SOME COMMENTS ON NATURALITY IN CATEGORY THEORY

SUPPLEMENT TO PETER MAY'S TQFT COURSE

Abstract. The goal of these notes is to make the concept of a natural transformation more natural. I apologize for the horrible pun, but there is no better way to put it. Also, I don't know if "naturality" is really a word.

1. Naturality with respect to isomorphisms

1.1. Let us start with the most naive approach to the concept of naturality. What does it mean for an algebraic construction to be "natural"? As a first approximation, let us recall that in algebra we often try not to distinguish between isomorphic objects. For instance, the standard Euclidean plane $V = \mathbb{R}^2$, and the subspace W of $\mathbb{R}^3$ formed by vectors (x_1, x_2, x_3) such that $x_1 + x_2 + x_3 = 0$, are isomorphic as abstract vector spaces over $\mathbb{R}$. Thus, if we wish to call some construction with vector spaces "natural", it should produce answers which "look the same" for the vector spaces V and W. Now let us try to make sense out of this vague idea.

1.2. The next step is to realize that if we are given an isomorphism between two algebraic objects, it usually makes sense to ask whether a given construction is compatible with this isomorphism or not. It would be futile to try to define these words in all-encompassing generality. Instead, let us look at some examples.

1.3. One thing we know one can do with vector spaces is to choose bases in them. For simplicity, let us stick to finite dimensional vector spaces. One can ask whether one can "naturally" choose a basis in each finite dimensional vector space. According to the philosophy explained above, this means the following. Does there exist a way to choose, for every finite dimensional vector space V, a basis $(v_1, \dots, v_n)$ of V in such a way that if $\phi : V \xrightarrow{\simeq} W$ is any isomorphism of vector spaces and $(v_1, \dots, v_n)$ and $(w_1, \dots, w_n)$ are the two "naturally chosen" bases[1] of V and W, respectively, then $\phi(v_j) = w_j$ for all $1 \leq j \leq n$.

The answer is NO already for one-dimensional vector spaces[2]. Namely, if V is any one-dimensional vector space, then any basis for V consists of just one nonzero

[1] We know the bases must have the same cardinality because the vector spaces are isomorphic.

[2] If you know about vector spaces over arbitrary fields, you will notice that I am being somewhat sloppy. Namely, if the ground field has only two elements, then, in fact, every one-dimensional vector space over this field has a natural basis. So, when the ground field is $\mathbb{F}_2$, you have to work with two-dimensional vector spaces to prove that no natural choice of bases is possible.

vector v, and the map $V \longrightarrow V$ given by multiplication by a scalar $\lambda \neq 0, 1$ is an isomorphism of V onto itself which does not take v to itself.

1.4. As another example, let us revisit one of the questions that came up in Peter's first lecture. Namely, for a given finite dimensional vector space V, does there exist a natural isomorphism between V and its dual, V^*?

If you interpret "natural" in the stronger sense of "natural transformation" of functors defined on the category of finite dimensional vector spaces (the morphisms being all linear maps of vector spaces), then, as Peter explained, the question itself does not make sense, because the identity functor $V \longmapsto V$ is co variant, while the "duality functor" $V \longmapsto V^*$ is contra variant.

However, with our weaker interpretation of "naturality", the question can be turned into a well posed one. Namely, one can ask if there exists a way to associate to every finite dimensional vector space V an isomorphism $\eta_V : V \xrightarrow{\simeq} V^*$ with the property that if $\phi : V \xrightarrow{\simeq} W$ is any isomorphism of finite dimensional vector spaces, then

$$(\phi^*)^{-1} \circ \eta_V = \eta_W \circ \phi.$$

The trick we are using here is as follows. We are trying to imitate the definition of a natural transformation between functors. Unfortunately, the adjoint map $\phi^* : W^* \longrightarrow V^*$ "goes the wrong way". However, since we are assuming that ϕ (and hence also ϕ^*) is an isomorphism, we can invert ϕ^* and write down a meaningful equation. So now we have a well defined question.

The answer to this question is also negative. You should understand why (once again, play with spaces of dimensions 1 and 2); this is a rather instructive exercise, and morally the answer is somewhat important, in spite of being negative.

1.5. Let us return to the question of why an algebraic construction should be called "natural" if it is compatible with isomorphisms between the objects that appear in the construction. I already gave one answer, namely, one should try not to distinguish between isomorphic objects. However, this does not quite explain why one should consider *all possible* isomorphisms.

Let us, however, approach the same question from a slightly different angle. Let us go back to a single vector space V. If we want to perform a "natural construction" with V, then intuitively it should be clear that we should only be allowed to play with symbols like $v, w, \dots$ (to denote some unspecified elements of V) and $\lambda, \mu, \dots$ (to denote scalars in the ground field), and also to use the vector space operations.

For example, how do we get a natural linear map $V \longrightarrow V^{**}$? We say: pick an arbitrary element v of V. We want to associate to it a linear functional T_v on V^*. This means we want to produce a scalar $T_v(f)$ for every $f \in V^*$. So we just put $T_v(f) := f(v)$. This construction "feels natural", and it is also natural in the precise sense of being compatible with all isomorphisms of vector spaces, as you can easily check. (That is, if $\phi : V \longrightarrow W$ is an isomorphism of vector spaces, then $T_{\phi(v)} = \phi^{**}(T_v)$ for all $v \in V$. Of course, here it is not necessary to restrict attention to isomorphisms: we might as well work with all linear maps ϕ.)

More generally, suppose we have any construction with vector spaces whatsoever which only involves playing with symbols like $v, w, \lambda, \mu, \ldots$ and also with vector space operations. Now suppose we have an isomorphism $\phi : V \longrightarrow W$ of vector spaces. Do the same construction, but playing with the symbols $\phi(v), \phi(w), \lambda, \mu, \ldots$ instead. Note that as v runs through the elements of V, $\phi(v)$ runs through the elements of W because ϕ is an isomorphism. Also, the assignment $v \mapsto \phi(v)$ is compatible with all vector space operations by the definition of a linear map. Thus, we should get the same result whether we first play with the symbols $v, w, \lambda, \mu, \ldots$ and then apply the map ϕ, or we play with the symbols $\phi(v), \phi(w), \lambda, \mu, \ldots$ directly and see what we get.

The notion of an algebraic construction being compatible with all isomorphisms between the objects involved in it is meant to formalize this very intuitive idea.

1.6. Another example. Let me take this opportunity to review something which came up during Peter's second lecture. If A is a finitely generated abelian group, Peter explained that we can define the torsion subgroup $T(A) \subset A$ so that the quotient $A/T(A)$ is a *free* finitely generated abelian group. He also showed that one can always find a splitting of the projection $A \xrightarrow{\pi} A/T(A)$. Then he asked two questions. First of all, can you find a *natural* splitting[3] of this projection? And second, if you cannot, is it at least possible to find some *monomorphism* $A/T(A) \hookrightarrow A$ which is natural with respect to A?

Peter gave an example showing that the second question has a negative answer. Notice, by the way, that Peter's example used morphisms between the abelian groups that he constructed which were NOT isomorphisms. It is an instructive exercise to try to analyze the same question from the point of view of our weaker notion of naturality. Namely, is it possible to choose, for every finitely generated abelian group A, an injective homomorphism $\eta_A : A/T(A) \hookrightarrow A$, so that if $\phi : A \xrightarrow{\simeq} B$ is any isomorphism of finitely generated abelian groups, then $\phi \circ \eta_A = \eta_B \circ \overline{\phi}$, where $\overline{\phi} : A/T(A) \xrightarrow{\simeq} B/T(B)$ is the isomorphism induced by ϕ in the obvious way?

Let me also write down Peter's comment about the possibility of doing this if you restrict attention to *finitely many* finitely generated abelian groups $A_1, \ldots, A_N$. Namely, each of the torsion subgroups $T(A_j)$ is finite and therefore annihilated by a natural number d_j; taking their product, we obtain a natural number d which annihilates each $T(A_j)$. This means that the homomorphism of multiplication by d induces a homomorphism $A_j/T(A_j) \longrightarrow A_j$, for every j, which must be injective

[3]Recall that a splitting of π is a homomorphism $\sigma : A/T(A) \longrightarrow A$ such that $\pi \circ \sigma = \mathrm{id}_{A/T(A)}$.

(check this!). Of course, this construction *is* natural in the sense of being compatible with arbitrary homomorphisms between the various A_j's, simply because any homomorphism of abelian groups commutes with multiplication by d.

The last comment is that the first question, namely, that of the existence of a natural *splitting* of the projection $A \longrightarrow A/T(A)$, is in fact much easier to answer. Of course, the answer is negative, and it is negative even with our weak notion of naturality. Moreover, one does not have to work with infinitely many groups: the group $A = \mathbb{Z} \oplus (\mathbb{Z}/2\mathbb{Z})$ will suffice. Indeed, let us use the projection onto the first coordinate, $A \longrightarrow \mathbb{Z}$, to identify $A/T(A)$ with $\mathbb{Z}$. Suppose that this projection does have a natural splitting $f : \mathbb{Z} \longrightarrow A$. We must then have $f(1) = (1, r)$, where $r \in \mathbb{Z}/2\mathbb{Z}$. Check that the following map is an automorphism of A: $\phi(m, t) = (m, t+\overline{m})$, where $\overline{m}$ is the image of m in $\mathbb{Z}/2\mathbb{Z}$. Clearly ϕ induces the identity automorphism of the quotient $A/T(A) = \mathbb{Z}$. However, $\phi \circ f \neq f$, because $\phi(1, r) \neq (1, r)$. This implies that f cannot be natural.

1.7. To end this section, let me mention that for some people the weak notion of naturality that we introduced, i.e., naturality only with respect to isomorphisms (as opposed to all morphisms), might seem more natural (sorry to keep doing this) than the notion of naturality with respect to all morphisms. There are, in fact, some important algebraic constructions which are natural at least in the colloquial sense of this word, but which are not natural with respect to *all* morphisms.

Here is an example. In group theory there is a notion of the "upper central series" of a given group. Namely, if G is a group, let $Z_1(G)$ denote the center of G. It is a normal subgroup of G, so we can form the quotient $G/Z_1(G)$. Let $Z_2(G)$ denote the preimage of the center of $G/Z_1(G)$ under the quotient homomorphism $G \longrightarrow G/Z_1(G)$. Continue this process inductively. You end up with an ascending chain[4] of normal subgroups $Z_1(G) \subseteq Z_2(G) \subseteq Z_3(G) \subseteq \dots$ of G.

This construction is clearly natural in the colloquial sense of this word. It is also natural with respect to isomorphisms: if $\phi : G \longrightarrow H$ is an isomorphism of groups, then it is obvious that $\phi(Z_n(G)) = Z_n(H)$ for all $n \in \mathbb{N}$. However, this construction is not natural with respect to *all* group homomorphisms: namely, if we only require ϕ to be a homomorphism, then, in fact, neither of the two subgroups $\phi(Z_n(G))$, $Z_n(H)$ of H needs to be contained in the other one (find counterexamples!).

Note that, strictly speaking, the upper central series construction does not fit into the framework of natural or unnatural transformations between functors. However, I am using it only for the purpose of illustrating general ideas. In any case, the concept of naturality in category theory goes beyond morphisms between functors.

[4]This chain does not have to be *strictly* ascending. For instance, we could have $Z_1(G) = \{1\}$, and then the chain become stationary right away. On the other hand, if $Z_n(G) = G$ for some $n \in \mathbb{N}$, then we say that G is a *nilpotent* group.

2. Naturality with respect to all morphisms

2.1. Let us now return to the definition of a natural transformation between functors given in Peter's course. Recall that if $\mathscr{C}$, $\mathscr{D}$ are categories and $F, G : \mathscr{C} \longrightarrow \mathscr{D}$ are (covariant) functors, a natural transformation from F to G is a collection of morphisms $\eta_X : F(X) \longrightarrow G(X)$ in $\mathscr{D}$, one for each object X of $\mathscr{C}$, with the property that for any morphism $f : X \longrightarrow Y$ in $\mathscr{C}$, the following diagram commutes:

$$\begin{array}{ccc} F(X) & \xrightarrow{\eta_X} & G(X) \\ {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle G(f)} \\ F(Y) & \xrightarrow{\eta_Y} & G(Y) \end{array}$$

We note right away that the weaker concept of naturality, introduced in the previous section, fits into this framework as well. Namely, we will now rephrase the requirement that the diagram above is commutative for all *isomorphisms* f in $\mathscr{C}$ in a different way by modifying the category $\mathscr{C}$.

2.2. There is a general way to obtain a groupoid (a category where every arrow is an isomorphism) from a given category $\mathscr{C}$. Namely, let us consider the subcategory $\mathscr{C}^\times$ whose objects are the objects of $\mathscr{C}$, but where $\mathrm{Hom}_{\mathscr{C}^\times}(X, Y)$ is defined as the set of arrows $f \in \mathrm{Hom}_{\mathscr{C}}(X, Y)$ which have a (left and right) inverse.

You should check that this is really a subcategory of $\mathscr{C}$. Of course, if $\mathscr{C}$ is not already a groupoid, then $\mathscr{C}^\times$ is not a full subcategory of $\mathscr{C}$.

Think about some examples. What happens if we do this to the category of sets? The category of abelian groups? A category with one object?

The process of passing from a category $\mathscr{C}$ to the groupoid $\mathscr{C}^\times$ is morally similar to the process of passing from a ring R to the group $R^\times$ of units in R.

2.3. Now suppose $\mathscr{C}$, $\mathscr{D}$ are categories and $F, G : \mathscr{C} \longrightarrow \mathscr{D}$ are (covariant) functors, as before. Check that we can "restrict" F and G to functors $F', G' : \mathscr{C}^\times \longrightarrow \mathscr{D}$. (In fact, you should also check that F' and G' automatically take $\mathscr{C}^\times$ into $\mathscr{D}^\times \subset \mathscr{D}$, but this is not important for our story.) Now you should be able to see that a transformation $\eta : F \longrightarrow G$ is natural in the weaker sense of being compatible with all isomorphisms in $\mathscr{C}$ if and only if it is a natural transformation between the functors $F', G' : \mathscr{C}^\times \longrightarrow \mathscr{D}$ in the sense that Peter defined.

2.4. Here is an exercise to help you learn how to play with these abstract concepts. Given a category $\mathscr{D}$, let us define a new category, $Ar(\mathscr{D})$, called the "category of arrows" in $\mathscr{D}$, as follows. The objects of $Ar(\mathscr{D})$ are really all the arrows of $\mathscr{D}$; or, if you wish to be a little bit more careful and pedantic, they are triples (X, Y, f), where X, Y are objects of $\mathscr{D}$ and $f \in \mathrm{Hom}_{\mathscr{D}}(X, Y)$. A morphism between two such triples,

(X_1, Y_1, f_1) and (X_2, Y_2, f_2), is a pair of arrows, $\phi : X_1 \longrightarrow Y_1$ and $\psi : X_2 \longrightarrow Y_2$, making the following diagram commute:

$$\begin{array}{ccc} X_1 & \xrightarrow{\ \phi\ } & X_2 \\ {\scriptstyle f_1}\big\downarrow & & \big\downarrow{\scriptstyle f_2} \\ Y_1 & \xrightarrow{\ \psi\ } & Y_2 \end{array}$$

You should be able to define the composition of morphisms and verify that $Ar(\mathscr{D})$ is indeed a category.

The exercise is to prove the following statement. Let $\mathscr{C}$ and $\mathscr{D}$ be categories. *To give a pair of functors $F, G : \mathscr{C} \longrightarrow \mathscr{D}$ and a natural transformation $F \longrightarrow G$ is the same as to give one functor $\mathscr{C} \longrightarrow Ar(\mathscr{D})$.*

2.5. Let me now present a completely different point of view both on the notion of a natural transformation and on the question of why this notion is important. From this new point of view, compatibility with *all* morphisms will in fact seem like a very natural requirement.

The philosophy that I will try to explain has very important and nontrivial applications in algebra, although I will only present toy examples of this philosophy.

2.6. In mathematics, many categories arise as categories of sets equipped with extra structures, where the morphisms are maps of sets that are compatible with these extra structures. For example, the categories of groups, abelian groups, vector spaces, topological spaces, manifolds, and so on, are all of this nature. To make this slightly more precise, note that in each of the examples of categories $\mathscr{C}$ that I mentioned, one has the forgetful functor from $\mathscr{C}$ to the category $\mathscr{S}$ of sets, and this forgetful functor is faithful (though not necessarily full). This has been discussed during Peter's first lecture.

Now if $\mathscr{C}$ is *any* category equipped with a faithful functor $F : \mathscr{C} \longrightarrow \mathscr{S}$ (where $\mathscr{S}$ is still the category of sets), we could (intuitively)think of $\mathscr{C}$ as being a category whose objects are sets equipped with some extra structure, and we could think of F as being the forgetful functor which "forgets" this extra structure.

Here is the main moral of the story. In a situation of the type I described, *it is often useful to play with all endomorphisms of the functor F, that is, natural transformations from F to itself, as a way of determining what natural structures are present on the sets $F(X)$, where X runs through all objects of $\mathscr{C}$.*

2.7. What is the point here? Let M denote the collection of all natural transformations $\eta : F \longrightarrow F$. Probably, to avoid set-theoretical issues, we should assume that $\mathscr{C}$ is a *small* category, i.e., the objects of $\mathscr{C}$ form a set. Then M is also a set. You should check that M is actually a *monoid* under composition of natural transformations. Now M acts on each of the sets $F(X)$ in an obvious way. Namely, we define a map

$$M \times F(X) \longrightarrow F(X) \qquad \text{by} \qquad (\eta, x) \longmapsto \eta_X(x).$$

Check that this map satisfies the axioms for the action of a monoid on a set. (If you know what a group action on a set is, the axioms for a monoid action are exactly the same: they do not involve inverting the group elements. If you don't know what a group action is, you should learn the definition.)

Moreover, if $f : X \longrightarrow Y$ is any morphism in $\mathscr{C}$, we obtain a map of sets $F(f) : F(X) \longrightarrow F(Y)$. In fact, this map commutes with the action of M! Check this; moreover, convince yourself that this statement is in fact *equivalent* to the requirement that M consists only of the *natural* transformations from F to itself.

2.8. To summarize, we started with a functor $F : \mathscr{C} \longrightarrow \mathscr{S}$, and we found a way to equip each of the sets $F(X)$ with the additional structure of an M-action, where M is the monoid of all natural transformations from F to itself. Moreover, F takes morphisms in the category $\mathscr{C}$ to maps of sets that commute with the M-action. Thus, tautologically, we found a way to "upgrade" F to a functor from $\mathscr{C}$ to the category of sets equipped with an M-action (or, for brevity, "M-sets").

2.9. Why is this useful? For instance, in some cases this upgraded functor is an equivalence of categories. Here is a simple, but morally important, exercise. Suppose that N is an arbitrary monoid, and let $\mathscr{C}$ denote the category of all N-sets. Let $F : \mathscr{C} \longrightarrow \mathscr{S}$ denote the obvious forgetful functor to the category of sets. Show that the monoid M of natural transformations from F to itself is isomorphic to N, and that the upgraded functor from $\mathscr{C}$ to the category of M-sets is an equivalence.

2.10. In other situations, games of this kind allow us to discover new structures in cases where we might not have noticed them otherwise. For example, if you know something about representation theory of groups, you know that the notion of the "group algebra" of a group plays an important role in this theory. Here is an (arguably natural) way to "invent" this definition.

Fix a group G and a field k. Let $\mathscr{C}$ denote the category of all representations of G over k, and let Vect_k denote the category of all vector spaces over k. We have the obvious forgetful functor $F : \mathscr{C} \longrightarrow \mathrm{Vect}_k$. Let A denote the monoid of all natural transformations from F to itself. Check, first of all, that A is also a vector space over k. Namely, we can define addition and multiplication by scalars in A "objectwise". (It is part of the exercise to make precise sense out of this; the definition is not difficult.) Next, show that A is in fact a k-algebra, that is, the composition operation in A is bilinear with respect to this vector space structure. Finally, construct an isomorphism between the group algebra of G over k and A.

To put it differently, what this exercise shows is that by playing with natural transformations from the forgetful functor $F : \mathscr{C} \longrightarrow \mathrm{Vect}_k$ to itself, we have discovered that representations of G over k have more structure (which is not entirely obvious from the definition of a group representation), namely, they have the structure of modules over the group algebra $k[G]$ of G.

2.11. To become more familiar with games of this sort, try to play with as many examples of forgetful functors as you can. For instance, can you describe the monoid of endomorphisms of the forgetful functor from the category of groups to the category of sets? What about the one from abelian groups to sets?

2.12. Finally, let me mention another type of games of a similar nature, but which are played with a different goal in mind. Fix a (small) category $\mathscr{C}$, and look at the monoid of endomorphisms of the *identity* functor from $\mathscr{C}$ to itself. This monoid is called[5] the Bernstein center of $\mathscr{C}$.

The word "center" is partially justified by the following exercise. Prove that the Bernstein center of any category is a *commutative* monoid (w.r.t. composition).

As far as I know, the word "Bernstein" is justified by the fact that the notion in question was introduced by Joseph Bernstein for the purpose of applying it to p-adic representation theory, whatever that means.

Another very instructive exercise is to compute the Bernstein centers of as many "interesting" categories as you can. Start with the category of abelian groups. More generally, do this for the category of left A-modules, where A is an arbitrary (not necessarily commutative) ring. What about the category of M-sets, where M is a monoid (or a group, if you wish)? The category of all groups?

Finally, if you know enough to make sense out of this exercise, use the notion of the Bernstein center to show that if A and B are commutative rings such that the category of A-modules is equivalent to the category of B-modules, then $A \cong B$. Is the commutativity requirement necessary?

This exercise is the beginning of a subject in algebra called "Morita theory".

[5]At least, this terminology is standard when $\mathscr{C}$ is an pre-additive category, which means that the set of arrows, $\mathrm{Hom}_{\mathscr{C}}(X, Y)$, between every pair of objects, X, Y, of $\mathscr{C}$, is equipped with an abelian group structure so that composition of morphisms is bi-additive.