

## Yoneda lemma

We begin with an intuitive introduction to the mathematical content of Yoneda lemma (Lawvere and Rosebrugh, 2003, pp. 175-176, 249). With simple illustrations of figures-and-incidences (along with [its dual] properties-and-determinations) interpretations of mathematical objects, we prove the Yoneda lemma (Lawvere and Schanuel, 2009, pp. 361, 370-371). Broadly speaking, Yoneda lemma is about [properties of] objects [of a category] and their mutual determination.

First, let us consider a function

$$f: A \rightarrow B$$

We can think of the function  $f$  as (i) a figure of shape  $A$  in  $B$ , i.e., an  $A$ -shaped figure in  $B$ . For example, in the category of graphs, a map

$$d: D \rightarrow G$$

from a graph  $D$  (consisting of one dot) to any graph  $G$  is a  $D$ -shaped figure in  $G$ , i.e., a dot in the graph  $G$ . We can also think of the same function  $f$  as (ii) a property of  $A$  with values in  $B$ , i.e., a  $B$ -valued property of  $A$  (Lawvere and Schanuel, 2009, pp. 81-85). For example, with sets, say,  $\text{Fruits} = \{\text{apple, grape}\}$  and  $\text{Color} = \{\text{red, green}\}$ , a function

$$c: \text{Fruits} \rightarrow \text{Color}$$

(with  $c(\text{apple}) = \text{red}$  and  $c(\text{grape}) = \text{green}$ ) can be viewed as  $\text{Color}$ -valued property of  $\text{Fruits}$ .

Now let us consider two figures: an  $X$ -shaped figure in  $A$

$$x_A: X \rightarrow A$$

and a Y-shaped figure in A

$$y_A: Y \rightarrow A$$

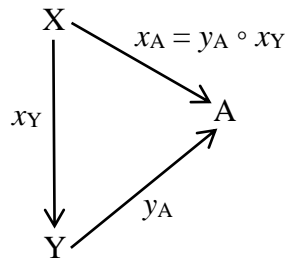
Given a transformation from the shape X to the shape Y, i.e. an X-shaped figure in Y

$$x_Y: X \rightarrow Y$$

we find that the X-shaped figure in Y ( $x_Y$ ) induces a transformation of a Y-shaped figure in A into an X-shaped figure in A via composition of maps

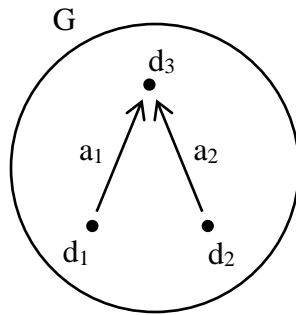
$$y_A \circ x_Y = x_A$$

(where ‘ $\circ$ ’ denotes composition) displayed as a commutative diagram

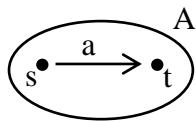


showing the transformation of a Y-shaped figure in A ( $y_A$ ) into an X-shaped figure in A ( $x_A$ ) by an X-shaped figure in Y ( $x_Y$ ) via composition of maps.

As an illustration, consider an object (of the category of graphs) i.e., a graph G (shown below):



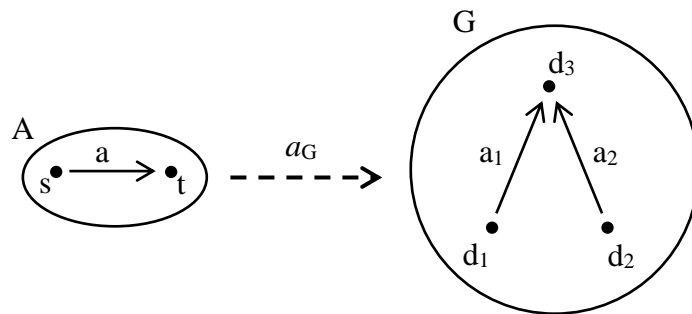
and a shape graph [arrow] A with exactly one arrow 'a', along with its source 's' and target 't', as shown:



along with an A-shaped figure in G

$$a_G: A \rightarrow G$$

displayed as:



with, say,

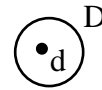
$$a_G(a) = a_1$$

This A-shaped figure in G, i.e. the graph map  $a_G$  maps the [only] arrow 'a' in the shape graph A to the arrow 'a<sub>1</sub>' in the graph G, while respecting the source (s) and target (t) structure of the arrow 'a', i.e., with arrow 'a' in shape A mapped to arrow 'a<sub>1</sub>' in the graph G, the source 's' and

target 't' of the arrow 'a' are mapped to the source 'd<sub>1</sub>' and target 'd<sub>3</sub>' of arrow 'a<sub>1</sub>', respectively.

Next, consider another shape graph [dot] D with exactly one dot 'd'

along with a D-shaped figure in A

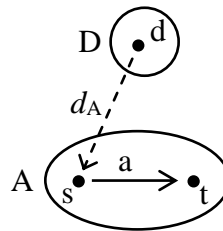


$$d_A: D \rightarrow A$$

with

$$d_A(d) = s$$

i.e., the graph map  $d_A$  maps the dot 'd' in the graph D to the dot 's' in the graph A, i.e. the source dot 's' of the arrow 'a', as shown below:



This graph map  $d_A$  from shape D to shape A induces a transformation of the (above) A-shaped figure in graph G

$$a_G: A \rightarrow G$$

into a D-shaped figure in G

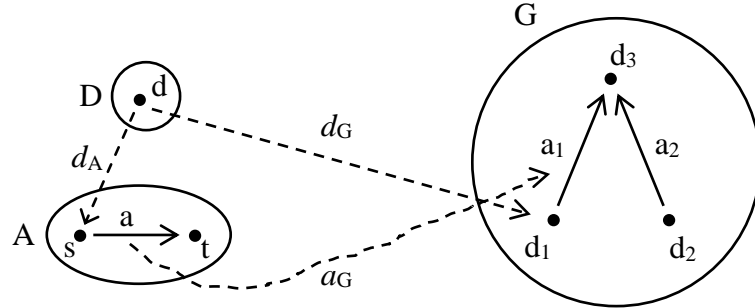
$$d_G: D \rightarrow G$$

via composition of graph maps

$$d_G = a_G \circ d_A$$

i.e.,  $d_G(d) = a_G \circ d_A(d) = a_G(s) = d_1$

as depicted below (Lawvere and Schanuel, 2009, pp. 149-150):



In general, every X-shaped figure in Y transforms a Y-shaped figure in A into an X-shaped figure in A i.e., every map

$$x_Y: X \rightarrow Y$$

induces a map in the opposite direction (contravariant; Lawvere, 2017; Lawvere and Rosebrugh, 2003, p. 103; Lawvere and Schanuel, 2009, p. 338)

$$A^{x_Y}: A^Y \rightarrow A^X$$

where  $A^Y$  is the map object of the totality of all Y-shaped figures in A,  $A^X$  is the map object of the totality of all X-shaped figures in A, and with the map  $A^{x_Y}$  of map objects defined as

$$A^{x_Y}(y_A: Y \rightarrow A) = y_A \circ x_Y = x_A: X \rightarrow A$$

assigning a map  $x_A$  in the map object  $A^X$  to each map  $y_A$  in the map object  $A^Y$ . Thus, the figures in an object A of all shapes (all X-shaped figures in A for every object X of a category) along with their incidences

$$A^{xY}: A^Y \rightarrow A^X$$

induced by all changes of figure shapes

$$x_Y: X \rightarrow Y$$

(i.e. every map in the category) together constitute the geometry of figures in A, i.e., a complete picture of the object A. Summing up, we have the complete characterization of the geometry of every object A of a category in terms of the figures of all shapes (objects of the category) and their incidences (induced by the maps of the category) in the object A (Lawvere and Schanuel, 2009, pp. 370-371).

Let us now examine how figures of a shape X in an object A are transformed into figures of the [same] shape X in an object B. We find that an A-shaped figure in B

$$a_B: A \rightarrow B$$

induces a transformation of an X-shaped figure in A

$$x_A: X \rightarrow A$$

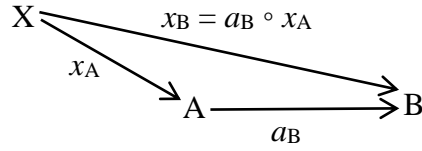
into an X-shaped figure in B

$$x_B: X \rightarrow B$$

via composition of maps

$$a_B \circ x_A = x_B$$

displayed as a commutative diagram (shown below):



showing the transformation of an X-shaped figure in A ( $x_A$ ) into an X-shaped figure in B ( $x_B$ ) by an A-shaped figure in B ( $a_B$ ) via composition of maps. Thus, every map

$$a_B: A \rightarrow B$$

induces a map in the same direction (covariant; Lawvere and Rosebrugh, 2003, pp. 102-103, 109; Lawvere and Schanuel, 2009, p. 319)

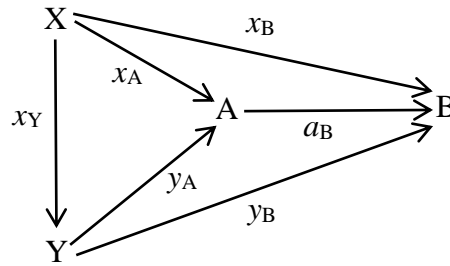
$$a_B^X: A^X \rightarrow B^X$$

where  $A^X$  is the map object of all X-shaped figures in A,  $B^X$  is the map object of all X-shaped figures in B, and with the map  $a_B^X$  defined as

$$a_B^X(x_A: X \rightarrow A) = a_B \circ x_A = x_B: X \rightarrow B$$

assigning a map  $x_B$  in the map object  $B^X$  to each map  $x_A$  in the map object  $A^X$ . Thus, the totality of maps  $a_B^X$  of map objects (for all objects and maps of the category) induced by a map  $a_B$  from A to B constitutes a covariant transformation of the figure geometry of object A into that of B, i.e., specifies how figures-and-incidences in A are transformed into figures-and-incidences in B.

Putting together these two transformations: (i) the covariant transformation of X-shaped figures in A into X-shaped figures in B induced by an A-shaped figure in B, and (ii) the contravariant transformation of Y-shaped figures in A into X-shaped figures in A induced by an X-shaped figure in Y, we obtain the diagram (Lawvere and Schanuel, 2009, p. 370):



from which we notice that there are two paths to go from a Y-shaped figure in A ( $y_A$ ) to an X-shaped figure in B ( $x_B$ ):

Path 1. First we map the Y-shaped figure in A ( $y_A$ ) into an X-shaped figure in A ( $x_A$ ) along the X-shaped figure in Y ( $x_Y$ ) via composition of the maps

$$y_A \circ x_Y$$

and then map the composite X-shaped figure in A ( $y_A \circ x_Y$ ) into an X-shaped figure in B along the A-shaped figure in B ( $a_B$ ) via composition

$$a_B \circ (y_A \circ x_Y)$$

Path 2. First we map the Y-shaped figure in A ( $y_A$ ) into a Y-shaped figure in B ( $y_B$ ) along the A-shaped figure in B ( $a_B$ ) via composition of the maps

$$a_B \circ y_A$$

and then map the composite Y-shaped figure in B ( $a_B \circ y_A$ ) into an X-shaped figure in B along the X-shaped figure in Y ( $x_Y$ ) via composition

$$(a_B \circ y_A) \circ x_Y$$



Based on the associativity of composition of maps (Lawvere and Schanuel, 2009, pp. 370-371), we find that

$$a_B \circ (y_A \circ x_Y) = (a_B \circ y_A) \circ x_Y$$

i.e., the two paths of transforming a Y-shaped figure in A

$$y_A: Y \rightarrow A$$

into an X-shaped figure in B give the same map

$$a_B \circ y_A \circ x_Y = x_B: X \rightarrow B$$

Since the associativity of composition of maps hold for all maps of any category (Lawvere and Schanuel, 2009, p. 17), we find that every A-shaped figure in B induces a covariant transformation of the figure geometry of A into the figure geometry of B. More explicitly, each A-shaped figure in B

$$a_B: A \rightarrow B$$

induces a commutative diagram (of maps of map objects)

$$\begin{array}{ccc}
 A^X & \xrightarrow{a_B^X} & B^X \\
 \uparrow A^{x_Y} & & \uparrow B^{x_Y} \\
 A^Y & \xrightarrow{a_B^Y} & B^Y
 \end{array}$$

satisfying

$$a_B^X \circ A^{x_Y} = B^{x_Y} \circ a_B^Y$$

for every map in the category, and hence a natural transformation (compatible with the composition of maps) of the figure geometry of A into the figure geometry of B. To see the commutativity, consider a Y-shaped figure in A, i.e. a map  $y_A$  of the map object  $A^Y$  and evaluate the above two composites:

$$a_B^X \circ A^{xY} (y_A) = a_B^X (y_A \circ x_Y) = a_B \circ (y_A \circ x_Y)$$

$$B^{xY} \circ a_B^Y (y_A) = B^{xY} (a_B \circ y_A) = (a_B \circ y_A) \circ x_Y$$

Again, according to the associativity of the composition of maps

$$a_B \circ (y_A \circ x_Y) = (a_B \circ y_A) \circ x_Y = a_B \circ y_A \circ x_Y$$

and hence both composites map each Y-shaped figure in A (a map in the map object  $A^Y$ )

$$y_A: Y \rightarrow A$$

to the X-shaped figure in B (a map in the map object  $B^X$ )

$$a_B \circ y_A \circ x_Y = x_B: X \rightarrow B$$

Since we have the above commutativity for every shape (object) and figure (map), i.e. for all objects and maps of the category, we conclude that an A-shaped figure in B corresponds to a natural transformation (respectful of figures-and-incidences) of the figure geometry of A into the figure geometry of B.

Now we formally show that every A-shaped figure in B

$$a_B: A \rightarrow B$$

of a category  $\mathcal{C}$  can be represented as a natural transformation

$$n^{aB}: \mathcal{C}(-, A) \rightarrow \mathcal{C}(-, B)$$

from the domain functor  $\mathcal{C}(-, A)$  constituting the figure geometry of the object A to the codomain functor  $\mathcal{C}(-, B)$  constituting the figure geometry of the object B, which is the core mathematical content of the Yoneda lemma (Lawvere and Rosebrugh, 2003, p. 249): “maps in any category can be represented as natural transformations” (Lawvere and Schanuel, 2009, p. 378). Since natural transformations represent structure-preserving maps between objects, the domain (codomain) functor of a natural transformation represents the domain (codomain) object of the structure-preserving map.

Let us define the (domain) functor

$$\mathcal{C}(-, A): \mathcal{C} \rightarrow \mathcal{C}$$

as: for each object X of the category  $\mathcal{C}$

$$\mathcal{C}(-, A)(X) = A^X$$

where  $A^X$  is the map object of all X-shaped figures in A

$$x_A: X \rightarrow A$$

and, for each map

$$x_Y: X \rightarrow Y$$

of the category  $\mathcal{C}$

$$\mathcal{C}(-, A)(x_Y: X \rightarrow Y) = A^{x_Y}: A^Y \rightarrow A^X$$

where  $A^Y$  is the map object of all Y-shaped figures in  $A$ , and with the map  $A^{x_Y}$  of map objects defined as

$$A^{x_Y} (y_A: Y \rightarrow A) = y_A \circ x_Y = x_A: X \rightarrow A$$

assigning a map  $x_A$  in the map object  $A^X$  to each map  $y_A$  in the map object  $A^Y$ . Thus the functor

$$C(-, A): C \rightarrow C$$

in assigning to each map

$$x_Y: X \rightarrow Y$$

(of the domain category  $C$ ) its [induced] map [of map objects]

$$C(-, A) (x_Y: X \rightarrow Y) = C(-, A) (Y) \rightarrow C(-, A) (X) = A^{x_Y}: A^Y \rightarrow A^X$$

(of the codomain category  $C$ ) is contravariant, i.e. a transformation of a shape  $X$  into a shape  $Y$  induces a transformation (in the opposite direction) of  $Y$ -shaped figures in  $A$  into  $X$ -shaped figures in  $A$  (Lawvere and Rosebrugh, 2003, pp. 236-237).

Now, we check to see if  $C(-, A)$  preserves identities, i.e. whether

$$C(-, A) (1_X: X \rightarrow X) = 1_{C(-, A)(X)}$$

for every object  $X$ . Evaluating

$$C(-, A) (1_X: X \rightarrow X) = A^{1_X}: A^X \rightarrow A^X$$

at a map

$$x_A: X \rightarrow A$$

we find that

$$A^{1_X} (x_A: X \rightarrow A) = (x_A \circ 1_X) = x_A: X \rightarrow A$$

(for every map  $x_A$  in the map object  $A^X$ ). Next, evaluating

$$1_{C(-, A)(X)} = 1_{A^X}: A^X \rightarrow A^X$$

at the map

$$x_A: X \rightarrow A$$

we find that

$$1_{A^X} (x_A: X \rightarrow A) = (x_A \circ 1_X) = x_A: X \rightarrow A$$

(for every map  $x_A$  in the map object  $A^X$ ). Since

$$A^{1_X} = 1_{A^X}$$

i.e.

$$C(-, A)(1_X: X \rightarrow X) = 1_{C(-, A)(X)}$$

for every object  $X$  of the category  $\mathcal{C}$ , we say  $C(-, A)$  preserves identities.

Next, we check to see if  $C(-, A)$  preserves composition. Since  $C(-, A)$  is contravariant, we check whether

$$C(-, A)(y_Z \circ x_Y) = C(-, A)(x_Y) \circ C(-, A)(y_Z)$$

where  $y_Z: Y \rightarrow Z$ . Evaluating

$$\mathbf{C}(-, \mathbf{A})(y_Z \circ x_Y) = \mathbf{A}^{(y_Z \circ x_Y)}$$

at any map  $z_A$  in the map object  $\mathbf{A}^Z$ , we find that

$$\mathbf{A}^{(y_Z \circ x_Y)}(z_A) = z_A \circ (y_Z \circ x_Y)$$

Next, we evaluate

$$\mathbf{C}(-, \mathbf{A})(x_Y) \circ \mathbf{C}(-, \mathbf{A})(y_Z) = (\mathbf{A}^{x_Y} \circ \mathbf{A}^{y_Z})$$

also at the map  $z_A$

$$(\mathbf{A}^{x_Y} \circ \mathbf{A}^{y_Z})(z_A) = \mathbf{A}^{x_Y}(z_A \circ y_Z) = (z_A \circ y_Z) \circ x_Y$$

Since

$$z_A \circ (y_Z \circ x_Y) = (z_A \circ y_Z) \circ x_Y$$

by the associativity of the composition of maps, we have composition preserved

$$\mathbf{C}(-, \mathbf{A})(y_Z \circ x_Y) = \mathbf{C}(-, \mathbf{A})(x_Y) \circ \mathbf{C}(-, \mathbf{A})(y_Z)$$

Having checked that

$$\mathbf{C}(-, \mathbf{A}): \mathbf{C} \rightarrow \mathbf{C}$$

with

$$\mathbf{C}(-, \mathbf{A})(X) = \mathbf{A}^X$$

$$\mathbf{C}(-, \mathbf{A})(x_Y: X \rightarrow Y) = \mathbf{A}^{x_Y}: \mathbf{A}^Y \rightarrow \mathbf{A}^X$$

where  $\mathbf{A}^{x_Y}(y_A) = y_A \circ x_Y$ , is a contravariant functor, we consider another contravariant functor

$$C(-, B): C \rightarrow C$$

with

$$C(-, B)(X) = B^X$$

$$C(-, B)(x_Y: X \rightarrow Y) = B^{x_Y}: B^Y \rightarrow B^X$$

where  $B^{x_Y}(y_B) = y_B \circ x_Y$ .

With the two functors  $C(-, A)$  and  $C(-, B)$  representing the [figure geometry of] objects  $A$  and  $B$ , respectively, we now show that every structure-preserving map

$$a_B: A \rightarrow B$$

is represented by a natural transformation

$$n^{a_B}: C(-, A) \rightarrow C(-, B)$$

More explicitly, given a map  $a_B$ , we can construct a natural transformation  $n^{a_B}$ . A natural transformation  $n^{a_B}$  from the functor  $C(-, A): C \rightarrow C$  to the functor  $C(-, B): C \rightarrow C$  assigns to each object  $X$  of the domain category  $C$  (of both domain and codomain functors) a map

$$a_B^X: A^X \rightarrow B^X$$

(in the common codomain category  $C$ ) from the value of the domain functor at the object  $X$ , i.e.  $C(-, A)(X) = A^X$  to the value of the codomain functor at  $X$ , i.e.  $C(-, B)(X) = B^X$ ; and to each map  $x_Y: X \rightarrow Y$  (in the common domain category  $C$ ), a commutative square (in the common codomain category  $C$ ) shown below:

$$\begin{array}{ccc}
 A^X & \xrightarrow{a_B^X} & B^X \\
 \uparrow A^{xY} & & \uparrow B^{xY} \\
 A^Y & \xrightarrow{a_B^Y} & B^Y
 \end{array}$$

satisfying

$$a_B^X \circ A^{xY} = B^{xY} \circ a_B^Y$$

(Lawvere and Rosebrugh, 2003, p. 241; Lawvere and Schanuel, 2009, pp. 369-370). We have already seen that with the composition-induced maps (of map objects):

$$A^{xY}(y_A) = y_A \circ x_Y$$

$$a_B^X(x_A) = a_B \circ x_A$$

$$a_B^Y(y_A) = a_B \circ y_A$$

$$B^{xY}(y_B) = y_B \circ x_Y$$

the required commutativity:

$$a_B^X \circ A^{xY}(y_A) = a_B^X(y_A \circ x_Y) = a_B \circ (y_A \circ x_Y)$$

$$B^{xY} \circ a_B^Y(y_A) = B^{xY}(a_B \circ y_A) = (a_B \circ y_A) \circ x_Y$$

is given by the associativity of the composition of maps



$$a_B \circ (y_A \circ x_Y) = (a_B \circ y_A) \circ x_Y = a_B \circ y_A \circ x_Y$$

Thus, each A-shaped figure in B ( $a_B$ ) is a natural transformation ( $n^{a_B}$ ; homogenous with respect to composition of maps) of the figure geometry  $\mathcal{C}(-, A)$  of A into the figure geometry  $\mathcal{C}(-, B)$  of B.

Furthermore, we can obtain the set  $|B^A|$  of all A-shaped figures in B based on the 1-1 correspondence between A-shaped figures in B and the points (i.e. maps with terminal object T of the category  $\mathcal{C}$  as domain; Lawvere and Schanuel, 2009, pp. 232-234) of the map object  $B^A$ . This 1-1 correspondence, which follows from the universal mapping property defining exponentiation, along with the fact that the terminal object T is a multiplicative identity (Lawvere and Schanuel, 2009, pp. 261-263, 313-314, 322-323), involves the following two 1-1 correspondences between three maps:

$$\begin{array}{c} T \rightarrow B^A \\ \hline T \times A \rightarrow B \\ \hline A \rightarrow B \end{array}$$

Yoneda lemma says, in terms of our figures-and-incidences characterization of objects, that the set  $|B^A|$  of A-shaped figures in B

$$a_B: A \rightarrow B$$

is isomorphic to the set  $|\mathcal{C}(-, B)^{\mathcal{C}(-, A)}|$  of natural transformations

$$n^{a_B}: \mathcal{C}(-, A) \rightarrow \mathcal{C}(-, B)$$

of the figure geometry of A into that of B. The required isomorphism of sets

$$|B^A| = |C(-, B)^{C(-, A)}|$$

follows from the 1-1 correspondence between A-shaped figures in B and the natural transformations (compatible with all figures and their incidences) of the figure geometry of A into that of B, which we have already shown (see also Lawvere and Rosebrugh, 2003, p. 104, 174).

Dually, a map

$$A \rightarrow B$$

viewed as a B-valued property on A induces a natural transformation

$$C(B, -) \rightarrow C(A, -)$$

of the function algebra of B into that of A (Lawvere and Rosebrugh, 2003, p. 249). Here also the proof of Yoneda lemma involves two transformations: (i) Contravariant: a map from an object A to an object B induces a transformation of properties of B into properties of A, for each type (object) of the category, and (ii) Covariant: a map from a type T to a type R (of properties) induces a transformation of T-valued properties into R-valued properties, for every object of the category. The calculations involved in proving Yoneda lemma in this case of function algebras are same as in the case of figure geometries, except for the reversal of arrows due to the duality between function algebra and figure geometry (Lawvere and Rosebrugh, 2003, p. 174; Lawvere and Schanuel, 2009, pp. 370-371). More specifically, function algebras and figure geometries are related by adjoint functors (Lawvere, 2016).

## References

Lawvere, F. W. (2016) Birkhoff's theorem from a geometric perspective: A simple example, *Categories and General Algebraic Structures with Applications* 4: 1-7.

Lawvere, F. W. (2017) Everyday physics of extended bodies or why functionals need analyzing, *Categories and General Algebraic Structures with Applications* 6: 9-19.

Lawvere, F. W. and Rosebrugh, R. (2003) *Sets for Mathematics*, Cambridge, UK: Cambridge University Press.

Lawvere, F. W. and Schanuel, S. H. (2009) *Conceptual Mathematics: A First Introduction to Categories*, (2nd ed.) Cambridge, UK: Cambridge University Press.