

NATURALITY AND DEFINABILITY, I

WILFRID HODGES AND SAHARON SHELAH

Eilenberg and Mac Lane [1] explained the notion of a ‘natural’ embedding by giving a categorical definition. Starting from their examples, we argue that one could equally well explain natural as meaning ‘uniformly definable in set theory’. But do the categorically natural embeddings coincide with the uniformly definable ones? This is partly a technical question about whether certain well-known algebraic constructions are definable in set theory. It can also be seen as a test of the adequacy of ZFC as a foundation for algebra.

After laying out the groundwork in Section 1, we analyse some examples (such as divisible hulls of abelian groups) in Section 2, and in Section 3 we formalise the main problem. The first partial answer is in Section 4: naturality does imply uniform definability with parameters, provided only a set of isomorphism types are involved. The proof also partially answers a model-theoretic question of Gaifman about definable operations. In Section 5 we prove a partial converse: uniform definability can imply naturality if urelements are allowed. Section 6 puts an upper bound on possible strengthenings of this result by showing that if uniform definability implies definable naturality in a model M of set theory, then M has a global choice function. In part II of this paper we hope to give an example of an unnatural but uniformly definable construction.

The results of Section 6 are due to the second author. The first author takes responsibility for the rest; he thanks the Royal Society for a grant which took him to Jerusalem for three weeks in 1981, where the results in the paper were assembled in discussion between the authors.

1. *Naturality*

Let L be a finite-dimensional vector space over the real numbers and let $T(L)$ be its dual space. We quote:

- (1) ‘For the iterated conjugate space $T(T(L))$. . . it is well known that one can exhibit an isomorphism between L and $T(T(L))$ *without* using any special basis in L . This exhibition of the isomorphism $L \cong T(T(L))$ is “natural” in that it is given *simultaneously* for *all* finite-dimensional vector spaces L . This simultaneity can be further analyzed. . . .’

This passage occurs in the opening paragraphs of ‘General theory of natural equivalences’ by S. Eilenberg and S. Mac Lane [1], the paper which founded category theory.

In the ‘further analysis’ that they refer to, Eilenberg and Mac Lane go on to show that T can be defined on linear maps as well as on vector spaces, so that it forms a functor. Then they point out that the family h of natural isomorphism $h_L: L \rightarrow T(T(L))$

Received 8 February 1985.

1980 *Mathematics Subject Classification* 08A35.

J. London Math. Soc. (2) 33 (1986) 1–12

is a natural isomorphism from the identity functor to T ; in other words, all the following diagrams commute.

$$(2) \quad \begin{array}{ccc} L & \xrightarrow{h_L} & T^2(L) \\ s \downarrow & & \downarrow T^2(s) \\ M & \xrightarrow{h_M} & T^2(M) \end{array}$$

They describe the commutativity of (2) as the *naturality condition* for h .

Now any set theorist, asked to formalise (1), would come up with something quite different from what Eilenberg and Mac Lane proposed. In the first place, to ‘exhibit an isomorphism between L and $T(T(L))$ ’, if it means anything, means to *define* such an isomorphism. Next, to say that it is given ‘simultaneously for all L ’ is to say that the same definition works uniformly for all L . Finally, to say that the isomorphism is exhibited ‘without using any special basis in L ’ means something like: the definition refers only to L and not to any other parameters depending on L . So (1) above already has a perfectly reasonable formalisation in terms of set-theoretic definability, as follows:

- (3) There is a formula $\phi(x, y)$ of set theory such that for every finite-dimensional vector space L over the reals there is a unique h_L such that $\phi(L, h_L)$ is true; for each such L this h_L is an isomorphism from L to $T(T(L))$.

True, there are models M of set theory in which (3) holds for irrelevant reasons. If M has a definable well-ordering $<$ of the universe, then $\phi(x, y)$ need only say that y is the $<$ -first isomorphism from x to $T^2(x)$. In M there are even uniformly definable isomorphisms i_L from finite-dimensional vector spaces L to their first duals $T(L)$. These maps i_L are certainly not natural. But it is clear how we can remove this irrelevance: we should require that $\phi(x, y)$ satisfies (3) in *all* models (or perhaps all transitive models) of ZFC.

So now there is a mathematical question whether (3) (made precise in a suitable way) is equivalent to the commutativity of (2). Putting it in more general terms: are the natural constructions precisely the definable ones? From a set-theoretic point of view the more interesting half of this question is whether definability implies naturality. This for two reasons: first, if definability *does* imply naturality, then a number of well-known algebraic constructions cannot be defined in set theory, because they are known not to be natural. Some examples are given in the next section. One would expect these constructions not to be definable, but we do not know how to prove it, and the problem seems hard. Secondly, if definability in ZFC *does not* imply naturality, this will have foundational implications: it can be understood as saying that ZFC imposes an internal structure on mathematical objects which clashes with the proper generality of algebra.

Eilenberg’s and Mac Lane’s example of vector spaces is atypical in several ways. For a start one should drop the restriction to finite dimension; then h_L becomes a natural embedding instead of an isomorphism. But before presenting a general analysis we turn to look at some rather different examples.

2. Examples

In all the examples of this section, a class K of algebraic objects is given, and for each algebra A in K we construct an algebra $F(A)$ which contains A as a part; $F(A)$ is specified up to isomorphism over A . The embedding map $h_A: A \rightarrow F(A)$ is just inclusion, so it is always definable at once from A and $F(A)$. The rub is that $F(A)$ may not be uniformly definable from A , even though it is definable up to isomorphism.

Following Eilenberg and Mac Lane, we shall say that F is *natural on automorphisms*, or for brevity just *natural*, if and only if for every A in K and every automorphism α of A there is an automorphism $f(\alpha)$ of $F(A)$ such that (4) and (5) hold:

(4) the diagram

$$\begin{array}{ccc} A & \hookrightarrow & F(A) \\ \alpha \downarrow & & \downarrow f(\alpha) \\ A & \hookrightarrow & F(A) \end{array}$$

commutes;

(5) f is functorial, that is, $f(1_A) = 1_{F(A)}$ and $f(\alpha\beta) = f(\alpha) \cdot f(\beta)$ for all automorphisms α, β .

We count the embedding of A in $F(A)$ as part of the structure of $F(A)$. Thus every automorphism β of $F(A)$ restricts to an automorphism $\rho_A(\beta)$ of A . Writing Aut for automorphism group, this defines a homomorphism $\rho_A: \text{Aut}(F(A)) \rightarrow \text{Aut}(A)$.

LEMMA 1. For F as above the following are equivalent:

(a) F is natural,

(b) for every $A \in K$ there is a group embedding $f_A: \text{Aut}(A) \rightarrow \text{Aut}(F(A))$ such that $\rho_A f_A = 1$.

Proof. If F is natural then $f: \text{Aut}(A) \rightarrow \text{Aut}(F(A))$ as in (4) is a group homomorphism by (5), and for each $\alpha \in \text{Aut}(A)$, $\rho_A f(\alpha) = \alpha$ by (4). Hence $\rho_A f = 1$, and this implies that f is an embedding. The converse is immediate.

Here follow some examples.

EXAMPLE 1. *Divisible hulls of abelian groups.* Every abelian group A can be extended to a divisible hull $F(A)$ which is unique up to isomorphism over A . If A is torsion-free then $F(A)$ can be constructed from A essentially as one constructs the rational numbers from the integers, and so there is a uniform and explicit set-theoretic definition of $F(A)$ in terms of A . Also when A is torsion-free, each automorphism of A extends uniquely to an automorphism of $F(A)$, and it follows that F is natural on torsion-free groups. Thus far, naturality and definability go hand in hand.

However, F is not natural on all abelian groups. Let us prove this. We need some preliminaries on linear groups. Let p be a prime and n a positive integer. Now $\text{GL}(2, p^n)$ is the general linear group of invertible 2×2 matrices over the ring $Z/(p^n)$. The natural homomorphism $h: Z/(p^{n+1}) \rightarrow Z/(p^n)$ induces a map $\phi: \text{GL}(2, p^{n+1}) \rightarrow \text{GL}(2, p^n)$. Since h preserves the ring operations, ϕ is a group homomorphism; its kernel consists of those matrices which equal $I \pmod{p^n}$. Thus we have a short exact sequence

$$I + p^n \text{GL}(2, p^{n+1}) \longrightarrow \text{GL}(2, p^{n+1}) \xrightarrow{\phi} \text{GL}(2, p^n).$$

The following lemma tells us that this sequence need not split.

LEMMA 2. *There is a matrix N in $\text{GL}(2, 5)$ of order 5, such that every pre-image of N under $\phi: \text{GL}(2, 25) \rightarrow \text{GL}(2, 5)$ has order 25.*

Proof. Take

$$N = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

The pre-images of N under ϕ are the matrices M in $\text{GL}(2, 25)$ of the form

$$M = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} + 5 \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad a, b, c, d \in \{0, \dots, 4\}.$$

A straightforward but tedious calculation shows that whatever the values of a, b, c, d , we have

$$M^5 = \begin{pmatrix} 1 & 0 \\ 5 & 1 \end{pmatrix}$$

which has order 5, so that M has order 25.

To show that the divisible hull construction is not natural, we apply Lemma 2 as follows. The abelian group $A = Z_5 \oplus Z_5$ has divisible hull $Z(5^\infty) \oplus Z(5^\infty)$. Any automorphism of this hull restricts to an automorphism of the characteristic subgroup $B = Z_{25} \oplus Z_{25}$, so it suffices to show that some automorphism of A cannot be extended to an automorphism of B of the same order. Let c, d be generators of B , so that $5c, 5d$ generate A . Let $M \in \text{GL}(2, 25)$ be the matrix of an automorphism α of B , using c, d as basis. Then α induces an automorphism of A with matrix $\phi(M)$ with respect to the basis $5c, 5d$. Since $5c, 5d$ have order 5, $\phi(M)$ will be the matrix ‘ $M \bmod 5$ ’ in $\text{GL}(2, 5)$, so that ϕ is precisely the homomorphism of Lemma 2. Then the automorphism of A with matrix N as in the lemma has order 5, but every automorphism of B which extends it has order at least 25, and so the same holds for the divisible hull of A . It follows that there is no group embedding $f: \text{Aut}(A) \rightarrow \text{Aut}(F(A))$ satisfying Lemma 1(b), and so by that lemma F is not natural.

Is the divisible hull construction F set-theoretically definable? It would be amazing if it was. But virtually all we know is that if $\phi(x, y)$ is a Σ_1 formula which (provably in ZF) defines a function, then there is a model of ZFC in which $\phi(x, y)$ fails to define divisible hulls (cf. [6], where a slightly stronger result is proved).

EXAMPLE 2. *Algebraic closures of fields.* There is no natural construction for these. By a theorem of Artin and Schreier [7, p. 316] if α is a non-trivial automorphism of an algebraically closed field and α has finite order, then α has order 2 and switches the two square roots of -1 . Hence if A is a field with an automorphism of order 3 (for example a pure transcendental extension of the prime field, with transcendence basis of cardinality 3), then A has an automorphism which cannot be extended to an automorphism of its algebraic closure with the same order. Now apply Lemma 1 as in Example 1.

As for set-theoretic definability of algebraic closures, we know exactly the same as we know for divisible hulls of abelian groups (cf. [5]). Again it would be astonishing if there was a set-theoretic formula which uniformly defined them.

EXAMPLE 3. *Left adjoints.* If G is a forgetful functor on structures and F is left adjoint to G , then F defines a natural construction as follows. Let $h: 1 \rightarrow GF$ be the unit of the adjunction. Then for each structure A in the domain of F , h_A maps A homomorphically to $GF(A)$. Thus from A we construct a two-part structure $(A, F(A))$ with a map h_A from the elements of A to the elements of $F(A)$. Condition (5) in the definition of naturality comes at once from the fact that h is a natural transformation. This recipe includes most of the natural constructions that one first thinks of, such as the free group on the set A of generators, or the abelianisation $A/[A, A]$ of the group A (considered as a two-part structure $(A, A/[A, A])$ with the so-called ‘natural homomorphism’ connecting them). Another example is the construction which takes each set A to its power-set boolean algebra $\mathcal{P}(A)$ together with the obvious map from A to the set of atoms of $\mathcal{P}(A)$. Divisible hulls of torsion-free groups are an example too.

Every construction of this sort is set-theoretically definable from a parameter. This is because it is a pointwise right Kan extension along a small functor (cf. Mac Lane [8]; Hodges [4] works out the details).

EXAMPLE 4. *Projective covers of R -modules for a perfect ring R .* Projective covers are the dual of injective hulls. They need not exist; the condition ‘ R is perfect’ means simply that they do (cf. [2, Vol. II, Chapter 22]).

In general there is no natural projective cover construction. For let R be the perfect commutative ring $Z/(25)$. Then if A and B are as in the discussion after Lemma 2 above, A is an R -module whose projective cover is the surjective homomorphism

$$(6) \quad B \xrightarrow{g} A,$$

where g operates by killing the elements of order 5. As in Example 1 we can represent automorphisms of B by matrices in $GL(2, 25)$, using a basis c, d of B . Any automorphism of B with a matrix M induces an automorphism of A whose matrix with respect to the basis gc, gd is $\phi(M)$ as defined for Lemma 2. So Lemma 2 again shows that some automorphism of A of order 5 cannot be extended to an automorphism of order 5 of the whole diagram (6), and it follows by Lemma 1 again that projective covers are not natural.

3. What is a construction?

The following definitions are relative to a model M of set theory; M can be the real world of sets, V .

Let K be a class of structures. To avoid coding-in irrelevant information, we always assume that K is closed under isomorphism. A *representing formula* is a formula $\phi(x, y)$ of set theory, possibly with parameters from M , such that

- (7) for each structure A in K there is a structure $F(A)$ containing A as a distinguished part, such that $M \models \phi(A, F(A))$;
- (8) for each structure A in K and each element B of M , $M \models \phi(A, B)$ if and only if B is a structure containing A and isomorphic to $F(A)$ over A ;
- (9) if A, C are isomorphic structures in K and $M \models \phi(A, B) \wedge \phi(C, D)$ then B is isomorphic to D .

Thus a representing formula ϕ defines a construction F on K , but only up to isomorphism. We say that ϕ *represents* F . In algebraic contexts we can speak of ‘the structure $F(A)$ ’ when only the isomorphism type of $F(A)$ over A matters.

We could have strengthened (9) to require that any isomorphism from A to C extends to an isomorphism from B to D . But nothing below needs this stronger assumption.

Let F be a construction on K , represented by the formula ϕ . We say that the formula $\psi(x, y)$ *defines* F if and only if for every structure A in K ,

$$(10) \quad M \models (\exists y \phi(A, y) \rightarrow \exists_{=1} y \psi(A, y)) \wedge \forall y (\psi(A, y) \rightarrow \phi(A, y)).$$

In other words, ψ uniformises ϕ . We say that F is *definable* if and only if some formula defines it.

By (7) every automorphism α of $F(A)$ restricts to an automorphism $\rho_A(\alpha)$ of A . In line with Lemma 1 above, we say that F is *natural* if and only if for every structure A in K there is a group embedding $f_A: \text{Aut}(A) \rightarrow \text{Aut}(F(A))$ in M such that $\rho_A f_A = 1$.

Eilenberg’s and Mac Lane’s second duals of vector spaces are actually rather hard to fit into this framework. One way of encasing the algebraic definition of ‘dual space’ is to make it a two-part structure $(A, T(A))$ with relations R_λ for the scalars λ , such that if $a \in A$ and $\gamma \in T(A)$ then $R_\lambda(a, \gamma)$ holds if and only if $\gamma(a) = \lambda$. Then the second dual will be a three-part structure $(A, T(A), T^2(A))$ with suitable relations, and the natural embedding h_A will run from A to $T^2(A)$ in this structure. The complication is that $T(A)$ is needed for the definition of h_A but embeddings from $\text{Aut}(A)$ to $\text{Aut}(T(A))$ are irrelevant to the naturality of h_A . The definitions of representing formula etc. could be adjusted to deal with this point, but no matter of principle is involved and it seems hardly worth the effort.

4. Natural implies definable?

THEOREM 3. *Let M be a model of set theory (ZFC, or possibly with urelements). Suppose some formula represents the construction F on the class K in M . If F is natural and there is only a set of isomorphism types of structures in K , then F is definable in M with parameters.*

Proof. We begin with the case where all structures in K are isomorphic. First choose a structure A in K and a corresponding structure $F(A)$. Choose an embedding $f: \text{Aut}(A) \rightarrow \text{Aut}(F(A))$ in M as in Lemma 1. The parameters in the definition of F will be $A, F(A), f$ and any parameters in the formula representing F .

If B is any structure in K and $\beta: A \rightarrow B$ an isomorphism, we build a structure $F(B, \beta)$ in M as follows. The elements of $F(B, \beta)$ are the ordered pairs $\langle d, \beta \rangle$ such that d is an element of $F(A)$, and relations are defined so that the map $d \mapsto \langle d, \beta \rangle$ is an isomorphism from $F(A)$ to $F(B, \beta)$.

Unfortunately there is generally no canonical choice of β . So we have to choose an average of the various $F(B, \beta)$ as β ranges over the set $\text{Iso}(A, B)$ of isomorphisms from A to B in M . Suppose that $\beta, \gamma \in \text{Iso}(A, B)$. Then $\gamma^{-1}\beta \in \text{Aut}(A)$, and so $f(\gamma^{-1}\beta)$ is an automorphism of $F(A)$. Define a relation \sim by

$$\langle d, \beta \rangle \sim \langle e, \gamma \rangle \quad \text{if and only if} \quad f(\gamma^{-1}\beta)(d) = e.$$

We claim that \sim is an equivalence relation.

First $f(\beta^{-1}\beta) = 1_{F(A)}$, and so \sim is reflexive. Next, since f is a homomorphism, $f(\gamma^{-1}\beta)^{-1} = f(\beta^{-1}\gamma)$ and hence \sim is symmetric. For transitivity, suppose that $\langle c, \alpha \rangle \sim \langle d, \beta \rangle \sim \langle e, \gamma \rangle$. Then $f(\beta^{-1}\alpha)(c) = d$, $f(\gamma^{-1}\beta)(d) = e$ and hence $f(\gamma^{-1}\alpha)(c) = f(\gamma^{-1}\beta)f(\beta^{-1}\alpha)(c) = e$, whence $\langle c, \alpha \rangle \sim \langle e, \gamma \rangle$ as required.

Now suppose that $\beta, \gamma \in \text{Iso}(A, B)$. The relation \sim sets up a bijection between $F(B, \beta)$ and $F(B, \gamma)$, and it can be checked that this bijection is an isomorphism. So we can define the elements of $F(B)$ to be the equivalence classes of \sim , and a unique structure is induced on this set to make it isomorphic to $F(A)$. We define $i: \text{dom } B \rightarrow \text{dom } F(B)$ by $i(\beta d) = \langle d, \beta \rangle \sim$. This map is well defined. For suppose that $\beta d = \gamma e$. Then $e = \gamma^{-1}\beta d = f(\gamma^{-1}\beta)d$ and so $\langle d, \beta \rangle \sim \langle e, \gamma \rangle$. Hence we can identify B with $i(B)$ inside $F(B)$. This gives an explicit definition of $F(B)$ from B and the stated parameters.

Finally when K has a set of isomorphism classes, we use the axiom of choice in M to choose a family A_i ($i \in I$) of representatives of these classes, together with appropriate $F(A_i)$ and f_i , and then proceed as before on each class separately. The choice set will be a parameter in the definition of F .

The proof of Theorem 3 shows more than we stated. We can define F in M so that for each isomorphism $\gamma: B \rightarrow C$ between structures in K , an isomorphism $F(\gamma): F(B) \rightarrow F(C)$ extending γ is also given; moreover this defined map F is functorial in the sense that $F(\beta\gamma) = F(\beta) \cdot F(\gamma)$ and $F(1_B) = 1_{F(B)}$ for all structures B in K and isomorphisms β, γ between them. If A is as in the proof and $\beta: A \rightarrow B$ is an isomorphism, we define $F(\beta)(d)$ to be $\langle d, \beta \rangle \sim$ for each element d of $F(A)$. Then, given any isomorphism $\gamma: B \rightarrow C$, we define $F(\gamma)$ to be $F(\gamma\beta)F(\beta)^{-1}$; this is independent of the choice of $\beta: A \rightarrow B$. In short F is definably natural on isomorphisms as well as automorphisms.

PROBLEM A. Can the restriction that K contains only a set of isomorphism types be removed in Theorem 3?

REMARK. For any model M of ZFC, the answer to Problem A is Yes if and only if the following holds. If X is a definable class in M , and every element of X is a non-empty set of pairwise isomorphic structures, then there is a definable function G in M such that for each $x \in X$, $G(x)$ is a structure isomorphic to the elements of x .

PROBLEM B. Under the hypotheses of Theorem 3, if F has representing formula ϕ , is it always possible to define F by a formula ψ whose parameters are those in ϕ and those needed to define K ?

REMARK. For any model M of ZFC, the answer to Problem B is Yes if and only if the following holds. If X is a set each of whose elements is a set of isomorphic structures, and X is definable by a formula with parameter c , then there is a formula with parameter c which defines a set Y of structures containing, for each $x \in X$, just one structure isomorphic to those in x .

Haim Gaifman raised the following question in several unpublished preprints. Let T be a complete theory in a first-order language L_2 with a 1-ary relation symbol P , and let L_1 be a language $\subseteq L_2 \setminus \{P\}$, such that every model B of T is determined by

the L_1 -reduct of the substructure defined by P , $B^P \upharpoonright L_1$, up to isomorphism over $B^P \upharpoonright L_1$. We call such a theory T an *operation* (it operates on $B^P \upharpoonright L_1$ to produce B). When does there exist a theory T^+ , in a language got by adding to L_2 a finite number of new function symbols g_i , so that (1) every model B of T expands to a model B^+ of T^+ , (2) every element of B^+ is of form $g_i(\bar{a})$ for some i and some tuple \bar{a} in $B^P \upharpoonright L_1$, and (3) T^+ states that the formulas satisfied by any sequence $(\bar{a}_0, g_{i_1}(\bar{a}_1), \dots, g_{i_n}(\bar{a}_n))$ in B^+ are determined by the formulas satisfied by $(\bar{a}_0, \bar{a}_1, \dots, \bar{a}_n)$ in $B^P \upharpoonright L_1$?

The method of Theorem 3 answers this question when the models of T are finite.

THEOREM 4. *Let T be an operation such that for each model B of T , $B^P \upharpoonright L_1$ is finite. Then T^+ exists as above if and only if for some (or, equivalently, every) model B of T there is a group embedding $f: \text{Aut}(B^P \upharpoonright L_1) \rightarrow \text{Aut}(B)$ such that each $\alpha \in \text{Aut}(B^P \upharpoonright L_1)$ is the restriction to $B^P \upharpoonright L_1$ of $f(\alpha)$.*

Proof. If T^+ exists, then define f by putting $f(\alpha)(g_i(\bar{a})) = g_i(\alpha\bar{a})$. The finiteness condition is not needed for this.

For the converse, since T is complete and the structures $B^P \upharpoonright L_1$ are finite, they are all isomorphic to some fixed L_1 -structure A . List the elements of A as a fixed n -tuple \bar{a} . Choose a model D of T such that $D^P \upharpoonright L_1 = A$. Since T is an operation, D is also finite. All models of T are isomorphic to D . We define T^+ as the theory of an expansion D^+ of D , as follows.

For each element d of D , introduce a new n -ary function symbol g_d . If $\beta \in \text{Aut}(A)$, put $g_d(\beta\bar{a}) = f(\beta)d$. If \bar{b} is an n -tuple in A which is not of the form $\beta\bar{a}$ for any $\beta \in \text{Aut}(A)$, put $g_d(\bar{b}) = b_0$. This defines D^+ , and (1), (2) are automatically satisfied.

For (3) it suffices to check that if

$$(11) \quad (A, \bar{a}_0, \bar{a}_1, \dots, \bar{a}_m) \equiv (A, \bar{b}_0, \bar{b}_1, \dots, \bar{b}_m)$$

then, for all ψ, d_1, \dots, d_m ,

$$(12) \quad D^+ \models \psi(\bar{a}_0, g_{d_1}(\bar{a}_1), \dots, g_{d_m}(\bar{a}_m)) \leftrightarrow \psi(\bar{b}_0, g_{d_1}(\bar{b}_1), \dots, g_{d_m}(\bar{b}_m)).$$

By (11) and some rearrangement, we can assume that \bar{a}_0 is \bar{a} and \bar{b}_0 is $\beta\bar{a}$ for some $\beta \in \text{Aut}(A)$. By (11) again, if some \bar{a}_i is not of the form $\gamma\bar{a}$ for any $\gamma \in \text{Aut}(A)$, then neither is \bar{b}_i , and so we can eliminate g_{d_i} in (12). Finally by (11) again, there are $\alpha_1, \dots, \alpha_m \in \text{Aut}(A)$ such that $(\bar{a}_0, \dots, \bar{a}_m)$ is $(\bar{a}, \alpha_1\bar{a}, \dots, \alpha_m\bar{a})$ and $(\bar{b}_0, \dots, \bar{b}_m)$ is $(\beta\bar{a}, \beta\alpha_1\bar{a}, \dots, \beta\alpha_m\bar{a})$. Now we use the fact that f is a group homomorphism:

$$(13) \quad g_{d_i}(\bar{b}_i) = g_{d_i}(\beta\alpha_i\bar{a}) = f(\beta\alpha_i)d_i = f(\beta) \cdot f(\alpha_i)d_i = f(\beta)g_{f(\alpha_i)d_i}(\bar{a})$$

and similarly $g_{d_i}(\bar{a}_i) = g_{f(\alpha_i)d_i}(\bar{a})$. Next we use the fact that $f(\beta)$ extends β , to see that the right-hand formula of (12) holds in D^+ if and only if

$$(14) \quad D^+ \models \psi(f(\beta)\bar{a}, f(\beta)g_{f(\alpha_1)d_1}(\bar{a}), \dots, f(\beta)g_{f(\alpha_m)d_m}(\bar{a})).$$

Applying $f(\beta)^{-1}$, we deduce (12).

Theorem 4 is unlikely to be true without the finiteness condition, but no counterexamples are known.

5. *Definable implies natural?*

For the next theorem we shall work in Fraenkel–Mostowski set theory with a proper class of atoms, FMP for short. Atoms are elements which are not sets. The axioms of FMP are the same as those of ZFC except that extensionality is restricted to sets, and a new function symbol $U_{(\cdot)}$ is added to the language, together with the following new axioms:

$$\begin{aligned} &(\forall \text{ ordinals } i) (U_i \text{ is a set and } |U_i| = |i|), \\ &\forall ij (i \neq j \rightarrow U_i \cap U_j = 0), \\ &\forall x (x \text{ is an atom} \rightarrow \exists i (x \in U_i)). \end{aligned}$$

It is routine to show that FMP is consistent relative to ZFC. In fact if ZFC has transitive models (as we shall assume) then so does FMP.

In any model V of FMP we put $U = \bigcup_i U_i$. We define the *transitive closure* of x , $TC(x)$, as the least transitive set containing x . We write V_i for the class of all elements x such that $TC(x) \cap U \subseteq \bigcup_{k < i} U_k$. Then $V = \bigcup_i V_i$.

THEOREM 5. *If V is a transitive model of FMP then any construction defined in V is natural.*

Proof. Let F be a construction defined in V by a formula $\psi(x, y, \bar{a})$, where \bar{a} are the parameters. Without loss we can assume that the domain K of F consists of all structures isomorphic to a certain fixed structure A . Using the axiom of choice we can launder K and ψ so that the symbols in the similarity type of K are simply an initial segment j_1 of the ordinals. Then every structure in K can be written as a pair $B = (\text{dom } B, g)$ where g is a map from j_1 to the set of all relations on $\text{dom } B$. Likewise we can assume that the symbols in the similarity type of $F(A)$ are an initial segment j_2 of the ordinals.

Consider $B = (\text{dom } B, g)$ in K , and let β be an automorphism of V such that $\beta(\text{dom } B) = \text{dom } B$. Then β fixes every ordinal, and hence it pointwise fixes the domain of g . It follows that $\beta(B) = B$ if and only if the restriction of β to $\text{dom } B$ is an automorphism of B . The same applies to structures isomorphic to $F(A)$.

Choose an ordinal $i \geq |\text{dom } A|$ such that the parameters \bar{a} lie in V_i . In V there is some structure B which is isomorphic to A and has domain $\subseteq U_i$. Since ψ defines F , there is a unique structure $C = F(B)$ such that $V \models \psi(B, C, \bar{a})$.

Each automorphism α of B extends to a permutation α^* of U so that $\alpha^*u = u$ for all $u \in U \setminus (\text{dom } B)$. Then α^* extends to an automorphism of V by writing $\alpha^*(x) = \{\alpha^*(y) : y \in x\}$ for every set $x \in V$. Thus $\text{Aut}(B)$ acts on V ; in fact the map $\alpha \mapsto \alpha^*$ is a group embedding from $\text{Aut}(B)$ to $\text{Aut}(V)$.

Let $\alpha \in \text{Aut}(B)$. Then $\alpha^*(B) = B$. If $x \in V_i$ then $\alpha^*x = x$, and in particular $\alpha^*\bar{a} = \bar{a}$. Since α^* is an automorphism of V , we have $V \models \psi(\alpha^*B, \alpha^*C, \alpha^*\bar{a})$, in other words $V \models \psi(B, \alpha^*C, \bar{a})$. But ψ defines F , and hence $\alpha^*C = C$. In particular α^* permutes the elements of $\text{dom } C$, and so by remarks above, $\alpha^* \upharpoonright C$ is an automorphism of C which extends the automorphism α of B . The map $\alpha \mapsto \alpha^* \upharpoonright C$ is the required embedding of $\text{Aut}(B)$ in $\text{Aut}(F(B))$.

Unfortunately Theorem 5 is a triviality. The significant question is the following.

PROBLEM C. Does ZFC have transitive models in which every definable construction is natural?

We have no idea how to answer this. The nearest approach so far is a result of Harvey Friedman [3]. He shows that ZFC has models in which every definable construction F has the following property, which we may call *semi-naturality*. For all A, B in the domain of F and every isomorphism $\alpha: A \rightarrow B$ there is an isomorphism from $F(A)$ to $F(B)$ extending α . This is equivalent to the property that if A is in the domain of F and α is an automorphism of A then α extends to an automorphism of $F(A)$.

Despite appearances, semi-naturality is a long way short of naturality. Most reasonable constructions are semi-natural anyway, including all the examples in Section 2 above.

6. *Definable implies definably semi-natural?*

Though we have no information on Problem C, we can at least show that one plausible improvement of it is false.

Let M be a model of set theory and F a construction defined in M by a formula ψ . Then for every structure A in the domain of F there is a unique B such that $M \models \psi(A, B)$; we write F_ψ for the function $A \mapsto B$. We say that F_ψ is *definably semi-natural* if and only if there is some formula $\chi(x, y)$ such that for each A , $\chi(A, -)$ defines in M a map $f: \text{Aut}(A) \rightarrow \text{Aut}(f_\psi(A))$ (not necessarily a group embedding) such that each $f(\alpha)$ extends α . We say that F is *definably semi-natural* if and only if there is some formula ψ defining F , such that F_ψ is definably semi-natural.

THEOREM 6. *There is a formula $\psi(x, y)$ of set theory such that if M is any model of ZFC then*

- (a) ψ defines a natural construction F in M , and
- (b) if F_ψ is definably semi-natural then M has a definable global choice function.

Proof. Let K be the class of all structures A of the following form: A has an equivalence relation E_A with two classes, these classes are of equal size, and A has no other functions or relations. Then $F(A)$ is formed by adding to each equivalence class of E_A the same number of elements as it already contains, and then supplying a new 1-ary relation symbol P to pick out the old elements. Clearly we can write down a formula $\psi(x, y)$ which defines such a construction F in any model M of ZFC, and F will be natural. This gives (a).

To get (b) we specify ψ more precisely. It should say that if the equivalence classes of A happen to be the sets $\{a\} \cup |a|$ and $\{|a\} \cup \{1\} \times |a|$, for some infinite set a , then the equivalence classes of $F_\psi(A)$ are $\{a\} \cup |a| \cup a$ and $\{|a\} \cup \{1, 2\} \times |a|$. Now suppose that F_ψ is definably semi-natural. Then there is a definable function G which takes each automorphism β of A to an automorphism $G(\beta)$ of $F_\psi(A)$ extending β . Now given any infinite set a , let A be as above, and let β_a be the automorphism of A which switches the equivalence classes, taking a to $|a|$ and each $i \in |a|$ to $\langle 1, i \rangle$. Then $G(\beta_a)$ switches the equivalence classes of $F_\psi(A)$, and hence it maps a bijectively to $\{2\} \times |a|$. Since $\{2\} \times |a|$ is well ordered, $G(\beta_a)$ defines a well-ordering of a . But $G(\beta_a)$ is uniformly definable from a . Thus (b) holds.

COROLLARY 7. *There is no model of ZFC in which, whenever a formula ψ defines a construction F , there is a definable map taking each A in the domain of F to a group embedding $f_A: \text{Aut}(A) \rightarrow \text{Aut}(F_\psi(A))$ as in Lemma 1.*

Proof. Suppose to the contrary that M is such a model, and consider ψ from the theorem. Since F_ψ is definably semi-natural by supposition, M has a definable global choice function by the theorem. But then we can use this choice function to define constructions which are not natural.

Can Corollary 7 be improved to say that there is no model of ZFC in which every definable construction is ‘definably natural’? This seem likely. Theorem 8 below improves the relevant part of Theorem 6 from ‘ F_ψ definably semi-natural’ to ‘ F definably semi-natural’; but the construction used is not natural, so that we had to assume a weak global choice in order to make it definable.

THEOREM 8. *Let M be a model of ZFC with a definable global choice function for finite sets. Then in M there is a definable construction F , which is definably semi-natural if and only if M has a definable global choice function.*

Proof. In M , let G be an abelian group and $\pi: G \rightarrow Z_2$ a surjective homomorphism. We define $A(G, \pi)$ to be a structure with three parts: $A(G, \pi) = (\{a, b\}, G, Z_2)$, where a, b are two distinct elements; besides the operations on G and Z_2 , $A(G, \pi)$ carries a function $\sigma_{A(G, \pi)}: G \rightarrow Z_2$ which coincides with π . We define K to be the class of all structures isomorphic to structures of form $A(G, \pi)$.

For each structure $A = A(G, \pi)$ we define a structure $B = B(G, \pi)$ as follows. B is a four-part structure $(\{a, b\}, G, Z_2, D)$, where $\{a, b\}, G, Z_2$ are as in A , and D is the set of all ordered pairs $g^* = \langle G, g \rangle$ with $g \in G$. Besides the operations of A , B carries functions $q_B: D \times G \rightarrow D$ and $p_B: D \rightarrow \{a, b\}$ defined by $q_B(g^*, h) = (gh)^*$, and $p_B(g^*) = a$ if and only if $\pi(g) = 0$.

CLAIM. If γ is an automorphism of $B(G, \pi)$ which fixes G pointwise, then there is a unique $h \in G$ such that, for every $g^* \in D$, $\gamma(g^*) = q_{B(G, \pi)}(g^*, h)$.

For let h^* be $\gamma(1^*)$. Then for each $g \in G$ we have

$$(15) \quad \gamma(g^*) = \gamma(q_B(1^*, g)) = q_B(h^*, g) = q_B((hg)^*, 1) = q_B(g^*, h).$$

The last equation uses the fact that G is abelian. This proves existence; uniqueness is immediate.

There is a formula $\psi(x, y)$ such that for all G and π , $\psi(A(G, \pi), -)$ defines $B(G, \pi)$. Moreover we can choose ψ so that it defines a construction F with $F_\psi(A(G, \pi)) = B(G, \pi)$. The only problem in choosing this ϕ is to determine, given a structure $(\{c, d\}, G, Z_2) = A(G, \pi)$, which of c, d corresponds to a . But global choice for pairs makes this decision for us.

Next, for any cardinal κ let R_κ be the set of sets of rank less than κ , let S_κ be the set of all well-orderings of R_κ , and let J_κ be the set of finite subsets of S_κ . We can make J_κ into an abelian group by putting $a + b = (a \setminus b) \cup (b \setminus a)$; the identity 0 is the empty set. We define a surjective homomorphism $\pi_\kappa: J_\kappa \rightarrow Z_2$ by setting $\pi_\kappa(a) = 0$ if and only if $|a|$ is even.

Now if the construction F is definably semi-natural, then it is defined by a formula θ such that F_θ is definably semi-natural. Therefore there is a definable function G such that for each cardinal κ , G maps $\text{Aut}(A(J_\kappa, \pi_\kappa))$ into $\text{Aut}(F_\theta(A(J_\kappa, \pi_\kappa)))$ so that each $G(\beta)$ extends β . For brevity we write C_κ for $F_\theta(A(J_\kappa, \pi_\kappa))$. Now C_κ forms a quadruple $(\{a, b\}, J_\kappa, Z_2, D_\kappa)$ with functions $\sigma_\kappa, p_\kappa, q_\kappa$ besides the group operations in J_κ and Z_2 .

For a fixed κ , let $\beta = \beta_\kappa$ be the automorphism of $A(J_\kappa, \pi_\kappa)$ which fixes J_κ pointwise but transposes a and b . Then $G(\beta)$ is an automorphism of C_κ which fixes J_κ pointwise and transposes a and b . For each $d \in D_\kappa$,

$$(16) \quad \sigma_\kappa(d) = 0 \Leftrightarrow p_\kappa(d) = a \Leftrightarrow p_\kappa(G(\beta)d) = \beta(a) = b \Leftrightarrow \sigma_\kappa(G(\beta)d) \neq 0.$$

It follows that $d \neq G(\beta)d$ for all $d \in D_\kappa$. So by the claim, there is some unique $h_\kappa \neq 0$ in J_κ such that for all $d \in D_\kappa$, $G(\beta)(d) = q_\kappa(d, h_\kappa)$. By global choice for finite sets, we can select an element x_κ of h_κ , uniformly for all κ . In short, for each cardinal κ we can uniformly define a well-ordering of R_κ . Clearly this supplies a definable global choice function for M .

Conversely it is straightforward to show that if M has a definable global choice function, then for any formula θ defining F , F_θ is definably semi-natural.

References

1. S. EILENBERG and S. MAC LANE, 'General theory of natural equivalences', *Trans. Amer. Math. Soc.* 58 (1945) 231–294.
2. CARL FAITH, *Algebra II: Ring theory* (Springer, Berlin 1976).
3. HARVEY FRIEDMAN, 'On the naturalness of definable operations', *Houston J. Math.* 5 (1979) 325–330.
4. WILFRID HODGES, 'A normal form for algebraic constructions II', *Logique et Anal.* (1975) 429–487; also in *Six days of model theory* (ed. Paul Henrard, Editions Castella, Albeuve, Switzerland 1977) pp. 197–255.
5. WILFRID HODGES, 'On the effectivity of some field constructions', *Proc. London Math. Soc.* (3) 32 (1976) 133–162.
6. WILFRID HODGES, 'Constructing pure injective hulls', *J. Symbolic Logic* 45 (1980) 544–548.
7. NATHAN JACOBSON, *Lectures in abstract algebra III: Theory of fields and Galois theory* (Van Nostrand, Princeton 1964).
8. SAUNDERS MAC LANE, *Categories for the working mathematician* (Springer, New York 1971).

School of Mathematical Sciences
Queen Mary College
Mile End Road
London E1 4NS

Department of Mathematics
University of Michigan
Ann Arbor
Michigan 48109
USA