Dyads, a generalisation of monads

Maarten Fokkinga

University of Twente, Dept INF, PO Box 217, NL 7500 AE Enschede
e-mail: fokkinga@cs.utwente.nl

Version of December 6, 1993

The concept of dyad is defined as the least common generalisation of monads and co-monads. So, taking some of the ingredients to be the identity, the concept specialised to the concept of monad, and taking other ingredients to be the identity it specialised to co-monads. Except for one axiom, all have a nice ("natural") form.

Introduction: monads

Let us first describe one way in which monads can be motivated. We shall motivate dyads by a similar argument later. (Other descriptions have been given by Barr and Wells [1] and Wadler [2].) We use a fairly standard notation: \( \mathcal{A}, \mathcal{B}, \ldots \) denote categories, \( a, b, \ldots \) objects, \( f, g, \ldots \) arrows, \( F, G, \ldots \) functors, greek letters denote natural transformations, and \( x, y, \ldots \) various things. Composition is denoted in diagrammatic order: \( f : g \) is normally written \( g \circ f \).

Example. Let \( L \) be the list functor: \( La \) is the set of lists over \( a \), and for \( f : a \to b \) we have \( Lf : La \to Lb \) as the well-known \( f \)-map. Given list producing functions \( f : a \to Lb \) and \( g : b \to Lc \), we often see the list producing "composition":

\[
\begin{align*}
\text{Example.} & \quad f : Lg \circ+ : a \to Lc .
\end{align*}
\]

Here \( \circ+ : LL \to L \) is the flattening, or concatenation, of lists of list into lists. Functions of type \( a \to Lb \), for varying \( a, b \), turn up frequently in actual programming. For example, unconditional list comprehensions can be described in that way:

\[
\begin{align*}
x \mapsto [z \mid y \leftarrow fx; \ z \leftarrow gy] & = f \circ Lg \circ+ .
\end{align*}
\]

Notice, moreover, the existence of a particularly nice list producing function of type \( a \to Lb \) with \( a = b \), namely the singleton former \( \eta = \lambda x :: [x] \). It has the property that \( f \circ L\eta \circ+ = f = \eta \circ Lf \circ+ \).
Generalisation. The situation above can be described more elegantly in categorical terms as follows. Let $F$ be an endofunctor. Under what conditions can we consider arrows of type $f: a \to Fb$, for varying $a, b$, to be arrows of type $a \to b$ in another category? This question means, amongst others, that there must be a way to “compose” arrows $f: a \to Fb$ and $g: b \toFc$ into an arrow of type $a \to Fc$, and that this composition is associative, and that there exists a function $\eta_a: a \to Fa$ for each $a$ that is the identity for the new “composition”.

The new category is known as the Kleisli category, and we shall indicate it by $\mathcal{K}(F)$, or simply $\mathcal{K}$ if the functor $F$ is understood. The construction of $\mathcal{K}$ is straightforward. Define:

- $a$ object in $\mathcal{K} \equiv a$ object in the given category
- $f$ arrow in $\mathcal{K} \equiv f$ arrow in the given category of type $a \to Fb$ for some $a, b$
- $f: a \to \mathcal{K} b \equiv f: a \to Fa$
- $f \circ K g = f: Fg \mu$ whenever $f: a \to Fb$, $g: b \to Fc$
- $id_{\mathcal{K}, a} = \eta_a$

where natural transformations $\mu$ and $\eta$ are assumed to exist:

$$
\mu : FF \to F \\
\eta : I \to F 
$$

The associativity of composition $\circ K$, and the neutrality of $id_{\mathcal{K}}$ for this composition, are equivalent to the following three properties of $\mu$ and $\eta$:

$$
F\mu \circ \mu = \mu F \circ \mu \\
F\eta \circ \mu = \eta F \circ \mu = id F 
$$

The first two of these say that $F\mu$ equals $\mu F$, and $F\eta$ equals $\eta F$, when followed by $\mu$. Indeed, a proof of associativity reads:

$$
(f \circ K g) \circ K h = f \circ K (g \circ K h) \\
\equiv \text{definition } \circ K \\
(f : Fg \mu) \circ Fh : \mu = f : F(g \circ Fh \mu) : \mu \\
\equiv \text{category, functor} \\
f : Fg \mu \circ Fh : \mu = f : Fg \circ FFh \circ F\mu : \mu \\
\equiv \text{in lhs: naturality } \mu \\
f : Fg \circ FFh \circ F\mu : \mu = f : Fg \circ FFh \circ F\mu : \mu \\
\equiv \text{for } \Leftarrow : \text{Leibniz;} \text{ for } \Rightarrow : \text{ instantiate with } f, g, h \text{ all equal to } id \\
\mu F \circ \mu = F\mu \circ \mu 
$$
A proof of neutrality of $id_K$ for $\iota_K$ reads:

\[
\begin{align*}
\iota K \ id K &= f = id_K \colon f \\
\equiv & \text{ definition } \iota K \text{ and } id_K \\
f \colon F \eta \colon \mu = f = \eta \colon F \colon \mu \\
\equiv & \text{ naturality } \eta \\
f \colon F \eta \colon \mu = f = \eta F \colon \mu \\
\equiv & \text{ for } \Leftarrow: \text{ Leibniz; for } \Rightarrow: \text{ instantiate with } f = id \\
F \eta \colon \mu = id F = \eta F \colon \mu .
\end{align*}
\]

The three remaining category axioms for $K$ read:

\[
\begin{align*}
f \colon a \to_K b, \ g \colon b \to_K c \Rightarrow f \circ_K g \colon a \to_K c \\
id_K \colon a \to_K a \\
f \colon a \to_K b, \ f \colon a' \to_K b' \Rightarrow a = a' \land b = b' .
\end{align*}
\]

The first two of these are evidently true; the latter one isn’t true in general. So the construction yields a pre-category, and by a standard construction one obtains a category, the Kleisli category $K$.

In summary, the ingredients that enable the construction of a Kleisli category are:

\[
(F, \ \mu \colon FF \to F, \ \eta \colon Id \to F)
\]

satisfying

\[
\begin{align*}
F \mu \colon \mu &= \mu F \colon \mu \\
F \eta \colon \mu &= \eta F \colon \mu = id F .
\end{align*}
\]

Such a triple is known as a monad.

**A generalisation wanted**

Loosely formulated the above construction of the Kleisli category enables us to “compose” arrows of type $a \to Fb$ (considering them to be of type $a \to t b$) for fixed $F$ and varying $a, b$. Doaitse Swierstra posed the problem of “composing” arrows of type $a \times b \to Fe$, for fixed $F$ and varying $a, b, e$. The Kleisli construction doesn’t help here.

Abstracting a little bit, the new problem is to construct a Kleisli-like category for arrows of type $Fa \to Gb$ (considering them to be of type $a \to t b$), for fixed $F, G$ and varying $a, b$. Taking $F = Id$ we get the original problem that is solved by Kleisli’s construction and the existence of a monad for $G$. Taking $G = Id$ we get the dual problem; solved by the dual of Kleisli’s construction and the existence of a co-monad for $F$. Thus our current problem formulation is the least common generalisation of the “Kleisli problem” and its dual; and our solution will be the least common generalisation of the Kleisli construction and its dual. We will term the ingredients of the solution a dyad: the least common generalisation of a monad and a co-monad.
Dyads

Let \( F, G \) be functors with a common source and a common target. We will construct a category \( \mathcal{D}(F, G) \), or simply \( \mathcal{D} \), whose arrows (of type \( a \to_b \)) are the given arrows of type \( Fa \to Gb \). To this end we define:

- **Object in** \( \mathcal{D} \) \( \equiv \) **Object in source category of** \( F; G \)
- **Arrow in** \( \mathcal{D} \) \( \equiv \) **Arrow in target category of** \( F; G \)
- of type \( Fa \to Gb \) for some \( a, b \)
- \( f: a \to_b b \) \( \equiv f: Fa \to Gb \)
- \( f \circ g = \mu: Ff: Gg: \nu \) (explained below)
- \( id_\mathcal{D} = \eta \)

where natural transformations \( \gamma, \eta, \mu, \nu \) are assumed to exist:

\[
\begin{align*}
\gamma &: FG \to GF & F,G\text{-commuting transformation} \\
\eta &: F \to G & F\text{-to } G\text{-unit transformation} \\
\mu &: F \to FF & F\text{-generating transformation} \\
\nu &: GG \to G & G\text{-reducing transformation}
\end{align*}
\]

Notice that \( G \) and \( \nu \) play the role of \( F \) and \( \mu \) in the usual nomenclature for monads and the Kleisli construction. The above definition of \( f \circ g \) is the simplest general way to combine \( f \) and \( g \) in that order into an arrow of type \( Fa \to Gc \). Indeed, in order that a transformed \( f \) and transformed \( g \) both have the same ingredients in their target and source type, respectively, \( F \) has to be applied to \( f \) and \( G \) to \( g \); this gives \( Ff: FFa \to GFe \) and \( Gg: GFb \to GGg \):

\[
\begin{array}{c}
Fa \xrightarrow{\gamma} FFa \xrightarrow{Ff} FGb \xrightarrow{\gamma} GFb \xrightarrow{Gg} GGe \xrightarrow{\nu} Gb
\end{array}
\]

Now an arrow of type \( Fa \to Gc \) may be obtained by an \( F\)-generating transformation first, followed by \( Ff \) and \( Gg \) with an \( F,G\text{-commuting transformation} \) in between, and a \( G\)-reducing transformation at the end:

\[
\begin{array}{c}
Fa \xrightarrow{\mu} FFa \xrightarrow{Ff} FGb \xrightarrow{\gamma} GFb \xrightarrow{Gg} GGe \xrightarrow{\nu} Gb
\end{array}
\]

Clearly, the typing axioms, except for uniqueness of typing, are satisfied:

- \( f: a \to_b b, \; g: b \to_c c \Rightarrow f \circ g: a \to_c c \)
- \( id_\mathcal{D}: a \to_a a \)

So in order to prove that \( \mathcal{D} \) is a pre-category, it remains to show that \( \circ \) is associative and \( id_\mathcal{D} \) is neutral for \( \circ \). We shall now give those proofs, assuming suitable properties on
\( \gamma, \eta, \mu, \nu \) along the way. First, for associativity, let \( f : a \to b, g : b \to c, \) and \( h : c \to d \) be arbitrary; then:

\[
\begin{align*}
\quad f \circ (g \circ h) &= (f \circ g) \circ h \\
\quad \equiv & \quad \text{definition } D \\
\quad F_a \circ F_{F_a} \circ F_{F_{F_a}} \circ F_{G_{F_{F_a}}} \circ F_{G_{G_{F_{F_a}}}} \circ \gamma \circ G_{F_{G_{F_{F_a}}}} \circ G_F \circ \mu \circ F \circ \nu \circ F \circ g &= \circ (f \circ g) \circ \gamma \circ G_{F \circ \mu} \circ F \circ \nu \circ F \circ g \\
\quad \| &= \| \quad \text{commutes} \\
\quad F_a \circ F_{F_a} \circ F_{F_{F_a}} \circ F_{G_{F_{F_a}}} \circ F_{G_{G_{F_{F_a}}}} \circ F_{G_{G_{G_{F_{F_a}}}}} \circ F_{G_{G_{G_{G_{F_{F_a}}}}}} \circ G_{G_{G_{G_{G_{F_{F_a}}}}}} \circ G_F \circ \gamma \circ G_{F \circ \mu} \circ F \circ \nu \circ F \circ g &= \circ (f \circ g) \circ \gamma \circ G_{F \circ \mu} \circ F \circ \nu \circ F \circ g \\
\quad (\ast) \quad \iff & \quad \{ \text{in order to split the terms into several parts at isomorphic objects:} \} \\
\quad \text{assume } \gamma \text{ has inverse } \tilde{\gamma} \text{, so that } \gamma F : FGF \to GFF \text{ has inverse } \tilde{\gamma} F \\
\quad F_a \circ F_{F_a} \circ F_{F_{F_a}} \circ F_{G_{F_{F_a}}} \circ F_{G_{G_{F_{F_a}}}} \circ F_{G_{G_{G_{F_{F_a}}}}} \circ F_{G_{G_{G_{G_{F_{F_a}}}}}} \circ G_{G_{G_{G_{G_{F_{F_a}}}}}} \circ G_F \circ \gamma \circ G_{F \circ \mu} \circ F \circ \nu \circ F \circ g &= \circ (f \circ g) \circ \gamma \circ G_{F \circ \mu} \circ F \circ \nu \circ F \circ g \\
\quad \| &= \| \quad \text{commutes} \\
\quad F_a \circ F_{F_a} \circ F_{F_{F_a}} \circ F_{G_{F_{F_a}}} \circ F_{G_{G_{F_{F_a}}}} \circ F_{G_{G_{G_{F_{F_a}}}}} \circ F_{G_{G_{G_{G_{F_{F_a}}}}}} \circ G_{G_{G_{G_{G_{F_{F_a}}}}}} \circ G_F \circ \gamma \circ G_{F \circ \mu} \circ F \circ \nu \circ F \circ g &= \circ (f \circ g) \circ \gamma \circ G_{F \circ \mu} \circ F \circ \nu \circ F \circ g \\
\quad \equiv & \quad \text{diagram notation} \\
\quad \mu : F \circ \gamma : G \mu &= \mu : F \circ \mu : F F \circ \gamma : \gamma F \quad \land \\
\quad G F \circ \gamma : \gamma G &= \gamma F \circ F G \circ \gamma : \gamma F \quad \land \\
\quad \gamma G : G \gamma : G G h : G \nu : \nu &= F \nu : \gamma : G h : \nu \\
\quad \equiv & \quad \{ \text{for readability:} \} \\
\quad \text{define } \gamma_{1,1} &= F \circ \gamma : F F G \to G F F \quad \land \\
\quad \text{define } \gamma_{1,2} &= \gamma G : G G F \to G G F \\
\quad \mu : F F \circ \gamma : G \mu &= \mu : F \circ \mu : F F \circ \gamma : \gamma_{1,1} \quad \land \\
\quad G F \circ \gamma : \gamma G &= \gamma F \circ F G \circ \gamma : \gamma_{1,1} \quad \land \\
\quad \gamma_{1,2} : G G h : G \nu : \nu &= F \nu : \gamma : G h : \nu \\
\quad \equiv & \quad 2\text{nd conjunct: naturality } \tilde{\gamma} : G F \to G F ; \\
\quad 1\text{st, 3rd conjunct: } \{ \text{aiming at the next two steps,} \} \\
\quad \text{assume } \gamma : G \mu &= \mu G : \gamma_{1,1} \quad \land \quad F \nu : \gamma = \gamma_{1,2} : \nu F \\
\quad \mu : F F \circ \gamma : \gamma_{1,1} &= \mu : F \circ \mu : F F \circ \gamma : \gamma_{1,1} \quad \land \\
\quad \gamma_{1,2} : G G h : G \nu : \nu &= \gamma_{1,2} : \nu F : G h : \nu \\
\quad \equiv & \quad \text{naturality } \mu : F \to F F \quad \land \quad G G : G G F \to G \\
\quad \mu : F F \circ \gamma : \gamma_{1,1} &= \mu : F \circ \mu : F F \circ \gamma : \gamma_{1,1} \quad \land \\
\quad \gamma_{1,2} : G G h : G \nu : \nu &= \gamma_{1,2} : G G h : \nu G : \nu \\
\quad \iff & \quad \text{Leibniz} \\
\quad \mu : F F &= \mu : F \mu \quad \land \\
\quad G \nu : \nu &= \nu G : \nu \\
\quad \equiv & \quad \text{assume } \mu : F F &= \mu : F \mu \quad \land \quad G \nu : \nu &= \nu G : \nu \\
\quad \text{true .}
The $\gamma_{2,1}$ and $\gamma_{1,2}$ defined above are instances of a more general $\gamma_{m,n}: F^m G^n \to G^n F^m$ that one may define easily by induction on $m$ and $n$ in several distinct but semantically equal ways. In step (*) there is no other nontrivial way to split the giant terms into several parts than the way indicated: in general two intermediate objects can be isomorphic only if their denotations contain the same ingredients. More precisely, in step (*) we have assumed that $F G F b$ and $G F F b$ are isomorphic via $\gamma F$; an alternative is to assume that $F F G b$ and $G F F b$ are isomorphic, via $\gamma_{2,1}$. But since $\gamma_{2,1} = F \gamma : \gamma F$ and since the term $F \gamma$ is present at just the right place, this alternative assumption is equivalent to ours.

Here are the five assumptions made along the way (the equalities are assumed, the typing is provable):

\begin{align*}
\gamma & \text{ has inverse } \bar{\gamma} \\
\gamma : G \mu & = \mu G : \gamma_{2,1} : FG \to GFF \\
F \nu \cdot \gamma & = \gamma_{1,2} \cdot \nu F : FGG \to GF \\
\mu : \mu F & = \mu : F \mu : F \to FF \\
G \nu \cdot \nu & = \nu G : G \nu : GG \to G \\
\end{align*}

and

\begin{align*}
\gamma_{2,1} & = F \gamma : \gamma F : FFG \to GFF \\
\gamma_{1,2} & = \gamma G : G \gamma : FGG \to GGF
\end{align*}

where $\gamma_{2,1} = F \gamma : \gamma F$ and $\gamma_{1,2} = \gamma G : G \gamma : FGG \to GGF$.

Taking $F = Id$ and also $\mu, \gamma = id, id$, these assumptions specialise to those that make the Kleisli composition for $G, \nu$ associative.

The derivation of nice assumptions on $\eta$ in order that $id_P$ is neutral for $\eta$ is problematic. Let us first consider only the first equality in "$f : x \to \eta f = \eta f$". So, let $f : a \to b$ be arbitrary; then:

\begin{align*}
\eta f & \text{ has inverse } \bar{\eta} f \\
\equiv & \text{ definition } \mathcal{D} \\
F a \mu \eta F \eta F a & \xrightarrow{\mu} FFa \xrightarrow{\eta} GFa \xrightarrow{G \eta} GGB \xrightarrow{G \eta} Gb = \\
F a \mu F \eta F a & \xrightarrow{\mu} \eta F a \xrightarrow{\nu} FGa \xrightarrow{\gamma} G Fa \xrightarrow{G \eta} GGB \xrightarrow{G \eta} Gb \\
\Leftarrow & \text{ Leibniz } \{ \text{no nontrivial intermediate isomorphism seems plausible} \} \\
\text{(*)} & F \gamma : \gamma, G \eta = F \eta : \gamma, G \gamma \\
\Leftarrow & \text{ naturality } \eta : F \to G, \text{ so } F \gamma : \eta G = \eta F : G \gamma \\
\text{(*)} & \gamma : G \eta = \eta G \land F \eta : \gamma = \eta F
\end{align*}

Equation (\text{*}), for all $f$, is acceptably nice: it asserts a sort of naturality $F \to G$. Since we have already assumed natural transformation $\eta : F \to G$, it seems reasonable to require that $\eta : G \eta$ is $\eta G$, as in line (\text{*}). However, when instantiating with $F, \gamma := Id, id$, both line (\text{*}) and line (\text{*}) give requirements that are stronger than those for a monad: line (\text{*}) becomes $f : G \eta = \eta : G f$ (which is not just naturality $\eta : Id \to G$), and line (\text{*}) becomes
\( G\eta = \eta G \). So, we look for another sufficient condition that implies \( f \circ \text{id}_D = \text{id}_D \circ f \) for all \( f : a \to b \). The following line of reasoning has been suggested by Lambert Meertens.

\[
\begin{align*}
  f : & \quad \text{id}_D = \text{id}_D \circ f \\
\equiv & \quad \text{definition } D \\
  Fa \quad & \quad \mu \cdot F \gamma \cdot \eta \eta \cdot \nu = f \\
\equiv & \quad \text{assumption in previous calculation} \\
\text{(\star)} \quad & \quad \mu \cdot F f \cdot \eta G \cdot \nu = f \\
\equiv & \quad \text{naturality } \eta : F \to G \\
\text{(\star)} \quad & \quad \mu \cdot \eta F \cdot G f \cdot \nu = f .
\end{align*}
\]

Thus we are led to assume the equation of line (\star) (for all \( f : a \to b \)), or equivalently, of line (\star). However, both equations are too complicated to be called nice; in particular, the “\( f \)” occurs in the middle of the term and not at one end, as in naturality assertions. Fortunately, the instantiation with \( F, \gamma, \mu := \text{id}, \text{id}, \text{id} \) does give a monad law:

\[
\begin{align*}
  \mu \cdot F f \cdot \eta G \cdot \nu = f & \quad \text{for all } f : a \to b \\
\equiv & \quad \text{substitution } F, \gamma, \mu := \text{id}, \text{id}, \text{id} \\
  f : & \quad \eta G \cdot \nu = f & \quad \text{for all } f : a \to Gb \\
\equiv & \quad \text{for } \Leftarrow : \text{Leibniz; } \\
  & \quad \text{for } \Rightarrow : \text{take } a, f := Gb, \text{id}_{Gb} \\
  \eta G \cdot \nu = & \quad \text{id} G .
\end{align*}
\]
Summary. Let \( F, G \) be functors. Then we call \((F, G, \gamma, \eta, \mu, \nu)\) a dyad if:

\[
\begin{align*}
\gamma &: FG \to GF \\
\eta &: F \to G \\
\mu &: F \to FF \\
\nu &: GG \to G 
\end{align*}
\]

satisfy the following conditions:

\[
\begin{align*}
\mu \circ \mu F &= \mu \circ F \mu : F \to FF \\
G \nu \circ \nu &= \nu G \circ \nu : GG \to G \\
\gamma \circ G \eta \circ \nu &= \eta G \circ \nu \\
\mu \circ F \eta \circ \gamma &= \mu \circ \eta F \\
\mu \circ F f \circ \eta G \circ \nu &= f \quad \text{for all } f: a \to b \\
\gamma \text{ has inverse } \bar{\gamma} \\
\gamma \circ G \mu &= \mu \circ G, \gamma_{2,1} : FG \to GFF \\
F \nu \circ \gamma &= \gamma_{1,2} \circ \nu F : FGG \to GF 
\end{align*}
\]

where

\[
\begin{align*}
\gamma_{2,1} &= F \gamma \circ \gamma F : FFG \to GFF \\
\gamma_{1,2} &= \gamma G \circ G \gamma : FGG \to GGF
\end{align*}
\]

After substituting \( F, \gamma, \mu := Id, id, id \), these requirements are equivalent to the statement that \((G, \eta, \nu)\) is a monad.

References
