The Abstraction Theorem

The final topic in this section is the abstraction theorem for prefix points. Suppose $\oplus \in (\mathcal{A}{\leftarrow}\mathcal{A}){\leftarrow}\mathcal{B}$ and $f \in \mathcal{B}{\leftarrow}\mathcal{C}$. Let us adopt the convention once more that application of $\oplus$ to argument b is denoted by $b\oplus$, and application of this function to argument a is denoted by $b{\oplus}a$. Then, for all $c{\in}\mathcal{C}$, $(f.c)\oplus$ is a (monotonic) endofunction on $\mathcal{A}$ and therefore has a least prefix point $\mu((f.c)\oplus)$. Abstracting from the parameter c, we observe that the function $c \mapsto \mu((f.c)\oplus)$ is monotonic, because it is the composition of three monotonic functions — f, $\oplus$ and μ — , and so the function $c \mapsto \mu((f.c)\oplus)$ is in $\mathcal{A}{\leftarrow}\mathcal{C}$. The theorem is that the latter is itself a least prefix point. To be precise, let $\dot{\oplus}$ denote the function of type $((\mathcal{A}{\leftarrow}\mathcal{C}){\leftarrow}(\mathcal{A}{\leftarrow}\mathcal{C})){\leftarrow}(\mathcal{B}{\leftarrow}\mathcal{C})$ defined by

$$(g \mathbin{\dot{\oplus}} h).c \;=\; g.c \oplus h.c \quad .$$

Then $f\dot{\oplus} \in (\mathcal{A}{\leftarrow}\mathcal{C}){\leftarrow}(\mathcal{A}{\leftarrow}\mathcal{C})$ and an easy calculation shows that $c \mapsto \mu((f.c)\oplus)$ is the least prefix point of $f\dot{\oplus}$. First, it is a prefix point:

$$
\begin{array}{ll}
 & c \mapsto \mu((f.c)\oplus) \;\; \dot{\sqsupseteq} \;\; f \mathbin{\dot{\oplus}} (c \mapsto \mu((f.c)\oplus)) \\
\equiv & \quad \{ \quad \text{pointwise ordering, definition of } \dot{\oplus} \quad \} \\
 & \forall(c :: \;\; \mu((f.c)\oplus) \sqsupseteq f.c \oplus \mu((f.c)\oplus)) \\
\equiv & \quad \{ \quad \mu((f.c)\oplus) \text{ is a prefix point of } (f.c)\oplus \quad \} \\
 & \mathit{true} \quad .
\end{array}
$$

Second, it is at most any other prefix point. Suppose g is a prefix point of $f\dot{\oplus}$, i.e. $g \mathrel{\dot{\sqsupseteq}} f \mathbin{\dot{\oplus}} g$. Then

$$
\begin{array}{ll}
 & g \;\; \dot{\sqsupseteq} \;\; c \mapsto \mu((f.c)\oplus) \\
\equiv & \quad \{ \quad \text{pointwise ordering} \quad \} \\
 & \forall(c :: \;\; g.c \sqsupseteq \mu((f.c)\oplus)) \\
\Leftarrow & \quad \{ \quad \mu((f.c)\oplus) \text{ is the least prefix point of } (f.c)\oplus \quad \} \\
 & \forall(c :: \;\; g.c \sqsupseteq f.c \oplus g.c) \\
\equiv & \quad \{ \quad \text{pointwise ordering, definition of } \dot{\oplus} \quad \} \\
 & g \mathrel{\dot{\sqsupseteq}} f \mathbin{\dot{\oplus}} g \\
\equiv & \quad \{ \quad g \text{ is a prefix point of } f\dot{\oplus} \quad \} \\
 & \mathit{true} \quad .
\end{array}
$$

Note that (co)completeness of $\mathcal{A}$ implies (co)completeness of $\mathcal{A}\leftarrow\mathcal{C}$. Applying the Knaster-Tarski theorem there is a function, which we denote by $\dot{\mu}$, that constructs least prefix points of functions of type $(\mathcal{A}\leftarrow\mathcal{C})\leftarrow(\mathcal{A}\leftarrow\mathcal{C})$. The *abstraction theorem* we have proved may therefore be conveniently summarised by the statement:

(16) $\quad c \mapsto \mu((f.c)\oplus) \;=\; \dot{\mu}(f\dot{\oplus}) \quad .$

4 The "Beautiful Theorem"

4.1 The Proof

We are now ready to formulate and prove Dijkstra and Scholten's "beautiful theorem" [8, p. 159].

The theorem is about the preservation of properties of suprema when constructing least prefix points. (There is of course a dual theorem concerning infima and greatest postfix points.) In order to formulate the theorem precisely we need to introduce a few definitions.

Suppose $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are posets, and suppose that $\mathcal{A}$ and $\mathcal{B}$ are both complete. (In most applications $\mathcal{A}$ and $\mathcal{B}$ will be the same.) Let $\mathsf{Sup}_{\mathcal{A}}$ and $\mathsf{Sup}_{\mathcal{B}}$ denote the supremum operators of type $\mathcal{A}\leftarrow(\mathcal{A}\leftarrow\mathcal{C})$ and $\mathcal{B}\leftarrow(\mathcal{B}\leftarrow\mathcal{C})$, respectively. The only fact that we will use about these two operators is that $\mathsf{Sup}_{\mathcal{A}}$ is a lower adjoint in a Galois connection between $\mathcal{A}$ and $\mathcal{A}{\leftarrow}\mathcal{C}$.

Suppose $\oplus \in (\mathcal{A}{\leftarrow}\mathcal{A}){\leftarrow}\mathcal{B}$. Adopting the notational conventions detailed in section 3.3, the function $b \mapsto \mu(b\oplus)$ is thus in $\mathcal{A}{\leftarrow}\mathcal{B}$. Let † denote this function. By definition,

(17) $\quad \dagger.b \;=\; \mu(b\oplus) \quad .$

Our goal is to prove that † "commutes" with Sup if ⊕ "commutes" with Sup. To be precise, we prove that

$$\forall(f :: \;\dagger.(\mathsf{Sup}_{\mathcal{B}}.f) = \mathsf{Sup}_{\mathcal{A}}.(\dagger\bullet f)) \Leftarrow \forall(f,g :: \;(\mathsf{Sup}_{\mathcal{B}}.f)\oplus(\mathsf{Sup}_{\mathcal{A}}.g) = \mathsf{Sup}_{\mathcal{A}}.(f \;\dot{\oplus}\; g)) \quad .$$

Here and elsewhere the dummy f ranges over functions of type $\mathcal{B}{\leftarrow}\mathcal{C}$ and the dummy g over functions of type $\mathcal{A}{\leftarrow}\mathcal{C}$.

We begin our proof with a use of the abstraction theorem designed to get us into a position in which μ-fusion is applicable. For all $f \in \mathcal{B}{\leftarrow}\mathcal{C}$,

$$\begin{array}{ll}
 & \dagger \bullet f \\
= & \qquad \{ \quad \text{extensionality, definition of composition} \quad \} \\
 & c \mapsto \dagger.(f.c) \\
= & \qquad \{ \quad \text{definition of } \dagger\text{: (17)} \quad \} \\
 & c \mapsto \mu((f.c)\oplus) \\
= & \qquad \{ \quad \text{abstraction theorem: (16)} \quad \} \\
 & \dot{\mu}(f\dot{\oplus}) \quad .
\end{array}$$

Thus,

$$\begin{array}{ll}
 & \dagger.(\mathsf{Sup}_{\mathcal{B}}.f) \;=\; \mathsf{Sup}_{\mathcal{A}}.(\dagger\bullet f) \\
\equiv & \qquad \{ \quad \text{definition of } \dagger\text{: (17) applied to the lhs,} \\
 & \qquad\quad\;\; \text{above calculation applied to the rhs} \quad \} \\
 & \mu((\mathsf{Sup}_{\mathcal{B}}.f)\oplus) \;=\; \mathsf{Sup}_{\mathcal{A}}.\dot{\mu}(f\dot{\oplus}) \\
\Leftarrow & \qquad \{ \quad \mu\text{-fusion: (15) (which is applicable since } \mathsf{Sup}_{\mathcal{A}} \text{ is} \\
 & \qquad\quad\;\; \text{a lower adjoint) and extensionality} \quad \} \\
 & \forall(g :: \quad (\mathsf{Sup}_{\mathcal{B}}.f)\oplus(\mathsf{Sup}_{\mathcal{A}}.g) \;=\; \mathsf{Sup}_{\mathcal{A}}.(f \;\dot{\oplus}\; g)) \quad .
\end{array}$$

We have thus proved that

$$\forall(f :: \quad \dagger.(\mathsf{Sup}_{\mathcal{B}}.f) = \mathsf{Sup}_{\mathcal{A}}.(\dagger\bullet f) \;\Leftarrow\; \forall(g :: \quad (\mathsf{Sup}_{\mathcal{B}}.f)\oplus(\mathsf{Sup}_{\mathcal{A}}.g) = \mathsf{Sup}_{\mathcal{A}}.(f \;\dot{\oplus}\; g))) \quad .$$

The "beautiful theorem" follows by elementary calculus.

4.2 Discussion

We introduced this paper with the claim that our own *proof* of the "beautiful theorem" has more right to the adjective "beautiful" than the proof given by Dijkstra and Scholten. The argument for this claim is not based on comparing lengths of proofs —the proof presented in this section is indeed very short but only on account of a substantial amount of preparatory work— but on the way our proof identifies and exploits concepts and properties more fundamental than the "beautiful theorem" itself. We have shown that suprema are defined by a Galois connection and we have made explicit use of this fact by combining it with the μ-fusion and abstraction theorems. The combination of the existence of a specific Galois connection with μ-fusion is a technique that is surprisingly useful, in spite of its beguiling simplicity. What we have shown is that the abstraction theorem is the key to exploiting μ-fusion in the proof of Dijkstra and Scholten's theorem.

Another argument for reformulating the proof is that Dijkstra and Scholten's proof does not lend itself to (straightforward) generalisation to category theory. The fundamental difference is that their proof is a proof by mutual inclusion whereas ours is based strictly on reasoning with equalities. The coming sections are intended to provide the proof of the pudding.

5 Basic Category Theory

In this section we summarise the most elementary notions of category theory — category, functor and natural transformation. We emphasise the viewpoint that a category is a "constructive" preorder such that the "witnesses" to orderings are "coherent". Definitions are taken from [13] with some notational adaptations.

We begin with the definition of a category via the notions of a graph and a deductive system.

A *graph* consists of two classes: the class of *arrows* and the class of *objects* and two mappings from the class of arrows to the class of objects, called *codomain* and *domain*. The codomain and domain mappings are denoted by cod and dom, respectively, and we write $f \in x{\leftarrow}y$ for $\mathsf{cod}.f = x \wedge \mathsf{dom}.f = y$. We also often say that f is *to* x and *from* y. A graph is said to be *small* if the classes of objects and arrows are sets, and *locally small* if for each pair of objects x and y the class of arrows to x from y is a set.

A *deductive system* is a graph in which to each object x there is associated an arrow id_x with codomain and domain both equal to x, i.e.

(18) $\quad \mathsf{id}_x \in x{\leftarrow}x \quad ,$

and to each pair of arrows $f \in x{\leftarrow}y$ and $g \in y{\leftarrow}z$ there is associated an arrow $f{\circ}g$ with codomain x and domain z That is,

(19) $\quad f \in x{\leftarrow}y \;\wedge\; g \in y{\leftarrow}z \;\;\Rightarrow\;\; f{\circ}g \in x{\leftarrow}z \quad .$

The arrow id_x is called the *identity* arrow on x, and $f{\circ}g$ is called the *composition* of f with g.

A *category* is a deductive system in which the following equations hold for all arrows $f \in w \leftarrow x$, $g \in x \leftarrow y$ and $h \in y \leftarrow z$:

$$\mathsf{id}_w \circ f \;=\; f \;=\; f \circ \mathsf{id}_x \quad \wedge \quad (f \circ g) \circ h \;=\; f \circ (g \circ h) \quad .$$

Objects x and y in a category are said to be *isomorphic* if there are arrows $f \in x \leftarrow y$ and $g \in y \leftarrow x$ such that $f \circ g \;=\; \mathsf{id}_x$ and $g \circ f \;=\; \mathsf{id}_y$. If this is the case we write $f \in x \cong y$, $g \in y \cong x$, or just $x \cong y$.

Suppose we define the relation $\sqsupseteq$ on objects in a deductive system by

$$x \sqsupseteq y \equiv \exists (f :: \quad f \in x \leftarrow y) \quad .$$

Then $\sqsupseteq$ is reflexive (by (18)) and transitive (by (19)). Thus the objects of a (small) deductive system form a preordered set, and classes of isomorphic objects a poset, under this relation. Borrowing jargon from constructive type theory we can read " $f \in x \leftarrow y$ " as "f *witnesses* the ordering $x \sqsupseteq y$ ". In this sense, category theory is constructive: it is a theory about how to construct such witnesses rather than a theory about the existence of the witnesses.

The axioms of a category say that witnesses must be *coherent*. Consider the identity axiom: the requirement that $\mathsf{id}_w \circ f \;=\; f \;=\; f \circ \mathsf{id}_x$ whenever $f \in w \leftarrow x$. Assuming $w \sqsupseteq x$ there are (at least) three ways of concluding that $w \sqsupseteq x$: immediately, by arguing that $w \sqsupseteq w$ (by reflexivity) and $w \sqsupseteq x$ (by assumption) and hence by transitivity $w \sqsupseteq x$, or by arguing that $w \sqsupseteq x$ (by assumption) and $x \sqsupseteq x$ (by reflexivity) and hence by transitivity $w \sqsupseteq x$. If we augment these arguments with the construction of witnesses to the ordering we obtain f, $\mathsf{id}_w \circ f$ and $f \circ \mathsf{id}_x$. The identity axiom says that these must be the same. A similar argument applies to the associativity axiom: the requirement that $(f \circ g) \circ h \;=\; f \circ (g \circ h)$ whenever $f \in w \leftarrow x$, $g \in x \leftarrow y$ and $h \in y \leftarrow z$. Specifically, given that $w \sqsupseteq x$, $x \sqsupseteq y$ and $y \sqsupseteq z$ there are two different ways we can combine the orderings into the ordering $w \sqsupseteq z$: we may first combine $w \sqsupseteq x$ and $x \sqsupseteq y$ by transitivity to obtain $w \sqsupseteq y$ and then combine the latter with $y \sqsupseteq z$, or we may begin by combining $x \sqsupseteq y$ and $y \sqsupseteq z$ to obtain $x \sqsupseteq z$ and then combine this with $w \sqsupseteq x$. If we augment these arguments with the construction of witnesses to the ordering we obtain $(f \circ g) \circ h$ and $f \circ (g \circ h)$, respectively.

The definition of a category is decomposed into three parts: a graph, a deductive system and, finally, a category. Our own understanding of category theory is that all concepts in the theory admit the same decomposition: Underlying the category-theory concept is a concept in lattice theory; to this is added a constructive element, a mechanism for building "witnesses" to orderings; finally, the construction of witnesses is required to be "coherent".

Possibly the simplest illustration is the notion of an *initial* object in a category. Formally, an initial object in a category is an object x such that, for each object y in the category, there is a unique arrow to y from x. The corresponding concept in lattice theory is *least* element. The constructive element in the definition of initial object is that there *exists* an arrow from the object to each object in the category. The coherence requirement is the *uniqueness* of such arrows. We will denote the unique arrow to an object y from an initial object x by $(\!|x \;=:\; y|\!)$.

The category-theoretic concept corresponding to the notion of monotonic function is the notion of (covariant) *functor*. A functor F to category $\mathcal{A}$ from category $\mathcal{B}$ is a function mapping objects of $\mathcal{B}$ to objects of $\mathcal{A}$ and arrows of $\mathcal{B}$ to arrows of $\mathcal{A}$. The constructive element of the definition is that if $f \in x{\leftarrow}y$ is an arrow in $\mathcal{B}$ then $F.f \;\in\; F.x \leftarrow F.y$. (F is thus monotone in the sense that if $x \sqsupseteq y$ then $F.x \sqsupseteq F.y$ but, in addition, F constructs from the witness f to the former ordering a witness $F.f$ to the latter ordering.) The coherence requirement is that

$$\mathsf{id}_{F.x} \;=\; F.\mathsf{id}_x$$

for all objects x, and

$$F.(f \circ g) \;=\; F.f \circ F.g$$

whenever f and g are composable arrows. The first requirement expresses the coherence of two different proofs of $F.x \sqsupseteq F.x$: a direct application of reflexivity (to $F.x$), and an application of reflexivity (to x) followed by the monotonicity of F. The second requirement expresses the coherence of two different proofs of $F.x \sqsupseteq F.z$ given that $x \sqsupseteq y$ and $y \sqsupseteq z$: either apply transitivity first followed by monotonicity of F, or apply monotonicity of F twice and then apply transitivity.

For each category $\mathcal{A}$ there is an identity functor defined in the obvious way. We use $\mathsf{Id}_{\mathcal{A}}$ to denote this functor.

A *contravariant functor* corresponds to an antimonotonic function. Just as in lattice theory, there is no theoretical need to identify the notion of contravariant functor[3]. It is nevertheless sometimes more convenient to do so. A contravariant functor satisfies the same requirements as a *covariant* functor except that the requirement

$$F.f \;\in\; F.x \leftarrow F.y \quad\Leftarrow\quad f \in x{\leftarrow}y$$

is replaced by

$$F.f \;\in\; F.y \leftarrow F.x \quad\Leftarrow\quad f \in x{\leftarrow}y \quad ,$$

and the requirement

$$F.(f \circ g) \;=\; F.f \circ F.g$$

is replaced by

$$F.(f \circ g) \;=\; F.g \circ F.f \quad .$$

[3]In lattice theory an antimonotonic function with domain $(A, \sqsupseteq)$ is a monotonic function if viewed as having domain $(A, \sqsubseteq)$. One just turns the ordering around in the domain (or in the codomain). Correspondingly, in category theory one can replace the domain of the functor by the opposite category, the category with the same class of objects and the same class of arrows but with the domain and codomain functions turned around.

(Note the reversal of x and y, and of f and g.)

The category-theoretic concept corresponding to the pointwise ordering between monotonic functions is the notion of *natural transformation* between functors. Suppose F and G are both functors in $\mathcal{A}{\leftarrow}\mathcal{B}$. Then α is a natural transformation to F from G, written $\alpha \in F{\leftarrow}G$ if α is a family of arrows, one for each object $x{\in}\mathcal{B}$, such that

$$\forall(x : \quad x{\in}\mathcal{B} : \quad \alpha_x \ \in \ F.x \leftarrow G.x) \ \ ,$$

and for all arrows $f \in x{\leftarrow}y$ in the category $\mathcal{B}$,

$$F.f \circ \alpha_y \ = \ \alpha_x \circ G.f \ \ .$$

The first of these requirements formulates the existence of a constructive pointwise ordering between the functors, and the second that two proofs of $F.x \sqsupseteq G.y$ given that $x \sqsupseteq y$ are coherent (begin by applying either the monotonicity of F or the monotonicity of G and then combine the result with the ordering between F and G).

6 Adjunctions and Colimits

6.1 Functor Category

Given posets $\mathcal{A}$ and $\mathcal{B}$, the set of monotonic functions to $\mathcal{A}$ from $\mathcal{B}$ forms a partially ordered set under the pointwise ordering of functions. Correspondingly, let $\mathcal{C}$ and $\mathcal{D}$ be categories. Then $\mathsf{Fun}(\mathcal{C},\mathcal{D})$, constructed as follows, is a category. The objects are the functors to $\mathcal{C}$ from $\mathcal{D}$ and the arrows are the natural transformations. Composition of arrows $\eta \in F{\leftarrow}G$ and $\tau \in G{\leftarrow}H$ is defined by

(20) $\quad (\eta \circ \tau)_x \ = \ \eta_x \circ \tau_x$, where $x{\in}\mathcal{D}$.

The identity arrows are the identity transformations. I.e. $\mathsf{id}F \in F{\leftarrow}F$ is defined by

(21) $\quad (\mathsf{id}F)_x \ = \ \mathsf{id}_{F.x}$, where $x{\in}\mathcal{D}$.

Later we use the notation $\mathsf{End}.\mathcal{C}$ to denote $\mathsf{Fun}(\mathcal{C},\mathcal{C})$. The objects of $\mathsf{End}.\mathcal{C}$ are called the *endofunctors* on $\mathcal{C}$.

If $f \in \mathcal{A}{\leftarrow}\mathcal{B}$ is a monotonic function then both $f\bullet$ of type $(\mathcal{A}{\leftarrow}\mathcal{C}){\leftarrow}(\mathcal{B}{\leftarrow}\mathcal{C})$ and $\bullet f$ of type $(\mathcal{C}{\leftarrow}\mathcal{B}){\leftarrow}(\mathcal{C}{\leftarrow}\mathcal{A})$ are monotonic functions. Correspondingly, if $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are categories and $F \in \mathcal{A}{\leftarrow}\mathcal{B}$ is a functor we can define functors $F\bullet \in \mathsf{Fun}(\mathcal{A},\mathcal{C}){\leftarrow}\mathsf{Fun}(\mathcal{B},\mathcal{C})$ and $\bullet F \in \mathsf{Fun}(\mathcal{C},\mathcal{B}){\leftarrow}\mathsf{Fun}(\mathcal{C},\mathcal{A})$ as follows: on functors $G \in \mathcal{B}{\leftarrow}\mathcal{C}$ (i.e. the objects of $\mathsf{Fun}(\mathcal{B},\mathcal{C})$) $F\bullet$ is defined in the obvious way:

$$(F\bullet).G \ = \ F \bullet G \ \ .$$

On natural transformations $\eta \in G{\leftarrow}H$ (where both $G{,}H \in \mathcal{B}{\leftarrow}\mathcal{C}$ — thus the arrows of $\mathsf{Fun}(\mathcal{B},\mathcal{C})$) application of $F\bullet$ is defined by:

$$((F\bullet).\eta)_x \ = \ F.(\eta_x) \text{ for all } \ x{\in}\mathcal{C} \ .$$

Similarly, on functors $G \in \mathcal{C} \leftarrow \mathcal{A}$ (thus the objects of $\mathsf{Fun}(\mathcal{C},\mathcal{A})$) $\bullet F$ is defined in the obvious way:

$$(\bullet F).G \;=\; G \bullet F \quad .$$

On natural transformations $\eta \in G \leftarrow H$ (where both $G{,}H \in \mathcal{C} \leftarrow \mathcal{A}$, i.e. the arrows of $\mathsf{Fun}(\mathcal{C},\mathcal{A})$) it is defined by:

$$((\bullet F).\eta)_x \;=\; \eta_{F.x} \text{ for all } x{\in}\mathcal{C} \;.$$

(Checking that these definitions do indeed define functors is tedious but straightforward.)

The conventional notation for the application of $F\bullet$ to natural transformation η is $F\eta$, and of $\bullet F$ is ηF in line with the convention that composition of functors is denoted by their juxtaposition —see for example [13]— . We shall, however, write $F{\bullet}\eta$ and $\eta{\bullet}F$, respectively, thus keeping the composition operator visible.

Having identified the functions $f\bullet$ and $\bullet f$ as entities in their own right leads to a multitude of ways of expressing the associativity of composition, for example:

$$f \bullet g \bullet h \;=\; f\bullet \,.\, (g\bullet \,.\, h) \;=\; f\bullet \,.\, (\bullet h \,.\, g) \;=\; \bullet h \,.\, (f\bullet \,.\, g) \;=\; \bullet h \,.\, (\bullet g \,.\, f) \quad .$$

These lead to a correspondingly large number of coherence properties relating natural transformations of the form $F{\bullet}\eta$ and $\eta{\bullet}F$, commonly referred to as Godement's rules.

6.2 Adjunctions

The category-theoretic concept corresponding to the notion of a Galois connection is that of an adjunction. There is a large number of equivalent definitions of an adjunction. The one we give here (see [9, 13]) is particularly well-suited to our goal.

Guided by the pseudo-inversality property of Galois connections we begin with a discussion that is more general than is strictly necessary to define an adjunction.

Suppose $F \in \mathcal{A} \leftarrow \mathcal{B}$ and $G \in \mathcal{A} \leftarrow \mathcal{C}$ are arbitrary functors. (These functors will remain fixed throughout this discussion.) Consider the function mapping $x{\in}\mathcal{B}$ and $y{\in}\mathcal{C}$ to the set[4] $F.x \overset{\mathcal{A}}{\leftarrow} G.y$ (the arrows in the category $\mathcal{A}$ to $F.x$ from $G.y$). This function can be extended to a binary functor that is covariant in the argument x and contravariant in the argument y. The codomain of the functor is the category **Set** (with objects sets, arrows functions, and function composition as the composition operator). Its domain is the cartesian product of the category $\mathcal{B}$ and (the opposite of) category $\mathcal{C}$. On arrows $h \in r \overset{\mathcal{B}}{\leftarrow} s$ and $k \in t \overset{\mathcal{C}}{\leftarrow} u$ we define $F.h \leftarrow G.k$ by

$$(F.h \leftarrow G.k).m \;=\; F.h \circ m \circ G.k \quad ,$$

for all $m \;\in\; F.s \overset{\mathcal{A}}{\leftarrow} G.t$. Straightforwardly,

$$F.h \leftarrow G.k \quad\in\quad (F.r \overset{\mathcal{A}}{\leftarrow} G.u) \leftarrow (F.s \overset{\mathcal{A}}{\leftarrow} G.t) \quad .$$

[4]For convenience we assume that the category is locally small. That is, the arrows between a pair of objects form a set. This assumption is not essential and can be circumvented.

(Note the switch in the order of t and u. The unlabelled arrow on the right denotes a class of arrows in the category **Set**. So, $F.h \leftarrow G.k$ is a function.) Moreover,

$$(F.\mathsf{id}_s \leftarrow G.\mathsf{id}_u).m \;=\; m \quad ,$$

and

$$(F.h \leftarrow G.k) \bullet (F.m \leftarrow G.n) \;=\; F.(h \circ m) \leftarrow G.(n \circ k) \quad ,$$

for all h, k, m and n of suitable type. (Note the switch in the order of k and n.) Thus we have indeed defined a functor. Let us denote this functor by $[x,y :: \;\; F.x \leftarrow G.y]$, the square brackets indicating abstraction from the dummies x and y, which range over objects and arrows of the category $\mathcal{A}$. (Category theoreticians will have no difficulty in recognising this functor as the composition of the hom functor $\mathcal{A}(+,-)$ after the cartesian product of the functor F and the functor G^{op} .)

Now we return to the definition of an adjunction. Suppose once more that F and G are functors but this time of type $\mathcal{A} \leftarrow \mathcal{B}$ and $\mathcal{B} \leftarrow \mathcal{A}$, respectively. Then $[x,y :: \;\; \mathsf{Id}_{\mathcal{A}}.x \leftarrow F.y]$ and $[x,y :: \;\; G.x \leftarrow \mathsf{Id}_{\mathcal{B}}.y]$ are functors of the same type. If they are isomorphic then F and G are said to be *adjoint functors*. Writing $[x,y :: \;\; x \leftarrow F.y]$ and $[x,y :: \;\; G.x \leftarrow y]$ instead of $[x,y :: \;\; \mathsf{Id}_{\mathcal{A}}.x \leftarrow F.y]$ and $[x,y :: \;\; G.x \leftarrow \mathsf{Id}_{\mathcal{B}}.y]$ we therefore have the definition:

(22) $\quad (F,G)$ is an adjunction $\equiv [x,y :: \;\; x \leftarrow F.y] \cong [x,y :: \;\; G.x \leftarrow y]$.

Functor F is called the *lower adjoint*, and functor G the *upper adjoint*.

As was the case for Galois connections, it is useful to "lift" the definition of an adjunction one level higher. Specifically, recalling that $F\bullet$ and $G\bullet$ are both functors (given that F and G are functors) and that $F\bullet y$ denotes application of $F\bullet$ to y (which may be a functor or a natural transformation), we have:

(23) $\quad (F,G)$ is an adjunction $\equiv [x,y :: \;\; x \leftarrow F\bullet y] \cong [x,y :: \;\; G\bullet x \leftarrow y]$.

Property (23) is just a stepping stone to the pseudo-inversality property of adjoint functors. Suppose, for $i \;=\; 0\;,1$, $\mathcal{A}_i$ and $\mathcal{B}_i$ are categories, and the pair of functors $F_i \in \mathcal{A}_i \leftarrow \mathcal{B}_i$ and $G_i \in \mathcal{B}_i \leftarrow \mathcal{A}_i$ forms an adjunction. Then (recalling that application of the functor $\bullet F$ to x is denoted by $x\bullet F$) we have:

(24) $\quad [x,y :: \;\; x\bullet F_0 \leftarrow F_1\bullet y] \;\cong\; [x,y :: \;\; G_1\bullet x \leftarrow y\bullet G_0]$

You should compare (22) and (24) with (2) and (5).

Let us spell out some of the details of (22), (23) and (24). Unfolding the definition of an isomorphism in a functor category (22) states that there are two, inverse, natural transformations between the functors $[x,y :: \;\; x \leftarrow F.y]$ and $[x,y :: \;\; G.x \leftarrow y]$. Denoting them by the parentheses $\lfloor\;\rfloor$ and $\lceil\;\rceil$, we have: first, for each $x{\in}\mathcal{A}$ and $y{\in}\mathcal{B}$, $\lfloor\;\rfloor_{x,y}$ and $\lceil\;\rceil_{x,y}$ are both functions of the following types:

(25) $\quad f \in x \overset{\mathcal{A}}{\leftarrow} F.y \;\Rightarrow\; \lfloor f \rfloor_{x,y} \in G.x \overset{\mathcal{B}}{\leftarrow} y \quad ,$

and

(26) $\quad \lceil g \rceil_{x,y} \;\in\; x \overset{\mathcal{A}}{\longleftarrow} F.y \;\;\Leftarrow\;\; g \;\in\; G.x \overset{\mathcal{B}}{\longleftarrow} y \quad .$

Second, their being inverses is formulated by —omitting the subscripts x and y—

(27) $\quad \lfloor f \rfloor \;=\; g \;\;\equiv\;\; f \;=\; \lceil g \rceil \quad .$

Third, their being natural transformations (i.e. arrows in a functor category) boils down, after some unfolding of definitions, to:

(28) $\quad \lfloor f \circ g \circ F.h \rfloor \;=\; G.f \circ \lfloor g \rfloor \circ h \quad ,$

and

(29) $\quad \lceil G.f \circ g \circ h \rceil \;=\; f \circ \lceil g \rceil \circ F.h \quad ,$

for all f, g, and h of appropriate type. We shall refer to the natural transformations $\lfloor\;\rfloor$ and $\lceil\;\rceil$ as the *lower* and *upper adjungates*.

Given that F and G are adjoint functors, the adjungates $\llbracket\;\rrbracket$ and $\llbracket\;\rrbracket$ of the adjunction $(F\bullet\, ,\, G\bullet)$ are of the following type:

$$\eta \;\in\; K \leftarrow F\bullet H \quad\Rightarrow\quad \llbracket \eta \rrbracket_{K,H} \;\in\; G\bullet K \leftarrow H \quad ,$$

$$\llbracket \tau \rrbracket_{K,H} \;\in\; K \leftarrow F\bullet H \quad\Leftarrow\quad \tau \;\in\; G\bullet K \leftarrow H \quad .$$

The relationship between the adjungates of the adjunction $(F\, ,\, G)$ and the adjungates of the adjunction $(F\bullet\, ,\, G\bullet)$ can easily be constructed. Let η and τ be natural transformations as defined above, and suppose the lower and upper adjungates of the adjunction (F,G) are denoted by $\lfloor\;\rfloor$ and $\lceil\;\rceil$, then —omitting the subscripts K and H for greater clarity—

$$\llbracket \eta \rrbracket_x \;=\; \lfloor \eta_x \rfloor \quad\text{and}\quad \llbracket \tau \rrbracket_y \;=\; \lceil \tau_y \rceil \quad ,$$

for all x and y (the type of which will depend on the type of K). The construction of an adjunction $(F\, ,\, G)$ given the adjunction $(F\bullet\, ,\, G\bullet)$ is a simple matter of instantiating the functors K and H with the constant functors $\mathsf{K}.x$ and $\mathsf{K}.y$. This half of the equivalence in (23) is not so interesting so the details are omitted.

Turning to (24), it is useful to note first that an equivalent definition of an adjunction is a sextuple consisting of two categories $\mathcal{A}$ and $\mathcal{B}$, two functors $F \in \mathcal{A}\leftarrow\mathcal{B}$ and $G \in \mathcal{B}\leftarrow\mathcal{A}$, and two natural transformations unit and counit such that

$$\mathsf{unit} \;\in\; G\bullet F \leftarrow \mathsf{Id}_{\mathcal{B}} \quad ,$$

$$\mathsf{counit} \;\in\; \mathsf{Id}_{\mathcal{A}} \leftarrow F\bullet G \quad ,$$

and the following coherence requirement is satisfied:

$$G.f \circ \mathsf{unit}_x \;=\; g \;\;\equiv\;\; f \;=\; \mathsf{counit}_y \circ F.g \quad .$$

(Readers familiar with Galois connections will immediately recognise a generalisation of a standard theorem.)

Now suppose, for $i = 0,1$, $\mathcal{A}_i$ and $\mathcal{B}_i$ are categories and $F_i \in \mathcal{A}_i \leftarrow \mathcal{B}_i$ and $G_i \in \mathcal{B}_i \leftarrow \mathcal{A}_i$ are functors. Suppose that $(F_0\ ,\ G_0)$ is an adjunction; let the unit and counit of this adjunction be denoted by unit and counit, respectively. Suppose in addition that $(F_1\ ,\ G_1)$ is an adjunction; let the lower and upper adjungates of the adjunction $(F_1\bullet\ ,\ G_1\bullet)$ be denoted by $\llbracket\ \rrbracket$ and $\llbracket\ \rrbracket$, respectively. Then the isomorphism expressed in (24) is witnessed by two families of functions: the first is such that, for all functors $H \in \mathcal{B}_1 \leftarrow \mathcal{B}_0$ and $K \in \mathcal{A}_1 \leftarrow \mathcal{A}_0$, the natural transformation $\tau \in G_1\bullet K \leftarrow H\bullet G_0$ is mapped to $\llbracket (\tau\bullet F_0) \circ (H\bullet\mathsf{unit}) \rrbracket$. The second is such that natural transformation $\eta \in K\bullet F_0 \leftarrow F_1\bullet H$ is mapped to $\llbracket (K\bullet\mathsf{counit}) \circ (\eta\bullet G_0) \rrbracket$.

The size of the expressions describing these mappings is a precise indicator of the length of the proof of the pseudo-inversality property (5) of Galois connections. This is because there is a (1–1) correspondence between subterms in the two expressions and steps in the proof of (5): the process used to construct the witnesses to the isomorphism (24) from a proof of (5) is a handle-turning exercise requiring no creativity whatsoever. There is, however, much more to (24) than just the construction of witnesses. If one unfolds all the definitions involved in its formulation then one discovers that it encapsulates a very great deal of detail. Accordingly, its proof requires quite some effort. Our own proof [5], including the proof of (23) as lemma, extends to three pages of detailed calculations. The point is, however, that its formulation is inspired by the usefulness of (5), and knowledge of the proof of the latter makes the proof of (24) a mechanical exercise.

6.3 Colimits

The category-theoretic notion corresponding to the notion of supremum is *colimit*. In order to define colimits we need to first define the "diagonal" *functor*[5] K in place of the *function* K . This is straightforward enough. Given categories $\mathcal{A}$ and $\mathcal{B}$ we define $\mathsf{K}.x$, for each x in $\mathcal{A}$, to be the functor that maps objects y in $\mathcal{B}$ to x, and arrows $g \in y \overset{\mathcal{B}}{\longleftarrow} z$ to the arrow $\mathsf{id}_x \in (\mathsf{K}.x).y \overset{\mathcal{A}}{\longleftarrow} (\mathsf{K}.x).z$. It is left to the reader to check that this does indeed define $\mathsf{K}.x$ to be a functor of type $\mathcal{A}\leftarrow\mathcal{B}$ (in particular that the coherence properties are satisfied). Now we extend the definition of K to arrows in the category $\mathcal{A}$: if $f \in w \overset{\mathcal{A}}{\longleftarrow} x$ we define $\mathsf{K}.f$ to be a natural transformation to the functor $\mathsf{K}.w$ from the functor $\mathsf{K}.x$ by letting $(\mathsf{K}.f)_y$ equal f for all objects y in the category $\mathcal{B}$. It is left to the reader to check that this does indeed define a natural transformation. The conclusion is that $\mathsf{K} \in (\mathcal{A}\leftarrow\mathcal{B})\leftarrow\mathcal{A}$, i.e. K is a functor from the category $\mathcal{A}$ to the functor category $\mathcal{A}\leftarrow\mathcal{B}$.

[5]The name "diagonal functor" is that used by Mac Lane [14, p67]. Mac Lane uses the symbol $\triangle$, however, instead of K. Confusingly, on p. 62 Mac Lane also uses the name "diagonal functor" and the symbol $\triangle$ for the functor that doubles its argument (the justification presumably being that the former generalises the latter). The functor K defined here is the more general of the two; the "doubling" functor $\triangle$ is introduced later.

Suppose $F \in \mathcal{A}{\leftarrow}\mathcal{B}$. A *colimit* of F is a solution of the equation:

$$(30) \qquad x :: \;\; [a :: \;\; a{\leftarrow}x] \cong [a :: \;\; \mathsf{K}.a \leftarrow F] \;\; .$$

Here both dummies a range over objects and arrrows of $\mathcal{A}$. The functors $[a :: \;\; a{\leftarrow}x]$ and $[a :: \;\; \mathsf{K}.a \leftarrow F]$ are defined according to a minor variation on the definition of the functor $[x,y :: \;\; F.x \leftarrow G.y]$ introduced in section 6.2. Specifically, for all $a{\in}\mathcal{A}$, $[a :: \;\; a{\leftarrow}x].a$ is the set of arrows to a from x in the category $\mathcal{A}$, and $[a :: \;\; \mathsf{K}.a \leftarrow F].a$ is the set of all arrows to $\mathsf{K}.a$ from F in the category $\mathcal{A}{\leftarrow}\mathcal{B}$; also, for all $f \in u \overset{\mathcal{A}}{\leftarrow} v$, $[a :: \;\; a{\leftarrow}x].f$ is the function $f\circ$ (mapping arrows to v from x into arrows to u from x) and $[a :: \;\; \mathsf{K}.a \leftarrow F].f$ is the function $(\mathsf{K}.f)\circ$. (Unfolding the definition of $\mathsf{K}.f$, the function $(\mathsf{K}.f)\circ$ maps natural transformation $\eta \in \mathsf{K}.v \leftarrow F$ to the natural transformation $\mathsf{K}.f \circ \eta \in \mathsf{K}.u \leftarrow F$ where, for each object y, $(\mathsf{K}.f \circ \eta)_y = f \circ \eta_y$.)

The category $\mathcal{A}$ is said to be *$\mathcal{B}$-cocomplete* iff the diagonal functor $\mathsf{K} \in (\mathcal{A}{\leftarrow}\mathcal{B}){\leftarrow}\mathcal{A}$ has a lower adjoint. Thus, $\mathcal{A}$ is *$\mathcal{B}$-cocomplete* if there is a functor $\mathsf{Col} \in \mathcal{A}{\leftarrow}(\mathcal{A}{\leftarrow}\mathcal{B})$ such that

$$(31) \qquad [a,F :: \;\; a{\leftarrow}\mathsf{Col}.F] \;\cong\; [a,F :: \;\; \mathsf{K}.a \leftarrow F] \;\; .$$

The dummies a and F in (31) range over objects of the category $\mathcal{A}$ and objects of the category $\mathcal{A}{\leftarrow}\mathcal{B}$ (i.e. functors to $\mathcal{A}$ from $\mathcal{B}$), respectively.

6.4 Parameterised Colimits

Just as in lattice theory if (poset) $\mathcal{A}$ is $\mathcal{B}$-cocomplete then $\mathcal{A}{\leftarrow}\mathcal{C}$ is $\mathcal{B}$-cocomplete for all (posets) $\mathcal{C}$ it is the case that if category $\mathcal{A}$ is $\mathcal{B}$-cocomplete then the functor category $\mathsf{Fun}(\mathcal{A},\mathcal{C})$ is $\mathcal{B}$-cocomplete for all categories $\mathcal{C}$. As mentioned earlier, this is theorem 1 on p.111 of [14]. Let us present an alternative proof to illustrate the process of transforming proofs in lattice theory into proofs in category theory. The reader is encouraged to place a copy of section 2.4 alongside this subsection whilst reading on.

Suppose $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are categories. Suppose $\oplus \in (\mathcal{A}{\leftarrow}\mathcal{C}){\leftarrow}\mathcal{B}$. For brevity we denote application of $\oplus$ to arguments b in $\mathcal{B}$ and c in $\mathcal{C}$ by $b{\oplus}c$. Assume that, for all $c{\in}\mathcal{C}$, the functor $\oplus c \in \mathcal{A}{\leftarrow}\mathcal{B}$ defined by $(\oplus c).b = b{\oplus}c$ has a colimit denoted by $\mathsf{Col}.(\oplus c)$. Note that, since $\oplus c$ is a functor of type $\mathcal{A}{\leftarrow}\mathcal{B}$, the existence of $\mathsf{Col}.(\oplus c)$ is implied by $\mathcal{B}$-cocompleteness of $\mathcal{A}$. The question is whether this information is sufficient to guarantee that $\oplus$ has a colimit.

As in section 2.4 we need to differentiate between diagonal functors: where in section 6.3 we wrote K let us now write $\mathsf{K}_{\mathcal{A},\mathcal{B}}$. The assumption is thus that, for all $c{\in}\mathcal{C}$, there is an object $\mathsf{Col}.(\oplus c)$ in $\mathcal{A}$ satisfying

$$(32) \qquad [a :: \;\; a{\leftarrow}\mathsf{Col}.(\oplus c)] \;\cong\; [a :: \;\; \mathsf{K}_{\mathcal{A},\mathcal{B}}.a \leftarrow \oplus c] \;\; ,$$

The goal is to solve the equation:

$$(33) \qquad x :: \quad x \in \mathcal{A}{\leftarrow}\mathcal{C} \;\;\wedge\;\; [f :: \;\; f{\leftarrow}x] \cong [f :: \;\; \mathsf{K}_{\mathcal{A}\leftarrow\mathcal{C},\mathcal{B}}.f \leftarrow \oplus] \;\; .$$

A candidate for x is obtained by extending the mapping $c{\mapsto}\mathsf{Col}.(\oplus c)$ on objects of the category $\mathcal{C}$ to a functor. This is easily done if we observe that $c{\mapsto}\mathsf{Col}.(\oplus c)$ coincides with the object part of the functor $\mathsf{Col} \bullet \mathsf{Switch}.\oplus$ where Switch is the functor that "switches" the order of the arguments of a functor of type $(\mathcal{A}{\leftarrow}\mathcal{C}){\leftarrow}\mathcal{B}$ to form a functor of type $(\mathcal{A}{\leftarrow}\mathcal{B}){\leftarrow}\mathcal{C}$. (That is, $\mathsf{Switch}.\oplus$ is the functor $\hat{\oplus}$ of type $(\mathcal{A}{\leftarrow}\mathcal{B}){\leftarrow}\mathcal{C}$ such that $b{\oplus}c \;=\; c\hat{\oplus}b$.) Accordingly we verify that $\mathsf{Col} \bullet \mathsf{Switch}.\oplus$ satisfies (33). It clearly satisfies the first conjunct (since it is a composition of two functors of types $\mathcal{A}{\leftarrow}(\mathcal{A}{\leftarrow}\mathcal{B})$ and $(\mathcal{A}{\leftarrow}\mathcal{B}){\leftarrow}\mathcal{C}$). For the second conjunct we have:

$$
\begin{array}{ll}
 & [f :: \quad f \;\leftarrow\; \mathsf{Col} \bullet \mathsf{Switch}.\oplus] \\
\cong & \qquad \{ \qquad \mathsf{Col} \text{ has upper adjoint } \mathsf{K}_{\mathcal{A},\mathcal{B}} \text{ , by assumption,} \\
 & \qquad\qquad\quad \text{so, by (23), } (\mathsf{Col}\bullet) \text{ has upper adjoint } (\mathsf{K}_{\mathcal{A},\mathcal{B}}\bullet) \quad \} \\
 & [f :: \quad \mathsf{K}_{\mathcal{A},\mathcal{B}} \bullet f \leftarrow \mathsf{Switch}.\oplus] \\
= & \qquad \{ \qquad \mathsf{K}_{\mathcal{A},\mathcal{B}} \bullet f \;=\; \mathsf{Switch}.(\mathsf{K}_{\mathcal{A}\leftarrow\mathcal{C},\mathcal{B}}.f) \\
 & \qquad\qquad\quad \text{(just as in lattice theory)} \quad \} \\
 & [f :: \quad \mathsf{Switch}.(\mathsf{K}_{\mathcal{A}\leftarrow\mathcal{C},\mathcal{B}}.f) \leftarrow \mathsf{Switch}.\oplus] \\
\cong & \qquad \{ \qquad \mathsf{Switch} \text{ is clearly an isomorphism of categories} \quad \} \\
 & [f :: \quad \mathsf{K}_{\mathcal{A}\leftarrow\mathcal{C},\mathcal{B}}.f \leftarrow \oplus] \quad .
\end{array}
$$

What emerges very clearly from this proof is that the parameterised colimit theorem is an instance of (23).

Mac Lane's [14, p.111] proof of this theorem involves the explicit construction of the witnesses to the isomorphism followed by a verification of their naturality properties. The above, equational, proof includes a construction of the witnesses as a (mechanical) by-product. Specifically, letting $\llbracket\ \rrbracket$ and $\lceil\!\lceil\ \rceil\!\rceil$ denote the lower and upper adjungate of the adjunction $(\mathsf{Col}_{\mathcal{A}}\bullet, \mathsf{K}_{\mathcal{A},\mathcal{C}}\bullet)$ they are:

$$\lceil\!\lceil\ \rceil\!\rceil_{\_\ ,\ \mathsf{Switch}.\oplus} \bullet \mathsf{K}.\mathsf{Switch} \;\in\; [f :: \quad f \;\leftarrow\; \mathsf{Col} \bullet \mathsf{Switch}.\oplus] \cong [f :: \quad (\mathsf{K}_{\mathcal{A}\leftarrow\mathcal{C},\mathcal{B}}.f){\leftarrow}\oplus]$$

and

$$\mathsf{K}.\mathsf{Switch} \bullet \llbracket\ \rrbracket_{\_\ ,\ \mathsf{Switch}.\oplus} \;\in\; [f :: \quad (\mathsf{K}_{\mathcal{A}\leftarrow\mathcal{C},\mathcal{B}}.f){\leftarrow}\oplus] \cong [f :: \quad f \;\leftarrow\; \mathsf{Col} \bullet \mathsf{Switch}.\oplus] \quad .$$

(The compositions arise from the transitivity of isomorphism, and the four subterms are the witnesses to the two isomorphism steps in the above proof.)

Example

The standard example of a parameterised colimit is the coproduct functor. Let $\mathcal{B}$ be the category consisting of exactly two objects, 0 and 1, and two arrows, namely the identity arrows on 0 and 1. Let $\mathcal{C}$ be $(\mathcal{A}{\times}\mathcal{A} \;,\; \sqsupseteq{\times}\sqsupseteq)$. Define $\oplus \in (\mathcal{A}{\leftarrow}\mathcal{C}){\leftarrow}\mathcal{B}$ by:

on objects $0{\oplus}(a_0,a_1) \;=\; a_0$ and $1{\oplus}(a_0,a_1) \;=\; a_1$; on arrows $\mathsf{id}_0{\oplus}(f_0,f_1) \;=\; f_0$ and $\mathsf{id}_1{\oplus}(f_0,f_1) \;=\; f_1$. Assuming that $\mathcal{A}$ is $\mathcal{B}$-cocomplete, this defines the *coproduct* functor $+$ to be the functor $\mathsf{Col} \bullet \mathsf{Switch}.\oplus$. That is, by definition,

$$a_0{+}a_1 \;\;=\;\; \mathsf{Col}.(\oplus(a_0,a_1)) \quad .$$

Applying the abstraction theorem, coproduct is the colimit of the functor $\oplus$ we have just defined.

Expanding the definitions of colimit, adjunction, etc., we find that this gives several "theorems for free", as they have been called by Wadler [20]. Briefly, the lower adjungate in the adjunction defining Col boils down to a mapping from arrows $f \in a \leftarrow b{+}c$ to the pair $(f \circ \mathbf{inl}_{b,c}\ ,\ f \circ \mathbf{inr}_{b,c})$ where **inl** and **inr** are both natural transformations; the upper adjungate boils down to a mapping from pairs of arrows f and g with the same codomain to an arrow $f \triangledown g$ with codomain the common codomain of f and g and domain the coproduct of the domains of f and g. Moreover, the property (27) yields

$$f \circ \mathbf{inl} \;=\; g \;\wedge\; f \circ \mathbf{inr} \;=\; h \;\;\equiv\;\; f \;=\; g \triangledown h \quad ,$$

the property (28) yields the conjunction of

$$(h{+}k) \circ \mathbf{inl} \;=\; \mathbf{inl} \circ h$$

and

$$(h{+}k) \circ \mathbf{inr} \;=\; \mathbf{inr} \circ k \quad .$$

(Several simplifications are needed to reduce what is obtained from (28) to this compact form. The conjunctions arise from equalities between pairs.) Finally, (29) boils down to

$$(f \circ g \circ k) \triangledown (f \circ h \circ l) \;=\; f \circ (g \triangledown h) \circ (k{+}l) \quad .$$

7 Algebras

The notion in category theory that corresponds to the notion of a prefix point in lattice theory is known as an *algebra*[6].

This section has the same structure as section 3. We begin by defining an algebra and an initial algebra. We then establish a fundamental theorem about conversion of one type of algebra to another via a lower adjoint, this theorem then being used to establish a constructive fusion theorem for initial algebras. Finally, we establish an abstraction theorem for initial algebras.

[6]Warning: The definition we are about to give is weaker than that given in [14] and [13], the terminological confusion that this leads to having been deplored by Lambek [12]. Nevertheless it seems to have become standard among computing scientists. See, for example, [15].

7.1 Basic Properties

Suppose $\mathcal{A}$ is a category and F is an endofunctor on $\mathcal{A}$. An F-*algebra* is an arrow in $\mathcal{A}$ of type $x \leftarrow F.x$ for some object x of $\mathcal{A}$. Thus, for all arrows f of the base category $\mathcal{A}$,

$$f \text{ is an } F\text{-algebra} \quad \equiv \quad \mathsf{dom}.f = F.(\mathsf{cod}.f) \quad .$$

The codomain of an F-algebra is often referred to as the *carrier* of the algebra.

The F-algebras can be organised into a category $\mathsf{Alg}_{\mathcal{A}}.F$ as follows. The objects are the F-algebras and the arrows φ to F-algebra f from F-algebra g are characterised by the equation:

$$\varphi \in f \xleftarrow{\mathsf{Alg}_{\mathcal{A}}.F} g \quad \equiv \quad \varphi \in \mathsf{cod}.f \xleftarrow{\mathcal{A}} \mathsf{cod}.g \quad \wedge \quad f \circ F.\varphi \ = \ \varphi \circ g \quad .$$

We will drop the subscript $\mathcal{A}$, writing simply $\mathsf{Alg}.F$, whenever it can be deduced from the context. The implicit parameter $\mathcal{A}$ will be called the *base category*.

Note, that the coherence condition $f \circ F.\varphi \ = \ \varphi \circ g$ arises from the fact that there are two different ways to construct an arrow to $\mathsf{cod}.f$ from $F.(\mathsf{cod}.g)$.

The claim that $\mathsf{Alg}.F$ is a category should be proven, of course. It is a straightforward exercise, however, and the fact is well-known. One “trivial” element of the proof that we make abundant use of is the fact that the identity arrows in $\mathsf{Alg}.F$ are the identity arrows in the base category. More precisely, if f is an F-algebra then

$$\mathsf{id}_f^{\mathsf{Alg}.F} \ = \ \mathsf{id}_{\mathsf{cod}.f}^{\mathcal{A}} \quad ,$$

i.e. $\mathsf{id}_{\mathsf{cod}.f}^{\mathcal{A}} \in f \xleftarrow{\mathsf{Alg}.F} f$. The superscripts ‘$\mathsf{Alg}.F$’ and ‘$\mathcal{A}$’ are usually omitted.

From the definition of arrows in $\mathsf{Alg}.F$ it is easy to see that cod can be extended to a functor to the base category from $\mathsf{Alg}.F$: the functor maps algebras to their carriers, and on arrows between algebras it is just the identity mapping.

An *initial* F-*algebra* is an initial object in the category $\mathsf{Alg}.F$. Existence of initial F-algebras is harder to predict than the existence of least prefix points. Generalisations to category theory of the Knaster-Tarski theorem have been considered by Lambek [11, 12], but typically it is *not* the case that if category $\mathcal{A}$ is cocomplete then all endofunctors on $\mathcal{A}$ have initial algebras. This won’t be a concern to us here; we shall simply assume that all the initial algebras we require do indeed exist.

Corresponding to the theorem that a least *prefix* point is a *fixed* point we have the theorem that:

$$\varphi \text{ is an initial } F\text{-algebra} \quad \Rightarrow \quad \varphi \cong F.\varphi \quad .$$

This theorem is due to Lambek [11].

7.2 The Fusion Theorem

Suppose we are given three functors $F \in \mathcal{A} \leftarrow \mathcal{B}$, $G \in \mathcal{A} \leftarrow \mathcal{A}$ and $H \in \mathcal{B} \leftarrow \mathcal{B}$. Suppose that there is a natural transformation $\eta \in F \bullet H \leftarrow G \bullet F$. Then we can define a functor K to $\mathsf{Alg}.G$ from $\mathsf{Alg}.H$ in the following way

(34) $\quad K.h \;=\; F.h \circ \eta_{\mathsf{cod}.h}$, for every object h in $\mathsf{Alg}.H$.

(35) $\quad K.\varphi \;=\; F.\varphi$, for every arrow φ in $\mathsf{Alg}.H$.

In particular, if F is taken to be the identity functor, we have:

$$\mathsf{Alg}.\eta \in \mathsf{Alg}.G \leftarrow \mathsf{Alg}.H \;\Leftarrow\; \eta \in H \leftarrow G \;\;,$$

where, by definition,

$$(\mathsf{Alg}.\eta).h \;=\; h \circ \eta_{\mathsf{cod}.h} \;\;,$$

for each object h in $\mathsf{Alg}.H$, and

$$(\mathsf{Alg}.\eta).\varphi \;=\; \varphi \;\;,$$

for each arrow φ in $\mathsf{Alg}.H$. In this way we have extended the definition of Alg to a (contravariant) functor from the category of endofunctors on base category $\mathcal{A}$ to the category of categories.

Another observation we can draw from (34) and (35) is the following. Suppose $\mathsf{mu}H$ is an initial H-algebra. Then $K.\mathsf{mu}H$ is a G-algebra. If, furthermore, $\mathsf{mu}G$ is an initial G-algebra then there is a (unique) arrow in the category $\mathsf{Alg}.G$ to $K.\mathsf{mu}H$ from $\mathsf{mu}G$.

This simple observation has two consequences of note. The first is obtained by again instantiating F to the identity functor. The observation becomes that there is an arrow to $\mathsf{mu}H$ from $\mathsf{mu}G$ if there is a natural transformation to H from G. This immediately suggests that mu can be extended to a functor Mu. Indeed, such a functor can be constructed although the details are a bit complicated. For this reason we postpone their consideration until later in this section.

The second consequence is in the case that F is an upper adjoint in an adjunction. Let $F^\flat$ denote the lower adjoint. Then, by the pseudo-inversality property of adjoint functors (24), given a natural transformation $\tau \in H \bullet F^\flat \leftarrow F^\flat \bullet G$ we can construct a natural transformation $\eta \in F \bullet H \leftarrow G \bullet F$, and thus a functor K to $\mathsf{Alg}.G$ from $\mathsf{Alg}.H$. Strengthening the requirement on τ to that it be an *isomorphism* between the functors $H \bullet F^\flat$ and $F^\flat \bullet G$ we can also construct a functor $K^\flat$ to $\mathsf{Alg}.H$ from $\mathsf{Alg}.G$. It is then possible to verify that the pair $(K^\flat, K)$ forms an adjunction between the categories $\mathsf{Alg}.G$ and $\mathsf{Alg}.H$, with the lower and upper adjungates defined in the same way as those of the adjunction $(F^\flat, F)$. (This theorem has been independently observed by Hermida and Jacobs [10].) An immediate corollary is that if $\mathsf{Alg}.G$ has an initial object $\mathsf{mu}G$ then $K^\flat.\mathsf{mu}G$ is an initial object of $\mathsf{Alg}.H$ (since in general lower adjoints map initial objects to initial objects). Filling in the definition of the functor $K^\flat$, we have established the *categorical fusion theorem*:

Theorem 36 (Categorical μ Fusion) Given are the following: two categories, $\mathcal{A}$ and $\mathcal{B}$, three functors $F^\flat \in \mathcal{A}\leftarrow\mathcal{B}$, $G \in \mathcal{B}\leftarrow\mathcal{B}$ and $H \in \mathcal{A}\leftarrow\mathcal{A}$, and an isomorphism $\mathsf{swap} \in F^\flat{\bullet}G \cong H{\bullet}F^\flat$. Furthermore, it is assumed that the functor $F^\flat$ is the lower adjoint in an adjunction between categories $\mathcal{A}$ and $\mathcal{B}$. Finally, it is also assumed that $\mathsf{Alg}.G$ has an initial object $\mathsf{mu}G$. Then

$$F^\flat.\mathsf{mu}G \;\circ\; \mathsf{swap}_{\mathsf{cod.mu}G}$$

is an initial H-algebra.

□

(We recommend skipping the remainder of this subsection on a first reading.) From the above theorem it immediately follows that if $\mathsf{mu}H$ denotes a canonical initial object in $\mathsf{Alg}.H$, then

$$F^\flat.\mathsf{mu}G \;\circ\; \mathsf{swap}_{\mathsf{cod.mu}G} \;\;\cong\;\; \mathsf{mu}H \quad .$$

To derive the witnesses for this isomorphism we argue as follows. From the assumed initiality of $\mathsf{mu}H$ there is a canonical arrow to $F^\flat.\mathsf{mu}G \;\circ\; \mathsf{swap}_{\mathsf{cod.mu}G}$ from $\mathsf{mu}H$, which we denote by $(\!|\mathsf{mu}H \;=:\; F^\flat.\mathsf{mu}G \;\circ\; \mathsf{swap}_{\mathsf{cod.mu}G}|\!)$. Our task is to construct the inverse arrow.

Since swap is assumed to be an isomorphism, we have an arrow $\mathsf{swap}^\cup \in H{\bullet}F^\flat\leftarrow F^\flat{\bullet}G$. Let $\lfloor\;\rfloor$ denote the lower adjungate and counit the counit of the assumed adjunction $(F^\flat, F)$ and let $\llbracket\;\rrbracket$ denote the lower adjungate of the adjunction $(F^\flat{\bullet}, F{\bullet})$ — whose existence follows by (23). Then the natural transformation η described above equals $\llbracket(H{\bullet}\mathsf{counit})\circ(\mathsf{swap}^\cup{\bullet}F)\rrbracket$. Furthermore, since $\llbracket\eta\rrbracket_x \;=\; \lfloor\eta_x\rfloor$, on objects h of $\mathsf{Alg}.H$ the functor K is defined by $K.h \;=\; F.h \circ \lfloor H.\mathsf{counit}_{\mathsf{cod}.h} \circ \mathsf{swap}^\cup_{F.(\mathsf{cod}.h)}\rfloor$. By construction of α we now derive the unique arrow from $F^\flat.\mathsf{mu}G \;\circ\; \mathsf{swap}_{\mathsf{cod.mu}G}$ to any other object of $\mathsf{Alg}.H$. Let h be an object of $\mathsf{Alg}.H$, then

$$\begin{array}{lll}
 & \alpha \;\in\; h \xleftarrow{\mathsf{Alg}.H} F^\flat.\mathsf{mu}G \;\circ\; \mathsf{swap}_{\mathsf{cod.mu}G} & \\
\Leftarrow & \quad\{ & (K^\flat, K) \text{ forms an adjunction, denote its lower and} \\
 & & \text{upper adjungates by } \lfloor\;\rfloor \text{ and } \lceil\;\rceil \text{ respectively} \\
 & & \text{(defined in the same way as those of } (F^\flat, F)). \\
 & & \text{assumption: } \alpha \;=\; \lceil\beta\rceil \quad \} \\
 & \beta \;\in\; F.h \circ \lfloor H.\mathsf{counit}_{\mathsf{cod}.h} \circ \mathsf{swap}^\cup_{F.(\mathsf{cod}.h)}\rfloor \xleftarrow{\mathsf{Alg}.G} \mathsf{mu}G & \\
\Leftarrow & \quad\{ & \text{assumed initiality of } \mathsf{mu}G \quad \} \\
 & \beta \;=\; (\!|\mathsf{mu}G \;=:\; F.h \circ \lfloor H.\mathsf{counit}_{\mathsf{cod}.h} \circ \mathsf{swap}^\cup_{F.(\mathsf{cod}.h)}\rfloor|\!) \quad . &
\end{array}$$

So $\lceil(\!|\mathsf{mu}G \;=:\; F.\mathsf{mu}H \;\circ\; \lfloor H.\mathsf{counit}_{\mu H} \circ \mathsf{swap}^\cup_{F.\mu H}\rfloor|\!)\rceil$ is a candidate for the desired arrow. Finally, there is a unique arrow to every object in a category from an initial object of that category. In particular, the arrow to an initial object from that initial object is unique and thus equals the identity arrow on that initial object. So, it is evident that the composition of the candidate witnesses in either order equals the identity arrow, i.e. they are each others inverses.

Let us now outline some of the details of the claim that mu can be extended to be a functor. The first complication we have to overcome is that it is unrealistic to assume that all endofunctors on base category $\mathcal{A}$ have initial algebras. Let us avoid this problem by defining the functor Mu only for functors for which an initial algebra exists. Stronger than that, we shall assume that mu is a (partial) function that maps an unspecified subclass of the endofunctors on category $\mathcal{A}$ to a canonical initial algebra. We shall define Mu only for this subclass of endofunctors. In order to make the definition precise let us assume that the unique arrow in the category $\mathsf{Alg}.F$ to F-algebra f from $\mathsf{mu}F$ (assuming the latter exists) is denoted by $(\!|\mathsf{mu}F \;=:\; f|\!)$. We also let μF denote the carrier of $\mathsf{mu}F$.

The second complication is that the codomain of Mu depends on its argument. This problem is overcome by defining its codomain to be a (dependent) sum category. Specifically, we denote the codomain of Mu by $\Sigma\mathsf{Alg}$, where

$$\Sigma\mathsf{Alg} \;=\; \Sigma(F : \quad F \in \mathsf{End}.\mathcal{A} : \quad \mathsf{Alg}.F) \quad .$$

The objects of $\Sigma\mathsf{Alg}$ are algebras f tagged with a functor F to indicate that f is an F-algebra, i.e. the objects of $\Sigma\mathsf{Alg}$ are pairs $(F\,,\,f)$ such that:

$$(F\,,\,f) \in \Sigma\mathsf{Alg} \equiv f \in \mathsf{Alg}.F \quad .$$

The arrows in $\Sigma\mathsf{Alg}$ are also pairs consisting of a construction between the functors and a construction between the algebras. More specifically,

$$(\eta\,,\,\varphi) \in (F\,,\,f) \xleftarrow{\Sigma\mathsf{Alg}} (G\,,\,g) \quad \equiv \quad \eta \in F \xleftarrow{\mathsf{End}.\mathcal{A}} G \;\;\wedge\;\; \varphi \;\in\; (\mathsf{Alg}.\eta).f \;\xleftarrow{\mathsf{Alg}.G}\; g \quad .$$

(Recall that $\mathsf{Alg}.\eta$ is a contravariant functor to $\mathsf{Alg}.G$ from $\mathsf{Alg}.F$.) We now have to define Mu as a functor to $\Sigma\mathsf{Alg}$ from $\mathsf{End}.\mathcal{A}$. On objects Mu is defined by

$$\mathsf{Mu}.F \;=\; (F\,,\,\mathsf{mu}F) \quad ,$$

for each $F \in \mathsf{End}.\mathcal{A}$. On arrows Mu is defined by

$$\mathsf{Mu}.\eta \;=\; (\eta\,,\,(\!|\mathsf{mu}G \;=:\; \mathsf{mu}F \circ \eta_{\mu F}|\!)) \quad ,$$

for each $\eta \in F \xleftarrow{\mathsf{End}.\mathcal{C}} G$.

7.3 The Abstraction Theorem

The final topic in this section is the abstraction theorem for initial algebras. Suppose $\oplus \in (\mathcal{A}\leftarrow\mathcal{A})\leftarrow\mathcal{B}$ and $F \in \mathcal{B}\leftarrow\mathcal{C}$. Let us adopt the convention once more that application of $\oplus$ to argument b is denoted by $b\oplus$, and application of this function to argument a is denoted by $b\oplus a$. (b and a may be objects or arrows.) Then, for all $c\in\mathcal{C}$, we can define an endofunctor $(F.c)\oplus$ on $\mathcal{A}$ by

$$((F.c)\oplus).x \;=\; F.c \oplus x \quad ,$$

for all objects x in $\mathcal{A}$, and

$$((F.c)\oplus).f \;=\; F.\mathsf{id}_c \oplus f \quad ,$$

for all arrows f in $\mathcal{A}$.

Throughout this section we make the assumption that, for all $b\in\mathcal{B}$, the functor $b\oplus$ has an initial object which we shall denote by $\mathsf{mu}(b\oplus)$. The codomain of $\mathsf{mu}(b\oplus)$ will be denoted by $\mu(b\oplus)$.

The mapping from b to $b\oplus$ is the object part of a functor from the category $\mathcal{B}$ to the category $\mathsf{End}.\mathcal{A}$, which functor we will denote by λ. (The specification of the arrow part is left to the reader.) The mapping from c to $\mu((F.c)\oplus)$ is thus also the object part of a functor from the category $\mathcal{C}$ to the category $\mathcal{A}$, since it is the composition of five functors: (in order of application) F, λ, the functor Mu discussed in the last section, the projection functor exr mapping a pair (x, y) to its right component y, and the functor cod. We denote this functor by the symbol ϖ.

An example of a functor defined in this way is the List functor. Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ all be the category **Set**. Define the binary functor $\oplus$ by

$$x\oplus y \;=\; \mathbb{1}+(x\times y)$$

(where $+$ and $\times$ denote coproduct and product, respectively) and let the functor F be the identity functor. Then $\varpi.c \;=\; \mu(y \mapsto \mathbb{1}+(c\times y))$, for set c, is the set of all lists with elements drawn from c; for function f to set c from set d, $\varpi.f \;=\; \mu(y \mapsto \mathbb{1}+(f\times y))$ is the function that “maps” f over all the elements of a list of d’s in order to construct a list of $c's$.

The abstraction theorem for initial algebras states that the functor ϖ is itself the carrier (i.e. codomain) of an initial algebra. In order to state the theorem precisely let us introduce the following notation. With $(F\dot{\oplus})$ we denote the functor $G\mapsto(x \;\mapsto\; F.x \oplus G.x)$. Specifically,

$$((F\dot{\oplus}).G).x \;=\; F.x \oplus G.x \quad ,$$

for all functors $G \in \mathcal{A}\leftarrow\mathcal{C}$ and all (objects or arrows) x in $\mathcal{C}$, and

$$((F\dot{\oplus}).\eta)_x \;=\; \mathsf{id}_{F.x}\oplus\eta_x \quad ,$$

for all natural transformations η and objects x in $\mathcal{C}$. Note that $(F\dot{\oplus})$ is an endofunctor on the category $\mathsf{Fun}(\mathcal{A},\mathcal{C})$ (the category of functors to $\mathcal{A}$ from or arrows $\mathcal{C}$). The abstraction theorem states that the natural transformation π defined by $\pi_x \;=\; \mathsf{mu}((F.x)\oplus)$ is an initial $(F\dot{\oplus})$-algebra. Moreover, its carrier is the functor ϖ.

The proof of this theorem, if carried out conscientiously, involves a substantial amount of detail. We have to prove that π is an $(F\dot{\oplus})$-algebra and for all $(F\dot{\oplus})$-algebras η there is a unique $(F\dot{\oplus})$-algebra arrow to η from π. The latter is given by the natural transformation α where $\alpha_z \;=\; (\!|\eta_z \;=\!:\, \mathsf{mu}((F.z)\oplus)|\!)$. The construction of this arrow is a mechanical exercise given the proof of the abstraction theorem in lattice theory; the bulk of the proof consists of showing that it is indeed an arrow in $\mathsf{Alg}.(F\dot{\oplus})$ and that it is unique.

8 The Categorical "Beautiful Theorem"

We are now in a position to prove the categorical "beautiful theorem". The theorem is about the preservation of properties of colimits when constructing initial algebras.

Suppose $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are categories and suppose that $\mathcal{A}$ and $\mathcal{B}$ are both $\mathcal{C}$-cocomplete. Let $\mathsf{Col}_{\mathcal{A}}$ and $\mathsf{Col}_{\mathcal{B}}$ denote canonical colimit operators of type $\mathcal{A}\leftarrow(\mathcal{A}\leftarrow\mathcal{C})$ and $\mathcal{B}\leftarrow(\mathcal{B}\leftarrow\mathcal{C})$, respectively. The only fact that we will use about these two operators is that $\mathsf{Col}_{\mathcal{A}}$ is a lower adjoint in an adjunction between the categories $\mathcal{A}$ and $\mathcal{A}\leftarrow\mathcal{C}$.

Suppose $\oplus \in (\mathcal{A}\leftarrow\mathcal{A})\leftarrow\mathcal{B}$. We adopt the same notational conventions and assumptions as in section 7.3. Thus we assume that, for all $b\in\mathcal{B}$, the functor $b\oplus$ has an initial object denoted by $\mathsf{mu}(b\oplus)$. The codomain of $\mathsf{mu}(b\oplus)$ is again denoted by $\mu(b\oplus)$. Let $\dagger$ denote the functor whose action on objects is to map b to $\mu(b\oplus)$. More precisely, define

$$\dagger \;=\; \mathsf{cod} \bullet \mathsf{exr} \bullet \mathsf{Mu} \bullet \lambda \;\; .$$

(See section 7.3 for the definitions of the four functors on the right side of this equation.) Our goal is to prove that $\dagger$ "commutes" with Col if $\oplus$ "commutes" with Col. To be precise, we prove that

$$\forall(F :: \;\mathsf{Col}_{\mathcal{A}}.(\dagger\bullet F) \cong \dagger.(\mathsf{Col}_{\mathcal{B}}.F)) \Leftarrow \forall(F,G :: \;\mathsf{Col}_{\mathcal{A}}.(F \mathbin{\dot{\oplus}} G) \cong (\mathsf{Col}_{\mathcal{B}}.F)\oplus(\mathsf{Col}_{\mathcal{A}}.G)) \;\; .$$

Here and elsewhere the dummy F ranges over functors of type $\mathcal{B}\leftarrow\mathcal{C}$ and the dummy G over functors of type $\mathcal{A}\leftarrow\mathcal{C}$.

Let us consider the premise. Eliminating the quantification over G, we are given that, for all F,

$$\mathsf{Col}_{\mathcal{A}} \bullet (F\dot{\oplus}) \;\cong\; (\mathsf{Col}_{\mathcal{B}}.F)\oplus \bullet \mathsf{Col}_{\mathcal{A}} \;\; .$$

Suppose the witnessing isomorphism is τ. Then, by the fusion theorem (and the fact that $\mathsf{Col}_{\mathcal{A}}$ is a lower adjoint),

$$\mathsf{Col}_{\mathcal{A}}.\mathsf{mu}(F\dot{\oplus}) \circ \tau_{\mu(F\dot{\oplus})}$$

is an initial $((\mathsf{Col}_{\mathcal{B}}.F)\oplus)$-algebra whenever $\mathsf{mu}(F\dot{\oplus})$ is an initial $(F\dot{\oplus})$-algebra. But, by the abstraction theorem, π defined by $\pi_x \;=\; \mathsf{mu}((F.x)\oplus)$ is an initial $(F\dot{\oplus})$-algebra. Thus,

$$\begin{array}{ll} & \dagger.(\mathsf{Col}_{\mathcal{B}}.F) \cong \mathsf{Col}_{\mathcal{A}}.(\dagger\bullet F) \\ \equiv & \quad \{ \quad \dagger.(\mathsf{Col}_{\mathcal{B}}.F) \text{ is by definition the codomain of an initial} \\ & \quad\quad ((\mathsf{Col}_{\mathcal{B}}.F)\oplus)\text{-algebra; by definition, } \dagger\bullet F \;=\; \mathsf{cod}.\pi \quad \} \\ & \mathsf{cod}.(\mathsf{Col}_{\mathcal{A}}.\pi \circ \tau_{\mu(F\dot{\oplus})}) \cong \mathsf{Col}_{\mathcal{A}}.(\mathsf{cod}.\pi) \\ \equiv & \quad \{ \quad \text{codomain calculus} \quad \} \\ & \mathit{true} \quad . \end{array}$$

We have thus proved that

$$\forall(F :: \ \dagger.(\mathsf{Col}_{\mathcal{B}}.F) \cong \mathsf{Col}_{\mathcal{A}}.(\dagger\bullet F) \ \Leftarrow \ \forall(G :: \ \mathsf{Col}_{\mathcal{A}}.(F \mathbin{\dot{\oplus}} G) \cong (\mathsf{Col}_{\mathcal{B}}.F)\oplus(\mathsf{Col}_{\mathcal{A}}.G))) \quad .$$

The "beautiful theorem" follows by elementary calculus.

For completeness we should of course provide details of the witnesses to the constructed isomorphism. The details are complicated but entirely mechanical. Let $\tau^{\cup}$ denote the inverse of τ. Let $\lfloor\ \rfloor$ and $\lceil\ \rceil$ denote the lower and upper adjungate and counit the counit of the adjunction $(\mathsf{Col}_{\mathcal{A}}, \mathsf{K}_{\mathcal{A},\mathcal{C}})$. Then the unique arrow from $\mathsf{Col}_{\mathcal{A}}.\pi \circ \tau_{\mu(F\dot{\oplus})}$ to a $(\mathsf{Col}_{\mathcal{B}}.F)\oplus$-algebra f equals $\lceil(\!|\pi \ =: \mathsf{K}_{\mathcal{A},\mathcal{C}}.f \circ \lfloor(\mathsf{Col}_{\mathcal{B}}.F)\oplus\mathsf{counit}_{\mathsf{cod}.f} \circ \tau^{\cup}_{\mathsf{K}_{\mathcal{A},\mathcal{C}}.(\mathsf{cod}.f)}\rfloor|\!)\rceil$. (See the discussion immediately following theorem 36). Furthermore,

$$\dagger.(\mathsf{Col}_{\mathcal{B}}.F) \;=\; (\mathsf{cod} \bullet \mathsf{exr} \bullet \mathsf{Mu} \bullet \lambda).(\mathsf{Col}_{\mathcal{B}}.F) \;=\; \mathsf{cod}.(\mathsf{mu}((\mathsf{Col}_{\mathcal{B}}.F)\oplus))$$

Thus,

$$\lceil(\!|\pi \ =: \mathsf{K}_{\mathcal{A},\mathcal{C}}.\mathsf{mu}((\mathsf{Col}_{\mathcal{B}}.F)\oplus) \ \circ \ \lfloor(\mathsf{Col}_{\mathcal{B}}.F)\oplus\mathsf{counit}_{\mathsf{cod}.f} \circ \tau^{\cup}_{\mathsf{K}_{\mathcal{A},\mathcal{C}}.(\dagger.(\mathsf{Col}_{\mathcal{B}}.F))}\rfloor|\!)\rceil$$

is the unique arrow to $\dagger.(\mathsf{Col}_{\mathcal{B}}.F)$ from $\mathsf{cod}.(\mathsf{Col}_{\mathcal{A}}.\pi \circ \tau_{\mu(F\dot{\oplus})})$. (Recall that an arrow between two algebras in an algebra category is also an arrow between their carriers in the base category). Finally,

$$\mathsf{cod}.(\mathsf{Col}_{\mathcal{A}}.\pi \circ \tau_{\mu(F\dot{\oplus})}) \;=\; \mathsf{Col}_{\mathcal{A}}.(\mathsf{cod}.\pi) \;=\; \mathsf{Col}_{\mathcal{A}}.(\dagger\bullet F) \quad .$$

So, defining

$$\eta \;=\; \lceil(\!|\pi \ =: \mathsf{K}_{\mathcal{A},\mathcal{C}}.\mathsf{mu}((\mathsf{Col}_{\mathcal{B}}.F)\oplus) \ \circ \ \lfloor(\mathsf{Col}_{\mathcal{B}}.F)\oplus\mathsf{counit}_{\mathsf{cod}.f} \circ \tau^{\cup}_{\mathsf{K}_{\mathcal{A},\mathcal{C}}.(\dagger.(\mathsf{Col}_{\mathcal{B}}.F))}\rfloor|\!)\rceil$$

we have:

$$\eta \ \in \ \dagger.(\mathsf{Col}_{\mathcal{B}}.F) \cong \mathsf{Col}_{\mathcal{A}}.(\dagger\bullet F) \quad \Leftarrow \quad \tau \ \in \ \mathsf{Col}_{\mathcal{A}} \bullet (F\dot{\oplus}) \ \cong \ (\mathsf{Col}_{\mathcal{B}}.F)\oplus \bullet \mathsf{Col}_{\mathcal{A}} \quad .$$

9 Discussion

The goal of this paper has been to advance the view that category theory may be profitably regarded as constructive lattice theory in which the constructions are required to satisfy certain coherence properties. In order to support our argument we have presented two theorems on initial algebras, the abstraction theorem and the fusion theorem, both of which we believe are new, and both of which are inspired by theorems in lattice theory on least prefix points whose usefulness is without doubt.

We have applied the (categorical) abstraction and fusion theorems to generalise Dijkstra and Scholten's "beautiful theorem" to category theory. This generalisation has involved a two-way interaction between lattice theory and category theory: the emphasis in category theory on the notion of an adjunction inspired us to seek a novel proof of Dijkstra and Scholten's theorem in which the existence of a Galois connection defining suprema played a central rôle. Having established the theorem at the level of lattice theory we were then able to add the extra detail needed to establish its categorical counterpart. The general process is relatively mechanical even though the details are —as in this case— often intricate. The reliability of the process is significantly enhanced by splitting it into two separate phases: in the first phase a completely equational argument is used to establish the required isomorphism, and in the second phase the "witnesses" to each step in the equational calculational are filled in.

An important special instance of the categorical "beautiful theorem" is that ω-cocontinuity is preserved by the process of constructing initial algebras [17, p.289]. A striking feature of our proof is that no use is made whatsoever of the structure of the adjunction defining colimits; we have used the fact that $\mathcal{B}$-cocompleteness of a category is equivalent to the existence of an adjunction but the structure of the colimit functor or its upper adjoint has not entered the picture. This is in marked contrast to proofs given by others of the preservation of ω-cocontinuity [16, 15] where details of the construction of a category of cocones plays a prominent rôle. This emphasises the more fundamental nature of the abstraction and fusion theorems.

The abstraction and fusion theorems are not the only theorems that are needed to construct a rich (constructive) calculus of initial algebras. Elsewhere [5] we have generalised the following three fixed point rules to category theory. The *diagonal rule*:

$$\mu(x \mapsto x{\oplus}x) \;=\; \mu(x \mapsto \mu(y \mapsto x{\oplus}y)) \quad ,$$

the *rolling rule*:

$$f.\mu(g{\bullet}f) \;=\; \mu(f{\bullet}g) \quad ,$$

and the *exchange rule*: if g and h are both the lower adjoints in Galois connections then

$$\mu(f{\bullet}g) \;=\; \mu(f{\bullet}h) \;\Leftarrow\; g{\bullet}f{\bullet}h \;=\; h{\bullet}f{\bullet}g \quad .$$

(The diagonal rule has been proved for ω-functors in [15]. Lambert Meertens was the first, to our knowledge, to formulate and prove the rolling rule in category theory.) We have also shown that, taken together, the five rules can be used to give compact deriviations of solutions to several list programming problems. These derivations are based on the view that the theory of lists can be profitably regarded as constructive regular algebra. (See also [3].)

In spite of the fact that it is more than a quarter of a century since Lambek's call to draw inspiration from lattice theory for the construction of theorems in category theory we still feel that a number of unturned stones remain. As an example, we would draw attention to the relationship between closure operators and monads. Based on related work [3], we have shown that the *closure decomposition* theorem, which is central to a number of program derivations [1, 4] can be generalised to a *monad decomposition* theorem. Other theorems on closure operators reported in, for example, [18] have, so far as we know, as yet no categorical counterparts. We contend that viewing category theory as coherently constructive lattice theory can have a profound and lasting influence on programming methodology.

Acknowledgement

Thanks go to Jaap van der Woude for much background support and for his careful reading of an earlier draft of this paper. Grant Malcolm helped to fill in some references to the literature.

The paper was written on-screen using the MathSpad interface without which its completion would have been delayed by many months. Thanks go to Richard Verhoeven for his untiring support.

References

[1] R.C. Backhouse and B.A. Carré. Regular algebra applied to path-finding problems. *Journal of the Institute of Mathematics and its Applications*, 15:161–186, 1975.

[2] R.C. Backhouse and H. Doornbos. Mathematical induction made calculational. Technical Report Computing Science Note No. 94/16, Eindhoven University of Technology, 1994. Available via anonymous ftp from `ftp.win.tue.nl` in directory `pub/math.prog.construction`.

[3] Roland Backhouse. Constructive lattice theory. Department of Computer Science, Eindhoven University of Technology, Available via anonymous ftp from `ftp.win.tue.nl` in directory `pub/math.prog.construction`, October 1993.

[4] Roland C. Backhouse, J.P.H.W. van den Eijnde, and A.J.M. van Gasteren. Calculating path algorithms. *Science of Computer Programming*, 22(1–2):3–19, 1994.

[5] Marcel Bijsterveld. Categorical fixed point rules. Graduating Dissertation. Eindhoven University of Technology, Department of Mathematics and Comput-

ing Science, Available via anonymous ftp from `ftp.win.tue.nl` in directory `pub/math.prog.construction`, June 1994.

[6] Garrett Birkhoff. *Lattice Theory*, volume 25 of *American Mathematical Society Colloquium Publications.* American Mathematical Society, Providence, Rhode Island, 3rd edition, 1967.

[7] B. A. Davey and H. A. Priestly. *Introduction to Lattices and Order.* Cambridge Mathematical Textbooks. Cambridge University Press, first edition, 1990.

[8] E.W. Dijkstra and C.S. Scholten. *Predicate Calculus and Program Semantics.* Springer-Verlag, Berlin, 1990.

[9] P.J. Freyd and A. Scedrov. *Categories, Allegories.* North-Holland, 1990.

[10] Claudio A. Hermida and Bart Jacobs. An algebraic view of structural induction. Extended Abstract, May 1994.

[11] J. Lambek. A fixpoint theorem for complete categories. *Mathematische Zeitschrift*, 103:151–161, 1968.

[12] J. Lambek. Least fixpoints of endofunctors of cartesian closed categories. *Mathematical Structures in Computer Science*, 3:229–257, 1993.

[13] J. Lambek and P.J. Scott. *Introduction to Higher Order Categorical Logic*, volume 7 of *Studies in Advanced Mathematics.* Cambridge University Press, 1986.

[14] S. Mac Lane. *Categories for the Working Mathematician*, volume 5 of *Graduate Texts in Mathematics.* Springer-Verlag, 1971.

[15] D.J. Lehman and M.B. Smyth. Algebraic specification of data types: A synthetic approach. *Math. Syst. Theory*, 14(2):97–140, 1981.

[16] G. Malcolm. *Algebraic data types and program transformation.* PhD thesis, Groningen University, 1990.

[17] E.G. Manes and M.A. Arbib. *Algebraic Approaches to Program Semantics.* Texts and Monographs in Computer Science. Springer-Verlag, Berlin, 1986.

[18] Dana Scott. Data types as lattices. *SIAM J. of Computing*, 5(3):522–587, September 1976.

[19] Ed Voermans. A relational perspective on datatypes. Ph.D. Thesis (title provisional), Forthcoming, 1995.

[20] P. Wadler. Theorems for free! In *4'th Symposium on Functional Programming Languages and Computer Architecture, ACM, London*, September 1989.