SOME REMARKS CONCERNING CATEGORIES AND SUBSPACES

J. R. ISBELL

Introduction. This paper is primarily a brief elaboration on the axioms for a *bicategory* introduced in **(3)**. From this point of view, the main aim is the development of the structure of certain systems of topological and uniform spaces, and the present paper merely points out some very general properties which follow from axioms so weak that they are satisfied by any system likely to be considered. However, from the point of view of the general theory of categories, the main content of this paper consists of a definition and certain technical observations which tend to justify the particular axioms used. The following remarks must serve as introduction for both viewpoints.

A category is an algebroid system analogous to a group. Rather than define a category here we define a category of mappings, which is analogous to a group of transformations. A category of mappings (Q, A, B) consists of a collection Q of sets, called *spaces*, a collection A of functions on spaces into spaces, called *mappings*, and a subset B of $A \times A \times A$ consisting of those triples (f, g, h) such that h is the composed mapping $g \circ f$. The sole requirements are that A is closed under composition and contains, for each space X in Q, the identity function i: $X \to X$. In particular, a group A of transformations on a set X forms a category if we take $Q = \{X\}$ and $B = \{(a, b, ba) | a \text{ and } b \text{ in } A\}$.

With many mathematical structures there are naturally associated categories of mappings. For example, with a collection Q of groups we may associate the category A of all homomorphisms on elements of Q into elements of Q. Many natural correspondences involve transformations of one category into another. This is most familiar in algebraic topology; for example, in associating with a space X a homology group $H(X)$ one also associates to each continuous function f: $X \to Y$ a homomorphism f': $H(X) \to H(Y)$. However, the phenomenon is also common in general topology. For example, the Stone-Čech compactification induces such a transformation. Passage from a space X to its ring of real-valued continuous functions $C(X)$ is an instance of a *contravariant* transformation, in that a function f: $X \to Y$ induces a homomorphism in the opposite direction, f^*: $C(Y) \to C(X)$.

A category is in the first place an abstract algebra, or at least an abstract structure resembling an algebra. The first section of this paper establishes

Received January 11, 1957. Research supported in part by the Office of Naval Research under Contract N7onr 41904, George Washington University Logistics Research Project, and in part by a National Science Foundation fellowship.

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

some simple propositions on homomorphisms and congruence relations in categories. For example (as in algebras), a one-to-one homomorphism is an isomorphism (1.1). However, the general homomorphism is not determined by the congruence relation which it induces (1.2).

For the applications one may wish to consider more structure than is given by the law of composition. For example, in a category of mappings (Q, A, B) one may be concerned with those mappings $f\colon X \to Y$ which embed X as a subspace of Y. Such mappings have special properties expressible in terms of the algebra of composition; for example, such an f always satisfies the cancellation law $fg = fh$ implies $g = h$. In any particular category the concept of "subspace" may or may not be definable in terms of the algebra of composition. The question has been considered how to impose axioms on a subset I of A so that I may reasonably be interpreted to be the set of all embeddings of subspaces. Axioms have been given (involving more than this) by MacLane in **(5)** and by the author in **(3)**; the more elaborate structure so defined is called a *bicategory*.

The second section of this paper is a study of conditions on a subset I of a category A in order that A may be represented as a category of mappings in such a way that I is just the class of embeddings of subspaces. First we consider conditions for an *isomorphic* representation of A so that mappings in I become actual inclusion functions $f\colon X \to Y$, where X is a subset of Y. Five conditions are taken from **(5)**, and a sixth is shown to be necessary. We sketch a proof that the six conditions are sufficient (2.2). But at this point we have an already formidable battery of axioms, which still do not cover all the primitive terms of bicategory theory. In search of simplicity we turn in another direction.

A *skeleton* of a category is a certain kind of subcategory. Given a category of mappings (Q, A, B), a skeleton is obtained as follows. A mapping $f \in A$ is an *isomorphism* provided f is one-to-one onto and the function f^{-1} is an element of A. Then let K be a subset of Q consisting of just one space from each isomorphism type. The set of all mappings in A whose domain and range are in K is a skeleton of A. (Any two skeletons of A are isomorphic categories.) Then we define two categories to be *coextensive* if they have isomorphic skeletons. We seek conditions on a subset I of A in order that A be coextensive with a category of mappings in such a way that the mappings in I correspond to functions gfh, where f is an inclusion function and g and h are isomorphisms. The necessary and sufficient conditions are (1) for f in I, $fg = fh$ implies $g = h$, and (2) I contains all isomorphisms and is closed under composition with isomorphisms (2.4).

The third section of the paper gives the bicategory axioms of **(3)**, with a few elementary consequences and some discussion of examples. In particular, modulo the identification of coextensive categories, the subspace concept in groups and in compact spaces is definable in terms of the algebra of composition. This is not true in MacLane's more delicate theory **(5)**. It is not

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

proposed to supplant categorical isomorphism with coextension. However, it is suggested that considerable work remains to be done, at least in general topology, in the study of coextensive invariants of categories of continuous functions. For such work the present system of axioms has substantial advantages.

The author is indebted for discussions and suggestions, particularly to Saunders MacLane, and also to James Case, Pierre Conner, Melvin Henriksen, and Dana Scott.

I. Categories. We begin with a formal definition of a category which is virtually the same as the one given in **(2)**.

Definition. A *category* is an ordered pair (A, B) of sets, where B is a subset of $A \times A \times A$ and the following conditions are met.

(a) For each f, g, in A, there is at most one h in A such that (f, g, h) is in B; such an h is designated gf.

(b) For each f in A there exist (i) at least one i in A such that if exists and for all x in A, (1) if ix exists then $ix = x$, and (2) if xi exists then $xi = x$; and (ii) at least one j in A satisfying (1) and (2) and such that fj exists.

(c) (i) If fg and gh exist then $(fg)h$ exists, $f(gh)$ exists, and $(fg)h = f(gh)$; (ii) if $(fg)h$ exists then gh exists; (iii) if $f(gh)$ exists then fg exists.

Uniqueness of the i and j of condition (b) follows, as shown herewith. Let us call an element of A an *identity* if it satisfies the conditions (1) and (2). Suppose $i'f$ exists, and $if = f$. Then $i'f = i'(if)$, and $i'i$ exists by (c). If i is an identity then $i'i = i'$ and i' is not an identity unless $i' = i$.

The axioms are satisfied by any semigroup A with unit. However, that is not the most interesting sort of category. The sort of "category" one would like to study is illustrated by the "collection" of all continuous functions $f: X \to Y$, where X and Y are compact spaces and the composition gf is the functional composition $g \circ f$. Such a collection of course involves the paradoxes of set theory.

A perfectly proper description of categories which are too large to be sets can be given, for example, in terms of Hilbert-Bernays set theory. Eilenberg and MacLane pointed this out in **(2)**, and MacLane actually carried it out in **(5)**. Until the theory develops further it seems reasonable to duck the complications involved in this development, so far as possible. In this paper we can do this, in spite of the fact that we are concerned primarily with applications to proper classes. All the theorems are stated for sets. In most cases the application may properly be interpreted along the following simple line: a proposition asserted, for example, for (the class of) all continuous functions may as well be asserted for (the set of) all continuous functions on spaces whose points are a subset of a fixed set S, for each S. This interpretation is not right for the representation theorems; generalization of 2.4 or of 3.5, for example, to apply to proper classes, is an unsolved problem. Aside from this, the entire argument could be carried out in Zermelo set theory.

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

To introduce another convention: a category may reasonably be regarded as an ordered triple (Q, A, B), where Q is a collection of *spaces*, A a collection of *mappings*, and B a subset of $A \times A \times A$ giving the law of composition in A. Since the algebra of mappings is the center of interest, we have defined a category as a pair (A, B); one or another set Q of spaces may be considered to provide a *representation*. In conformity with algebraic (and topological) usage, we may speak of A alone as the category, letting the law of composition be understood. However, in examples, we may name Q alone, as in "the category of all groups"; in such a case it is to be understood that A consists of the usual mappings of such objects (if Q consists of the groups then A consists of their homomorphisms), and B gives the usual law of functional composition. Note, though, that other classes of mappings may be explicitly indicated, and in particular, in speaking of a subcategory there is no presumption that all possible mappings are included. For example, it may be convenient to refer to a subcategory consisting of one group G, one subgroup H, and one isomorphism of H into G.

Yet another convention: a *function* $f: G \to H$ is an ordered triple (f, G, H), where f is a single-valued relation in $G \times H$, G is the set of arguments of f, and H contains the set $f(G)$ of values of f. H is called the *range* of f; $f(G)$ has no particular name. In loose talk we may call $f(G)$ the image of G or of f, but we need the technical term *image* for another use.

In an abstract category A the terms *domain* and *range* are applied to the handiest objects which suggest the domain and range of a function. Specifically, the *domain* of f, $\delta(f)$, is that identity i such that fi exists; and the range $\rho(f)$ is that identity j such that jf exists.

Note that a category may be regarded as an "algebra" with one operation fg, or with three operations, including δ and ρ. In either case it is not precisely an algebra, since fg is not defined for all pairs. However, with the three operations one has a structure which is quite nearly algebraic; the necessary and sufficient condition for the existence of fg is that $\delta(f) = \rho(g)$. (Proof omitted.) One could throw in a zero and define $fg = 0$ if fg is not otherwise defined; however, $\delta(0)$ and $\rho(0)$ would raise new problems. So far as is known, the structure of categories is not adequately described by any strictly algebraic formulation.

Eilenberg and MacLane have shown (**2**, Appendix) that every abstract category may be represented as a category of sets and functions. Specifically, a *concrete category* is defined as an ordered pair (Q, A), where A is a set of functions on elements of Q into elements of Q, and the axioms are

0. For each f in A, the domain and range of f are in Q.

1. Every identity function $i: X \to X$ whose domain is in Q is a member of A.

2. For any $f: X \to Y$ and $g: Y \to Z$ in A, $gf: X \to Z$ is in A.

Either Q or A may be called the category when the meaning is apparent. Every concrete category (Q, A) determines an abstract category (A, B) in the obvious way; the representation theorem is that every abstract category

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

(A', B') is *isomorphic* with such a category, where isomorphism has the obvious meaning.

Specifically, an isomorphism of (A', B') upon (A, B) consists of a one-to-one correspondence τ of A' onto A such that the induced correspondence of $A' \times A' \times A'$ onto A^3,

$$(f, g, h) \rightarrow (\tau(f), \tau(g), \tau(h)),$$

maps B' onto B. A *homomorphism* is a mapping $\tau\colon A' \rightarrow A$ satisfying (a) if fg exists in A' then $\tau(f)\,\tau(g)$ exists and is $\tau(fg)$, and (b) if f is an identity in A' then $\tau(f)$ is an identity in A. One may replace (b) (in the presence of (a); proof omitted) with the conditions $\delta\tau = \tau\delta$ and $\rho\tau = \tau\rho$. A *subcategory* of A is a subset closed under composition, δ, and ρ. Clearly every intersection of subcategories is a subcategory; thus every subset generates a subcategory, and in particular for each homomorphism $\tau\colon A \rightarrow A'$ there is a least subcategory containing $\tau(A)$, which is called the *image* of A under τ. The homomorphism τ also determines an equivalence relation $\mathbf{r}$ in A, $x\mathbf{r}y$ if $\tau(x) = \tau(y)$; an equivalence relation obtainable in this way is called a *congruence relation*. A homomorphic image B of A is called an *identification category* of A, and $\tau\colon A \rightarrow B$ an *identification mapping*, in case the following is true: whenever $\sigma\colon A \rightarrow C$ is a homomorphism such that the congruence relation $\mathbf{s}$ determined by σ contains the congruence relation $\mathbf{r}$ determined by τ, then there exists a homomorphism $\xi\colon B \rightarrow C$ such that $\xi \circ \tau = \sigma$.

The rest of this section is devoted to establishing the following results.

1.1. (First Isomorphism Theorem). *If $\tau\colon A \rightarrow A'$ is a one-to-one homomorphism then A is isomorphic with its image under τ.*

1.2. *Every homomorphism determines an identification category, not necessarily isomorphic with the image.*

1.3. *The congruence relations on A form a complete lattice $L(A)$. However, if $\mathbf{r}$ is a particular member of $L(A)$, and A' the identification category determined by $\mathbf{r}$, the lattice $L(A')$ and the sublattice of $L(A)$ consisting of all relations containing $\mathbf{r}$ need not be isomorphic.*

1.4. *For homomorphisms*

$$\tau : A \rightarrow A', \quad \alpha : A' \rightarrow A'', \quad \beta : A' \rightarrow A'',$$

if $\alpha \circ \tau = \beta \circ \tau$ then α and β coincide on the image of A. Hence if $\tau\colon A \rightarrow B$ is an identification mapping and $\sigma\colon A \rightarrow C$ is a homomorphism divisible by τ then the solution of $\xi\tau = \sigma$ is unique.

1.5. *Every homomorphic image of a category A is an identification category of an identification category of A.*

Proposition 1.1 is valid for algebras and is sometimes called the First Isomorphism Theorem. Sometimes such names are applied to certain theorems which are significant only for systems having a zero. At any rate, the negative

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

statements in 1.2 and 1.3 assure that none of the results commonly called the Second Isomorphism Theorem is valid for categories.

Proof of 1.1. Let τ be a one-to-one homomorphism of A onto the subset B of A'. Then B is closed under δ and ρ. (This is true even if τ is not one-to-one.) In B, the general element has the form $\tau(x)$; and

$$\tau(x)\ \tau(y) \text{ exists in } A' \Leftrightarrow \delta\tau(x) = \rho\tau(y) \Leftrightarrow \tau\delta(x) = \tau\rho(y)$$
$$\Leftrightarrow \delta(x) = \rho(y) \Leftrightarrow xy \text{ exists in } A \Leftrightarrow \tau(x)\ \tau(y) = \tau(xy) \text{ in } B.$$

Therefore B is a subcategory, τ is one-to-one onto B, and $\tau: A \to B$ is an isomorphism.

For 1.2 and 1.3 we need the lemma

1.6. *An equivalence relation in a category A which determines the set C of equivalence classes c is a congruence relation if and only if*

(1) *the set product cd of any two members of C is a subset of a member of C, and*

(2) *the sets $\delta(c)$ and $\rho(c)$ are subsets of members of C.*

Proof. Clearly a congruence relation has these properties. Conversely let the partition C satisfy (1) and (2). Let the category B consist of the members of C, and other elements to be described. For $c \in C$, $\delta(c)$ in A is a subset of some element of C; on the other hand, the set $\delta(c)$ is not empty, and thus it lies in a unique member $\delta'(c)$ of C. Similarly ρ in A induces an operator ρ' in C. Altogether let B consist of all ordered n-tuples (words) of elements of C, $(c_1, \ldots, c_n)$, such that for $1 \leqslant i \leqslant n - 1$, $\delta'(c_i) = \rho'(c_{i+1})$, but the product in A of the sets c_i, c_{i+1}, is empty. (That is, $\delta(c_i)$ and $\rho(c_{i+1})$ are disjoint subsets of the same element of C.) Define

$$\delta'(c_1, \ldots, c_n) = \delta'(c_n), \quad \rho'(c_1, \ldots, c_n) = \rho'(c_1).$$

The product in B of $(c_1, \ldots, c_n)$ and $(b_1, \ldots, b_m)$ is defined if and only if $\delta'(c_n) = \rho'(b_1)$. If $\delta(c_n) \cap \rho(b_1) = 0$ then the product is $(c_1, \ldots, c_n, b_1, \ldots, b_m)$. Otherwise $c_n b_1$ is a non-empty subset of a unique member d of C; and, suppressing an induction, we describe the product as the word $(c_1, \ldots, c_{n-1}, d, b_2, \ldots, b_m)$, contracted as far as possible by further multiplication.

It is easily seen that we have defined a category B. The function $\tau: A \to B$ which takes each member of A to its C-equivalence class is a homomorphism, and thus C defines a congruence relation. Furthermore, B is an identification category. We have finished the proof of 1.6 and begin on the

Proof of 1.2. Given the situation above, with the homomorphism $\tau: A \to B$; and given a homomorphism $\sigma: A \to D$ constant on each equivalence class c of the partition C; to construct a homomorphism $\xi: B \to D$ such that $\xi\tau = \sigma$. For one-letter words $c \in B$, let $\xi(c)$ be the constant value of $\sigma(x)$, for any $x \in c$ in A. Since σ is a homomorphism, therefore

$$\delta\xi(c) = \xi\delta(c), \qquad \rho\xi(c) = \xi\rho(c).$$

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

Then if (c_1,c_2) is a word in B, necessarily $\xi(c_1)\xi(c_2)$ exists in D. Define $\xi(c_1, c_2)$ to be $\xi(c_1)\xi(c_2)$; and so on by induction. By definition $\xi\tau = \sigma$, and clearly ξ is a homomorphism.

That the identification category need not be isomorphic with the image is perhaps obvious, but we give an example. The homomorphism cannot be one-to-one on identities, for the freedom in the image arises only where new products are defined. Accordingly consider the category A with four elements, x_0, x_1, y_0, y_1;

$$\rho(x_i) = \delta(x_i) = x_0, \qquad \rho(y_i) = \delta(y_i) = y_0, \qquad i = 0, 1;$$

thus x_0 and y_0 act as identities; and finally, $x_1^2 = x_1$, $y_1^2 = y_1$. (A typical realization of A is on a pair of linear spaces, each with its identity mapping and one projection upon a proper subspace.) Consider the category B with four elements, z_0, z_1, z_2, z_{12}, all idempotent, z_0 an identity, z_{12} a zero, $z_1z_2 = z_2z_1 = z_{12}$. There is a homomorphism $\tau: A \to B$ given by

$$\tau(x_0) = \tau(y_0) = z_0, \qquad \tau(x_1) = z_1, \qquad \tau(y_1) = z_2.$$

B is the image; but the identification category determined by τ is neither finite nor commutative.

For the proof of 1.3, it is clear from 1.6 that every intersection of congruence relations is a congruence relation. Hence any equivalence relation generates a least containing congruence relation, and $L(A)$ is a complete lattice. In the example in the proof of 1.2, $L(A)$ is a finite lattice, while the identification category clearly has infinitely many congruence relations. This finishes 1.3.

The proof of 1.4 is a trivial induction.

For 1.5 we establish a lemma.

1.7. *Let $\tau: A \to B$ be a homomorphism which is one-to-one on identities. Then the identification category determined by τ is isomorphic with the image of A under τ.*

Proof. From the proof of 1.6 we see that the homomorphism τ^* of A upon the identification category A' is onto unless for some equivalence classes, c_1, c_2, the sets $\delta(c_1)$ and $\rho(c_2)$ are disjoint subsets of the same equivalence class. This is impossible when τ is one-to-one on identities. But then the quotient homomorphism $\xi: A' \to B$ is one-to-one, since otherwise $\tau = \xi\tau^*$ would determine a larger congruence relation on A. Then 1.1 applies, and 1.7 is proved.

Proof of 1.5. From the proof of 1.6 we see that every identity in an identification category is the image of an identity in A. Hence the induced mapping of the identification category upon the image is one-to-one on identities, 1.7 applies, and 1.5 is proved.

II. Subspaces. In **(3)** there is given a simplified version of MacLane's axioms for a *bicategory* (crudely: a category with subspaces), which will be used in the concluding portion of this paper. The simplified version has a

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

somewhat different motivation than (may be presumed for) the original, and it seems likely that a combination of the two may survive. Basically, the simplification involves a broader notion of equivalence. In **(5)** MacLane investigates properties invariant under *isomorphism*. Below we define a relation of *coextension*, and we shall be concerned with coextensive invariants. Isomorphic categories or bicategories are coextensive, but not conversely.

This section illustrates the two viewpoints—mainly the cruder one—in examining the question what axioms must be imposed on subspaces in order that they behave like subsets.

A function $f: X \to Y$ is called an *inclusion* function provided X is a subset of Y and $f(x) = x$ for all x in X. Given a category A and a subset I of A, under what conditions can A be represented as a concrete category so that the mappings in I, and no others, become inclusion functions? Five clearly necessary conditions are

(1) every identity is in I,

(2) I is closed under composition (by (1) and (2), I is a subcategory),

(3) $f = gh$ with f and g in I implies h is in I,

(4) $fg = fh$ with f in I implies $g = h$, and

(5) I contains at most one element with given domain and range.

These conditions have been recognized by MacLane and incorporated mutatis mutandis into his axioms **(5)**. A sixth condition is necessary and, for the immediate question, sufficient; but we shall merely sketch the proof (2.2).

Let two elements of A, f and g, be called *equivalent* if there is a finite chain $(h_1, \ldots, h_n)$, $h_1 = f$, $h_n = g$, such that for $1 \leqslant i \leqslant n - 1$, either $h_i = j_i h_{i+1}$ or $h_{i+1} = j_i h_i$, for some j_i in I. Supposing I to be the set of inclusion functions of a concrete category, equivalence of f and g implies that f and g have the same domain and values. Therefore we may demand (6) two equivalent elements of A having the same range are identical.

Remark 2.1. The conditions (1)—(5) do not imply (6). In fact, one can construct a system satisfying all the axioms and conventions of **(5)** for bicategories, in which the class of *injections* (in the language of **(5)**) does not satisfy (6). The construction is straightforward but too tedious and unsurprising to give here.

Remark 2.2. The conditions (1)—(6) imply that A is isomorphic with a concrete category in such a way that I corresponds precisely to the inclusion functions. The reasons for omitting the somewhat lengthy proof are (a) that the result seems to be useless both in the context of MacLane's theory, where it is not strong enough, and in the context of this paper, where it is irrelevant; and (b) it is a mere modification of the Eilenberg-MacLane representation of **(2)**. In fact, A is partitioned into equivalence classes by the relation of equivalence defined above; carry through the Eilenberg-MacLane construction and then choose a representative f_0 of each equivalence class $[f]$, and replace each occurrence of f by f_0. The representation is preserved

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

because of assumption (6), the elements of I become inclusion functions because of (4), and there are no other inclusion functions because of (1)—(3). (Condition (5) is an easy consequence of (6).)

In any category A, f is said to be an *isomorphism* if there exists a mapping f^{-1} in A such that ff^{-1} is an identity and $f^{-1}f$ is an identity. In a concrete category we call f an *injection* if f has the form gih, where i is an inclusion function and g and h are isomorphisms. Two injections f and g are *equivalent* if there is an isomorphism j such that $fj = g$. An equivalence class of injections into X is called a *subspace* of X. (Thus a subspace has a fixed range, and (speaking imprecisely) a fixed "image," but the domain is determined only up to isomorphism.)

One might ask under what conditions a category A can be represented with a prescribed family I of injections. Clearly I must contain all isomorphisms and be closed under composition with isomorphisms. Further, the cancellation condition (4) above must hold. More arcane properties can be found, for example, if X has $\mathbf{m}$ subspaces isomorphic with Y ($\mathbf{m}$ a cardinal number), then A contains $\mathbf{m}$ spaces isomorphic with Y. But this is not what we want.

Accordingly we define a *skeleton* K of a category A as follows. A subcategory S of A is *full* in case the hypotheses $\delta(f) \in S$ and $\rho(f) \in S$ imply $f \in S$. Two identities, i, j, are *isomorphic* or *equivalent* if there exist isomorphisms f, g, such that $fg = i$ and $gf = j$. Then a *skeleton* is a full subcategory K including exactly one identity from each equivalence class.

2.3. *All skeletons of a category A are pairwise isomorphic.*

Proof. Let K, K' be two skeletons of A. For each identity i in K there is exactly one equivalent identity i' in K', and at least one isomorphism f in A such that $f^{-1}f = i, ff^{-1} = i'$. For each i in K choose one such f. For any g in K, let f_1 be the isomorphism associated with $\delta(g)$, f_2 the isomorphism associated with $\rho(g)$. Then $g' = f_2 g f_1^{-1}$ is in K', and the transformation $g \rightarrow g'$ is evidently an isomorphism.

We define two categories to be *coextensive* if they have isomorphic skeletons. In the ordinary parlance of algebra and topology, outside of homology theory, the distinction between coextensive categories is commonly ignored. This is not to say that it ought to be ignored; but one may properly investigate those properties of categories which are coextensive invariants.

If we prescribe a class I of *injections* in a category A, and I satisfies the modest requirement of including all isomorphisms and being closed under composition with isomorphisms, then for any skeleton K of A the family $K \cap I$ also has these properties. Further, I is determined by $K \cap I$. (Proof omitted.) Note, though, that the definition of injections for a concrete category does not relativize to a skeleton in general. If we agree that in a skeleton of a concrete category (Q, B) the term *injection* is to be defined by reference to the whole category, then we have

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

2.4. *A category A with a distinguished subset I is coextensive with a concrete category under an isomorphism of skeletons identifying I with the class of injections, if and only if*

(1) *I contains all isomorphisms and is closed under composition with isomorphisms, and*

(2) *for f in I and g and h in A, $fg = fh$ implies $g = h$.*

We preface the proof with some remarks and a lemma. The construction is a modification of that of Eilenberg-MacLane **(2)**; like that one, it does not extend to proper classes. For the lemma, let us introduce the term *projection* in a somewhat unusual way. Relative to a prescribed class I of injections, f is a *projection* if the hypothesis $f = gh$, where g is an injection, implies g is an isomorphism.

2.5. *Under the conditions of* 2.4, *every isomorphism is a projection.*

Proof. It suffices to consider identities, for if k is an isomorphism having a factorization fg, f a proper injection, then the range of k is $f(gk^{-1})$. If the identity i is fg, f in I, we consider the mapping gf and its domain (= its range) j. We have $f(gf) = if = f = fj$; hence by condition (2), $gf = j$. Thus f must be an isomorphism, as was to be shown.

Proof of 2.4. The necessity of conditions (1) and (2) is clear. For the converse, choose a skeleton K of A, and let $J = I \cap K$. Let Q be the set of all identities of K. We must construct a concrete category (S, M) of spaces and mappings, having a skeleton K' involving a set Q' of spaces, with an isomorphism of K upon K' identifying J with the injections.

For each i in Q, let X_i be the set of all mappings in K with range i. Let $Q' = \{X_i | i \in Q\}$. For each f in K, define $f': X_{\delta(f)} \to X_{\rho(f)}$ by $f'(g) = fg$. Let $K' = \{f' | f \in K\}$. For each element f of J, f' is one-to-one, by condition (2); let the set $f'(X_{\delta(f)})$ be an element of S, and let S consist precisely of all these sets (f ranging over J) and all the elements of Q'. Let M consist of

(a) the functions in K',

(b) for each element f of J, the function f^* agreeing with f' on the domain $X_{\delta(f)}$, with range (cut down to) $f'(X_{\delta(f)})$, the function $(f^*)^{-1}$, and the inclusion function i: $f'(X_{\delta(f)}) \to X_{\rho(f)}$; and finally

(c) finite compositions of functions in (a) and (b).

Let us call the functions under (b) *b-mappings* (b for basic).

(S, M) satisfies the axioms **0**, **1**, **2**, for a concrete category, obviously. (The identities on spaces not in Q' are compositions $f^*(f^*)^{-1}$.) Next, if f is in J but not an isomorphism then the function f' is not onto; in fact, the identity $\rho(f)$ is not in the range of f^*, by 2.5. This shows that each b-mapping either is in K' or has a domain or range not in Q'. By induction, every mapping $g \in M$ whose domain and range are in Q' is an element of K'. Clearly each space in S is isomorphic with at least one space in Q'; K includes no two isomorphic identities; if f is not an identity in K then $f\delta(f) \neq \delta(f)$, and therefore

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

K' includes no isomorphism between two different spaces. Therefore K' is a skeleton of M.

Finally, it is clear that K and K' are isomorphic and every element f of J corresponds to an injection $f' = if^*$. An induction shows that every inclusion function in M is a b-mapping; and then every injection gih in K' corresponds to an element of J. This completes the proof of 2.4.

III. Bicategories. In a category A, a mapping f is *left cancellable* if $fg = fh$ implies $g = h$, *right cancellable* if $gf = hf$ implies $g = h$. A *bicategory* $\mathscr{C}$ is an ordered triple (A, I, P), where A is a category and I and P are subsets of A whose members are called *injections* and *projections*, respectively, and

3. Both I and P are subcategories containing all the isomorphisms.

4. Every mapping f in A is a composition gh of a projection h and an injection g. This decomposition, or factorization, is essentially unique; that is, the only other such expressions of f are of the form $(gj^{-1})(jh)$, where j is an isomorphism.

5. (a) Every injection is left cancellable. (b) Every projection is right cancellable.

The axioms are stated for an abstract category, but clearly

3.1. *Every bicategory (A, I, P) is coextensive with a concrete category under a correspondence representing I as the set of injections.*

For the axioms imply the hypothesis of 2.4. As for P, **(3**, Lemma 2.0**)** states

2.0. *In any bicategory, the mapping f is a projection if and only if $f = gh$, with g an injection, implies g is an isomorphism.*

One can also derive some of the results of **(5)** from these axioms. In particular, if $f = gh$ with f and g in I then h is in I, because of the uniqueness clause in Axiom 4. Thus we have the first four of the six properties of inclusion functions listed at the beginning of the previous section. Since the axioms **3—5** are preserved under passage to a skeleton K of A (replacing I with $I \cap K$, P with $P \cap K$), but the fifth and sixth properties in the cited list are not so preserved, therefore the axioms imply all those properties of inclusion functions which are coextensive invariants. I do not know if this remark can be made precise. (It cannot be done by constructing a coextensive concrete bicategory with I = inclusions, because there could be no non-identical isomorphisms.) But perhaps it conveys the idea.

What sort of category A can be made into a bicategory (A, I, P)? To begin with, it suffices if A is a concrete category in which for every mapping $f: X \to Y$ the set $f(X)$ is a space; the definitions are obvious and the proof is omitted. Any category coextensive with a bicategory is, in a natural way, a bicategory. Let us consider a more restrictive condition.

6. Every mapping which is left and right cancellable is an isomorphism.

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

3.2. *A category satisfying Axiom* **6** *can be made into a bicategory in at most one way.*

Proof. Suppose A satisfies **6**, and (A, I, P) and (A, I', P') are two bicategories. For any f in I, consider the factorization $f = gh$, g in I', h in P'. Since f is left cancellable, so is h (compute); since h is in P', it is also right cancellable and hence an isomorphism. Thus I is a subset of I'; by the same argument, I' is a subset of I, and by 2.0, $P = P'$ as well.

The one way is, of course, with I = all left cancellable mappings, P = all right cancellable mappings. The conditions for this to make a bicategory are Axioms **3**, **4**, and **5**; but **3** and **5** are trivial here. Axiom **4**, in this case, implies **6**; hence we may reduce the axioms to

3.3. The necessary and sufficient conditions on a category A in order that (A, I, P) form a bicategory, where I consists of all left cancellable mappings and P of all right cancellable mappings, are (1) I and P generate A, (2) for i in I and p in P such that pi exists, there are i' in I and p' in P such that $i'p' = pi$, and (3) for i and i' in I, if i' is not ij for any isomorphism j, then there do not exist p and p' in P such that $ip = i'p'$. If I' and P' are subcategories of I and of P, respectively, each containing all isomorphisms, then conditions (1)—(3) applied to I' and P' are necessary and sufficient in order that (A, I', P') be a bicategory.

The proof is omitted.

Axiom **6** may be regarded as a form of the First Isomorphism Theorem. It holds in many interesting categories; for example, in any *exact* category in the sense of **(1)**, and in any equationally definable class of algebras with zero, in the sense of **(4)**, where of course the mappings are the homomorphisms. In each case the proof is a routine exercise in the relevant theory. That the axiom is invalid for the most general types of algebras, and for some other types of systems, is illustrated by a rather trivial example. Consider all those algebras, (S, O), where S is a ground set and O a set of finitary operations on S, which is empty, and the (non-existent) operations in O are subjected to the one requirement $x = y$. There are precisely two algebras and three homomorphisms in any skeleton of this category, and Axiom **6** is clearly false. Thus if the axiom is to be satisfied one must exclude this sort of pathology. However, a category may contain this example and still satisfy Axiom **6**, as is shown by the compact Hausdorff spaces.

Another illustration is given by the category of all categories; precisely, the proper class $A \cup B$, where A is the class of all homomorphisms $f: X \to Y$, X and Y being categories which are sets, and B is the obvious subclass of $A \times A \times A$. In proving this, let us designate homomorphisms of categories by Latin letters, elements by Greek letters. Suppose $f: X \to Y$ is cancellable on both sides. Then f is one-to-one on identities; for if $f(\alpha) = f(\beta)$, α and β identities, then α and β form a two-element category Z which is mapped into X by the inclusion function $i: Z \to X$, and another homomorphism $j: Z \to X$ is defined

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

by $j(\alpha) = j(\beta) = \alpha$. Here $fi = fj$ but $i \neq j$, a contradiction. Therefore if $f(\alpha) = f(\beta)$ for any two elements α and β of X, we may conclude $\delta(\alpha) = \delta(\beta)$ and $\rho(\alpha) = \rho(\beta)$. If $\delta(\alpha) \neq \rho(\alpha)$ then there is a four-element category consisting of α, β, $\delta(\alpha)$ and $\rho(\alpha)$, which clearly has two different homomorphisms into X which have the same composition with f. There remains the case $\delta(\alpha) = \rho(\alpha) = \gamma$. Then α, β, and γ generate a semigroup Z with unit γ. Let W be the free semigroup with unit on two generators, σ, τ. Then W is a category, and there exist two homomorphisms h: $W \to Z$, k: $W \to Z$, determined by the conditions $h(1) = k(1) = \gamma$, $h(\sigma) = k(\tau) = \alpha$, $h(\tau) = k(\sigma) = \beta$. Composing h and k with the injection i: $Z \to X$, we obtain two category homomorphisms ih: $W \to X$ and ik: $W \to X$ such that $ih \neq ik$ but $fih = fik$. The contradiction establishes that f must be one-to-one. Then by 1.1, f is an isomorphism of X upon its image $Z \subset Y$. It remains to show that if Z is a proper subcategory of Y, then there exist a category U and two different homomorphisms of Y into U which coincide on Z. We omit the details of the argument, which turns on constructing a free sum of two copies of Y modulo the identification of the two copies of Z.

Thus the neat structure described by Axioms **0—6** is not uncommon. We do not have it, however, in non-compact topological spaces. As noted in **(3)**, what one typically finds in this example (say, all continuous mappings between Hausdorff spaces) is that the category A can be made a bicategory in two ways; once with all left cancellable mappings taken for injections, and again with every right cancellable mapping a projection. The common part, the two-sided cancellable mappings, consists of those one-to-one continuous functions whose image is a dense subspace of the range. The smaller classes of injections and of projections are then respectively the injections (in the ordinary sense) of closed subspaces, and the identification or quotient mappings.

We have avoided the term "quotient." The difficulty is in distinguishing between quotient and image. Now in groups, and in many other examples, the quotient and image in the usual sense are isomorphic; the distinction is a rather delicate one to make in an abstract setting, and the present bicategory axioms cannot do it. For work involving such distinctions one must use the original formulation of MacLane **(5**, see §11**)**. In topology, however, the quotient and image are typically quite different. They arise not in the factorization belonging to one bicategorical structure, but in two different ones. Note that a topological quotient mapping is categorically definable; $f: X \to Y$ is a quotient mapping if and only if the equation $f = gh$, with g left cancellable, implies g is an isomorphism. From each mapping h, of course, one can factor out the unique quotient mapping k such that $h = jk$ with j left cancellable. A similar, but more complicated, description of images can be given by reference to the one-point space.

Thus we have discriminated the two main uses of the terms. Clearly they conflict, and we cannot anticipate a revision either in topology or in algebra. We need a term for the blurred quotient-or-image given by projections accord-

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

ing to Axioms **0—5**. Let it be *quotient*; precisely, a quotient is an equivalence class of projections under the equivalence relation defined by $f \sim g$ when $f = ig$ for some isomorphism i. (This is perfectly analogous to the definition of a subspace.)

This choice frees the term "image," which happens to be wanted on several other counts. Some of these are (1) the use in the refined theory of bicategories, (2) the use, at least informally, for sets of values $f(X)$, (3) the use in connection with category homomorphisms (definition preceding 1.1), and (4) the following use. If $\phi = [f]$ is an equivalence class of injections into X, and $g: X \rightarrow Y$ a mapping, then the projection-injection factorization of gf yields a subspace of Y which is most naturally called the *image* of ϕ under g.

Now consider the propositions 1.3, on congruence relations, and 1.5, on identification categories. They are partially misleading, considered alone. But now we see that the trouble is that the congruence relations and identification categories have less to do with the categorical structure in this example than in either algebra or topology. If we replaced the concept of an identification category with the concept of a quotient, defined as an image under a mapping having no proper left cancellable left factor, then we should find 1.5 replaced by the proposition "Every homomorphic image is a quotient." Similarly the lattice isomorphism denied in 1.3 could be rediscovered by looking at the lattice of images instead of the lattice of congruence relations.

Next, the definition of a quotient in a bicategory (two paragraphs back) is more than merely analogous to the definition of a subspace; it is dual. The *dual* A^* of any category A is (a category) in one-to-one correspondence with A, $f \leftrightarrow f^*$, such that g^*f^* is defined if and only if fg is defined, and in that case $g^*f^* = (fg)^*$. It follows **(2)** that A^* is a category, $\delta(f^*) = \rho(f)^*$, and $\rho(f^*) = \delta(f)^*$. If (A, I, P) is a bicategory, then (A^*, I^*, P^*) is a bicategory **(3)**, where I^* is the image of P under $f \rightarrow f^*$, and P^* is the image of I. That is,

3.4. *Every bicategory has a dual, unique up to isomorphism, which is a bicategory.*

The proof is omitted.

We conclude with an important definition and a sketch of an embedding theorem. A subcategory $\mathscr{D}$ of a bicategory $\mathscr{C}$ is said to be *regular* if it is closed under factorization, i.e. if f is an injection in $\mathscr{C}$ and g a projection in $\mathscr{C}$, and fg is in $\mathscr{D}$, then f and g are in $\mathscr{D}$. A regular subcategory of a bicategory is of course a bicategory with the relativized sets of injections and projections. Every intersection of regular subcategories is regular, and therefore every subcategory (for that matter, every subset) is contained in a least regular subcategory.

3.5. Every concrete category which is a bicategory may be embedded as a regular subcategory of a bicategory satisfying Axiom **6**.

The embedding is an isomorphism; if the concreteness hypothesis is removed, one gets coextension from 2.4. The proof is too long to give here,

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press

mainly because of the first stage. In outline, the first stage is to enlarge the spaces suitably so that mappings which are not injections cease to be one-to-one. The third stage is to introduce a one-point space mapping into every space so that mappings which are not one-to-one cease to be left cancellable; one must precede this by a stage assuring that no existing one-point spaces are confused, which can be done by adding two zeros to each space. For the final stage, consider all pairs (X, Y), X a subspace of Y. In each case form a space Σ consisting of the sum of three copies of Y with the three copies of X identified. Two copies would be enough so that none of the old mappings, not a projection, remains right cancellable; to assure that $Y \to \Sigma$ (each of the three natural mappings) is not right cancellable, provide Σ with a group of six motions permuting the copies of Y. Only six mappings with domain Σ are admitted.

For the first stage, consider the general space X. Let $S(X)$ be the set of all ordered pairs (σ, τ), σ a subspace of X, i.e. an equivalence class of injections f: $Y \to X$, and τ an equivalence class of projections g: $Y \to Z$ under the relation $g \sim g'$ if $g' = agb$, a and b isomorphisms. The idea is that a mapping h: $X \to W$ which is not an injection has a right factor which is a proper projection; something which is surely narrowed by the mapping is the possibility of forming further projections. Thus we should like to transform quotients of X to quotients of W, which we could do directly (for projections h) if the quotients of a given domain formed a complete lattice. As it is, we must build a complete lattice. Accordingly call a subset T of $S(X)$ *residual* provided for each (σ, τ) in $T, f \in \sigma, g \in \tau$, T contains the equivalence classes of (1) all pairs (fi, g'), i an injection, g' the projection having the same domain as i arising in factorization of gi, and (2) all pairs (f, hg), h a projection. Replace X with the set X' consisting of the points of X and the residual subsets of $S(X)$. For any mapping h: $X \to W$, extend h over X' by taking for $h(T)$ the least residual set in $S(W)$ containing the equivalence classes of all (f', g') such that for some (f, g), $f \in \sigma$, $g \in \tau$, $(\sigma, \tau) \in T$, the following is true. The mapping hf has a factorization $f'k$, k a projection; i.e. σ' is the image of σ. And $g'k = g$, i.e. g' induces g. The empty set is a residual subset of $S(W)$ which may have to be used; however, the padded category is well defined and the straightforward verification of its properties may be omitted.

References

1. D. Buchsbaum, *Exact categories and duality*, Trans. Amer. Math. Soc., *80* (1955), 1–34.
2. S. Eilenberg and S. MacLane, *General theory of natural equivalences*, Trans. Amer. Math. Soc., *58* (1945), 231–294.
3. J. Isbell, *Algebras of uniformly continuous functions*, submitted to Annals of Math.
4. B. Jònsson and A. Tarski, *Direct Decompositions of Finite Algebraic Systems* (Notre Dame, 1947).
5. S. MacLane, *Duality for groups*, Bull. Amer. Math. Soc., *56* (1950), 485–516.

Institute for Advanced Study

https://doi.org/10.4153/CJM-1957-064-6 Published online by Cambridge University Press