Part IV.

Categories
10. Introduction to category theory

In this final part of the book, we look at some of the ways in which category theory offers a means of formalising some of the concepts of structure and equivalence that we have encountered so far; in particular, at how we can use categorical notions to bring together the logical and physical examples. In this chapter, we introduce the notion of a category, and look at some examples of categories.  

10.1. Motivation and definition

In our investigations, we have encountered various definitions of the form ‘a wotsit is a set, equipped with such-and-such bells and whistles’: for example, Tarski models, vector spaces, or groups. The result of such definitions is that we get a collection of ‘structured sets’ (e.g. groups), which have ‘structure-preserving mappings’ (e.g. group homomorphisms) between them. One way of approaching category theory is to note that a lot of the interesting mathematical information gets encoded in these mappings. For example, a subgroup $N$ of a group $G$ is a normal subgroup (i.e. is invariant under conjugation) iff there is some group homomorphism $\phi : G \rightarrow H$ such that $N$ is the kernel of $\phi$; so, roughly speaking, if you knew all the facts about group homomorphisms, then you could figure out which subgroups are normal. This motivates the study of the networks of structure-preserving mappings, and the development of a theory of such mappings: i.e., the postulation of axioms that any such collection of mappings should obey. We refer to such a network as a category, and axiomatise this notion as follows.

**Definition 27.** A category $\mathcal{C}$ consists of a class $|\mathcal{C}|$ of objects and a class $\text{Hom}(\mathcal{C})$ of arrows (also often referred to as morphisms), such that:

- Every arrow $f \in \text{Hom}(\mathcal{C})$ is associated with a pair of objects $A$ and $B$ of $\mathcal{C}$, referred to as its domain and codomain: we denote this by writing $f : A \rightarrow B$.

- For any two arrows $f : A \rightarrow B$ and $g : B \rightarrow C$, there is a third arrow from $A$ to $C$ called the composition of $f$ and $g$, and denoted $g \circ f$.

---

1For more on category theory, see Awodey (2010) or Halvorson (2019).
• Composition is associative: given three arrows $f : A \to B$, $g : B \to C$, and $h : C \to D$,

$$h \circ (g \circ f) = (h \circ g) \circ f$$  (10.1)

• Associated with every object $A$ in $C$ there is an identity arrow $\text{Id}_A : A \to A$: this arrow has the property that for any arrow $f : A \to B$, $\text{Id}_B \circ f = f = f \circ \text{Id}_A$.

Given any two objects $A$ and $B$ in a category $C$, the set of all arrows with domain $A$ and codomain $B$ will be denoted $C(A, B)$.

You should satisfy yourself that these axioms seem plausible conditions for the general notion of structure-preserving mappings between structured sets. Indeed, the following are all examples of categories:

• The category $\text{Grp}$, with groups as objects and group homomorphisms as arrows

• The category $\text{Vec}$, with vector spaces as objects and linear maps as arrows whose objects are vector spaces and whose arrows are linear maps

• The category $\text{Set}$, with sets as objects and functions as arrows

(Exercise: demonstrate that the above examples are indeed categories.)

However, having postulated those axioms, we can then study the structures that obey them without regard for whether all those structures are interpretable as collections of mappings or not; this is analogous to the way that group theory postulates some axioms intended to capture the notion of a set of transformations, but then goes on to study anything satisfying those axioms without necessarily thinking of it as a set of transformations. In other words, we abstract away from ‘concrete categories’ such as the above, to study all algebraic structures satisfying the axioms.² To illustrate this, here are some further examples of categories.

10.2. Examples

For categories with finitely many objects and arrows, we can explicitly describe the structure of the category. We start with a couple of examples of categories of this kind.

²This is not intended as a claim about the actual history of the development of category theory; for that history, see (Marquis, 2020, §2).
Example 2. The category $\mathbf{2}$ contains two objects, which we will label $X$ and $Y$; it contains the identity arrows $\text{Id}_X$ and $\text{Id}_Y$ (as it must) and one non-identity arrow $f : X \to Y$. We can depict this category, as follows:

\[
\begin{array}{c}
\text{X} \\
\text{f} \quad \text{Y} \\
\text{Id}_X \quad \text{Id}_Y
\end{array}
\]

$\mathbf{2}$ is quite a boring category, but not the most boring: that honour belongs (arguably) to its little cousin $\mathbf{1}$.

Example 3. The category $\mathbf{1}$ contains a single object $X$, and only the single arrow $\text{Id}_X$:

\[
\begin{array}{c}
X \\
\text{Id}_X
\end{array}
\]

Let’s go wild, and consider a category with three objects. (Incidentally, these names—$\mathbf{1}$, $\mathbf{2}$ and $\mathbf{3}$—are not especially canonical, so don’t be surprised if other books use different names for these categories, or use these names for different categories.)

Example 4. The category $\mathbf{3}$ has three objects $X, Y, Z$, and contains non-identity arrows $f : X \to Y, g : Y \to Z$, and $h : X \to Z$, where

\[ h = g \circ f \quad (10.2) \]

This category has the following diagram, where (as is usual) we no longer bother to draw the identity arrows:

\[
\begin{array}{c}
\text{X} \\
\text{f} \quad \text{Y} \quad \text{g} \\
\text{h} \quad \text{Z}
\end{array}
\]

Note that if we hadn’t specified the compositional relation (10.2), it would have been ambiguous how many arrows there are in the category in total; in general, specifying a category requires specifying the compositional structure among the arrows. This means that one has to be a little careful in using diagrams like the above to depict categories; in most instances, there is a (perhaps implicit) convention that the diagram commutes. Just to make this point clear, consider the following category:

Example 5. The category $\mathbf{1'}$ (whose name will be explained in the next chapter) has two objects $X$ and $Y$, and non-identity arrows $f : X \to Y$ and $g : Y \to X$, where

\[ g \circ f = \text{Id}_X \quad (10.3a) \]
\[ f \circ g = \text{Id}_Y \quad (10.3b) \]
We give the diagram of this category as follows:

\[
\begin{array}{c}
X \\
\circlearrowleft
\end{array}
\begin{array}{c}
f \\
\downarrow \\
f \circ g
\end{array}
\begin{array}{c}
Y
\end{array}
\]

Here, if we did not require that the compositional relations (10.3) held, then the category would have an infinite number of arrows: all the results of composing \( f \) and \( g \) together arbitrarily many times!

In the category \( 1' \), \( g \) and \( f \) are said to be *inverse* to one another. More generally:

**Definition 28.** For any arrow \( f : X \to Y \) in a category \( C \), an arrow \( g : Y \to X \) is said to be *inverse* to \( f \) if \( g \circ f = \text{Id}_X \) and \( f \circ g = \text{Id}_Y \).

This leads to the category-theoretic definition of *isomorphism*: an isomorphism is just an invertible arrow.

**Definition 29.** An arrow \( f : X \to Y \) in a category \( C \) is an *isomorphism* if there exists an inverse arrow \( f^{-1} : Y \to X \).

Note that identity arrows are self-inverse, and hence are always isomorphisms. A category like \( 1' \) in which every arrow is an isomorphism is referred to as a *groupoid*.

We’ve seen already that classes of mathematical objects with structure-preserving mappings between them often constitute a category; certain mathematical objects can themselves be regarded as categories, as we now discuss.

**Example 6.** A *partial order* is a set \( X \) equipped with a binary relation \( \leq \) that is reflexive, transitive, and antisymmetric: that is, for any \( x, y \in X \),

\[
\begin{align*}
x \leq x & \quad (10.4) \\
x \leq y, y \leq z \Rightarrow x \leq z & \quad (10.5) \\
x \leq y, y \leq x \Rightarrow x = y & \quad (10.6)
\end{align*}
\]

Any partial order can be considered to be a category, with the objects being the elements of \( X \), and with (exactly one) arrow between any pair of objects that stand in the relation \( \leq \). The composition of an arrow from \( x \) to \( y \) with an arrow from \( y \) to \( z \) is defined as the arrow from \( x \) to \( z \), which is guaranteed to exist by transitivity. For any \( x \), we take \( \text{Id}_x \) to be the arrow from \( x \) to itself (whose existence is guaranteed by reflexivity).

The category \( 2 \) can be regarded as a category of this kind, arising from the two-element partial order where one element is greater than the other.
Example 7. A preorder is a set $X$ equipped with a binary relation $\leq$ that is reflexive and transitive, but not (necessarily) antisymmetric: so there can be $x \neq y$ in $X$ such that $x \leq y$ and $y \leq x$. As with a partial order, any preorder can be regarded as a category with exactly one arrow between any pair of objects that stand in the relation $\leq$.

Example 8. Any set $X$ can be regarded as a category: one with no arrows other than the identity arrows. (Such a category is referred to as a discrete category.)

To be clear, the fact that any set can be regarded as a (discrete) category is distinct from the fact that there is a category Set of sets: it is both the case that any individual set can be regarded as a category, and that the collection of all sets forms a category.\(^3\)

Example 9. A group $G$ can be considered to be a category, with exactly one object and the group elements being the arrows (all of them arrows from that one object to itself). Composition of arrows is identified with group multiplication, and the identity arrow is identified with the group identity element; the group axioms then guarantee that the categorical axioms are satisfied.

Since every element of a group has an inverse, a group (regarded as a category) is one where every arrow is an isomorphism—i.e., a groupoid.

As with sets, the fact that every group can be regarded as a category is distinct from the fact that there is a category Grp of groups; rather, the point is that an individual group (say, the Lorentz group) may be regarded as a category. Note that when we regard a set as a category, the elements of the set get represented as objects; by contrast, when we regard a group as a category, the elements of the group get represented as arrows. This reinforces the point that in many (arguably, most) categories, it is the arrows that carry the interesting structure. Our final example also demonstrates this point (and will be discussed further in the next chapter).

Example 10. We define a category Mat as follows. The objects of Mat are the natural numbers, and the arrows in Mat from $m$ to $n$ are (all) the real $n \times m$ matrices; composition of arrows is given by matrix multiplication (i.e. given $M : m \to n$ and $N : n \to p$, $N \circ M = NM$) and the identity arrows are the identity matrices.\(^4\) This is a category: if $M$ is an $n \times m$ matrix and $N$ is a $p \times n$ matrix, then $NM$ is a $p \times m$ matrix; and matrix multiplication is an associative operation, for which the identity matrices act as identities.

---

\(^3\)Similarly, there is a category Pos which has all partially ordered sets as its objects, and order-preserving maps as its arrows; this should not be confused with the fact that any partially ordered set can be regarded as a category.

\(^4\)See Appendix A.
11. Functors between categories

11.1. Functors

In considering mathematical objects of a certain kind, one often wants to know how they relate to other kinds of mathematical object. In the context of category theory, the standard tool for making these kinds of comparison is the functor. A functor is a ‘homomorphism of categories’, its precise definition is as follows.

**Definition 30.** Let \( C \) and \( D \) be categories. A functor \( F \) from \( C \) to \( D \) consists of

1. a map \( F_{\text{obj}} : |C| \to |D| \); and
2. for every \( A, B \in |C| \), a map \( F_{AB} : C(A, B) \to D(F_{\text{obj}}(A), F_{\text{obj}}(B)) \)

such that the following two conditions hold:

- For any \( f : A \to B \) and \( g : B \to C \) in \( C \),

\[
F_{AC}(g \circ f) = F_{BC}(g) \circ F_{AB}(f)
\]

(11.1)

- For any \( A \in |C| \),

\[
F_{AA}(\text{Id}_A) = \text{Id}_{F_{\text{obj}}(A)}
\]

(11.2)

In the interests of reducing notational clutter, we will typically just use \( F \) to denote the maps \( F_{\text{obj}} \) and \( F_{AB} \) (for any \( A, B \in |C| \)); the argument of the map will make it clear which one is meant. Hence, for example, the equations (11.1) and (11.2) can be written as

\[
F(g \circ f) = F(g) \circ F(f)
\]

(11.3)

\[
F(\text{Id}_A) = \text{Id}_{F(A)}
\]

(11.4)
To illustrate the idea, let’s look at some examples of functors. Here is a functor $F : 2 \to 1'$:

\begin{align*}
F(A) &= X \\
F(B) &= Y \\
F(f) &= j
\end{align*}

(11.5)  
(11.6)  
(11.7)

Note that we don’t need to specify what $F$ does to the identity arrows: once we’ve specified $F$’s action on objects, it must send $\text{Id}_X$ to $\text{Id}_{F(X)}$, by condition (11.2). In the other direction, here’s a functor $G : 1' \to 2$:

\begin{align*}
G(X) &= A \\
G(Y) &= A \\
G(j) &= G(k) = \text{Id}_A
\end{align*}

(11.8)  
(11.9)  
(11.10)

Naturally, these examples are a bit trivial. Here is a less trivial example: as discussed in Appendix B, every vector space may be regarded as a group (with vector addition as the group operation). This means that there is a functor from the category $\text{Vec}$ of vector spaces to the category $\text{Grp}$ of groups: it maps any vector space to itself, or—perhaps better—to the ‘copy’ of itself that lives in the category of groups.

As already mentioned, functors are structure-preserving mappings between categories. As we discussed in the last chapter, collections of objects with structure-preserving mappings between them are paradigm cases of categories, and this example is no different. Indeed, if we compose two functors $F : \mathcal{C} \to \mathcal{D}$ and $G : \mathcal{D} \to \mathcal{E}$, then we find that the result is a functor $G \circ F : \mathcal{C} \to \mathcal{E}$. Moreover, this composition operation is associative $(H \circ (G \circ F) = (H \circ G) \circ F)$, and for every category $\mathcal{C}$, there is an identity functor $\text{Id}_\mathcal{C} : \mathcal{C} \to \mathcal{C}$ which maps every object and arrow in $\mathcal{C}$ to itself (you should convince yourself that this does, indeed, satisfy the definition of a functor). So there is a category of categories, denoted $\text{Cat}$, whose objects are categories and whose arrows are functors.

### 11.2. Equivalence functors

However, this does not mean that any two categories related by a functor should be regarded as equivalent, any more than the existence of a homomorphism between groups means those should be regarded as equivalent. So when should we regard two categories as structurally equivalent? One natural proposal is to do so when they are iso-
morphic, which we can cash out in category-theoretic terms (given that, as we’ve just observed, there is a category of categories):

**Definition 31.** Given categories $C$ and $D$, a functor $F : C \to D$ is an isomorphism of categories if there is a functor $G : D \to C$ such that $G \circ F = \text{Id}_C$ and $F \circ G = \text{Id}_D$. ♦

However, for many purposes a somewhat weaker notion is useful: that of categorical equivalence. Indeed, as we will discuss in the next chapter, the fact that categories admit this interestingly weaker notion accounts for much of the interest in category theory as a way of making precise certain ideas about equivalence in philosophy of science. Essentially, an equivalence functor is an ‘isomorphism up to isomorphism’; more precisely, the definition is as follows.¹

**Definition 32.** Given categories $C$ and $D$, a functor $F : C \to D$ is an equivalence of categories if $F$ is full, faithful, and essentially surjective, where:

1. $F$ is full if for any objects $A$ and $B$ of $C$, $F_{AB} : C(A, B) \to D(F(A), F(B))$ is surjective.
2. $F$ is faithful if for any objects $A$ and $B$ of $C$, $F_{AB} : C(A, B) \to D(F(A), F(B))$ is injective.
3. $F$ is essentially surjective if for any object $X$ of $D$, there is some object $A$ of $C$ such that $F(A)$ is isomorphic (in $D$) to $X$.

♦

Perhaps the most significant difference between isomorphism and equivalence of categories is that two categories can be equivalent even if they have different numbers of objects. For example, the following functor is an equivalence from $1'$ to $1$:

\[
F(X) = F(Y) = 1 \quad (11.11) \\
F(j) = F(k) = \text{Id}_1 \quad (11.12)
\]

This functor is surjective, hence essentially surjective; and although $F$ is not bijective on arrows overall, it is bijective on the arrows between any chosen pair of objects. In the

¹A more conceptually revealing (but less readily applicable) definition is that a functor $F : C \to D$ is an equivalence if there is an ‘almost inverse’ functor $G : D \to C$: a functor such that $G \circ F$ is ‘naturally isomorphic’ to $\text{Id}_C$ and $F \circ G$ is ‘naturally isomorphic’ to $\text{Id}_D$. This makes the relationship to categorical isomorphism clearer, but requires introducing natural transformations, which we don’t have the space to do here. For discussion (and a proof that these two notions coincide), see (Awodey, 2010, chap. 7).
other direction, the following functor is an equivalence from 1 to 1:

\[ G(1) = X \]  
(11.13)

(The fact that \( G(\text{Id}_1) = \text{Id}_X \) is implied by functoriality.) This functor is not surjective; but it is essentially surjective, since \( Y \) is isomorphic to \( X \).

Here are two further examples of equivalences, which again illustrate the fact that equivalence can be weaker than isomorphism.

**Proposition 9.** Any preorder \( X \) is categorically equivalent to some poset (where both are regarded as categories).

**Proof.** Two objects \( x \) and \( y \) of \( X \) are isomorphic if \( x \preceq y \) and \( y \preceq x \). We define the corresponding poset by first quotienting \( X \) by isomorphism, to obtain a set \( Y \): that is, elements of \( Y \) are equivalence classes of isomorphic elements of \( X \). Let us denote the equivalence class containing \( x \in X \) as \([x]\). We then define a binary relation \( \preceq \) on \( Y \), according to

\[ [x] \preceq [y] \iff x \preceq y \]  
(11.14)

The transitivity of \( \preceq \) guarantees that this definition is well-posed, in the sense of being independent of the choice of \( x \) and \( y \) from within an equivalence class. Furthermore, the reflexivity and transitivity of \( \preceq \) entail the reflexivity and transitivity of \( \preceq \). It remains only to confirm that \( \preceq \) is anti-symmetric. Indeed, if \([x] \preceq [y]\) and \([y] \preceq [x]\), then \( x \preceq y \) and \( y \preceq x \); hence \( x \) and \( y \) are isomorphic, and so \([x] = [y]\).

**Proposition 10.** The category \( \text{FinVect} \) of finite-dimensional vector spaces (with linear maps as arrows) is equivalent to the category \( \text{Mat} \) (of natural numbers with matrices as arrows).

**Proof.** First, equip every vector space \( V \) in \( \text{FinVect} \) with an (arbitrary) ordered basis \( e_i^V \). Now define a functor \( F : \text{FinVect} \to \text{Mat} \) as follows. For any finite-dimensional vector space \( V \), \( F(V) \) is the dimension of \( V \). For any linear transformation \( f : V \to W \), where \( \dim(V) = m \) and \( \dim(W) = n \), \( F(f) \) is the \( n \times m \) matrix representing \( f \) relative to those bases:

\[ f(e_i^V) = F(f)^i_\cdot e_j^W \]  
(11.15)

\(^2\)Strictly, I ought to write \( \tilde{e}_i^V \), but that looks terrible; I trust the reader to remember that these objects are vectors.

\(^3\)See Appendix A.
First, we show that $F$ is indeed a functor. If we have two linear transformations $f : U \to V$ and $g : V \to W$, then

\begin{align*}
F(g \circ f)^k_j e^W_k &= g(f(e^U_i)) \\
&= g(F(f)^l_i e^V_l) \\
&= F(g)^k_j F(f)^l_i e^W_l
\end{align*}

from which it follows that $F(g \circ f)^k_i = F(g)^k_j F(f)^l_i = (F(g)F(f))^k_i$, i.e. that $F$ preserves composition of arrows. Furthermore, if $I$ is the identity transformation on $V$, then

$$e^V_j = F(I)^l_i e^V_l$$

(11.16)

for which the only solution is $F(I)^l_i = \delta^l_i$; thus, $F$ preserves identity arrows.

We now show that $F$ is an equivalence functor. First, for any natural number $n$, there is some vector space $V$ in $\text{FinVect}$ of dimension $n$; so $F$ is essentially surjective (indeed, surjective).

Second, consider any vector spaces $V$ and $W$ in $\text{FinVect}$. If $f$ and $g$ are arrows from $V$ to $W$ in $\text{FinVect}$ such that $F(f) = F(g)$, then for every basis vector $e^V_i$ in $V$,

\begin{align*}
F(e^V_i) &= F(f)^l_i e^W_l \\
&= F(g)^l_i e^W_l \\
&= g(e^V_i)
\end{align*}

Thus, $f$ and $g$ agree on the basis vectors, and hence on all vectors: that is, $f = g$.

So $F$ is faithful.

Finally, again consider any vector spaces $V$ and $W$ in $\text{FinVect}$; suppose that their dimensions are $m$ and $n$ respectively. Then for any $n \times m$ matrix $M^l_i$, define a linear transformation $f : V \to W$ be defined by the condition that for any basis vector $e^V_i$ in $V$,

$$f(e^V_i) = M^l_i e^W_l$$

(11.17)

It follows immediately that $F(f) = M$. So $F$ is full.

One might feel somewhat disquieted by this example: on the face of it, it appears to suggest that a vector space has the same structure as a natural number, which seems
either false or nonsensical.\textsuperscript{4} However, we need to be careful, since (as was discussed in Chapter 10) categories will often encode their most interesting structure in their networks of arrows. And in fact, in this case, we can recover the internal vector-space structure from the categorical structure: the vectors in an \( n \)-dimensional space \( \mathbb{V} \) may be identified with the set of linear maps from a 1-dimensional vector space to \( \mathbb{V} \), and the vector-space operations (addition and scalar multiplication) can be defined using categorical data.\textsuperscript{5} This enables us to recover vectors from the category \textbf{FinVect}, or—what comes to the same thing—the category \textbf{Mat}.

11.3. Forgetful functors

A functor which is \textit{not} an equivalence functor is referred to as a \textit{forgetful functor}.\textsuperscript{6} A nice way to be precise about the concept of a forgetful functor, and about what it is that they are forgetting, is the so-called “stuff, structure, properties” perspective. The intuition here is that there are three ways in which a mathematical object can be interesting, and hence three ways of making it less interesting:

In math we’re often interested in equipping things with extra structure, stuff, or properties …. For example, a group is a set (stuff) with operations (structure) such that a bunch of equations hold (properties).\textsuperscript{7}

In other words, the \textit{stuff} is the raw materials out of which the object is ‘made’; the \textit{structure} comprises the relations and properties which organise the stuff into an interesting mathematical object; and the \textit{properties} are the properties that the object exhibits, in virtue of having its stuff organised thus-and-so by its structure. Philosophers would identify the stuff as the ‘ontology’, the structure as the ‘ideology’, and the properties as the ‘facts’. Given a \( \Sigma \)-model \( \mathcal{A} \), the stuff would be the domain \( |\mathcal{A}| \), the structure the signature \( \Sigma \) (or perhaps the set of formulae \( \text{Form}(\Sigma) \)), and the properties would be the sentences that \( \mathcal{A} \) satisfies (or perhaps all the facts about which formulae the tuples of \( \mathcal{A} \) satisfy).

Category theory offers a nice way of formalising this distinction (that’s somewhat more general than the model-theoretic formalisation). Given a functor \( F : C \to D \), we say that \( F \):

\textsuperscript{4}See Hudetz (2019) for an elaboration of this concern.
\textsuperscript{5}Dewar (nd)
\textsuperscript{6}At least, from a certain perspective. In the literature, the term “forgetful functor” is frequently used with a more vague denotation.
\textsuperscript{7}(Baez and Shulman, 2010, p. 15)
• forgets at most properties if it is full and faithful;
• forgets at most structure if it is faithful;
• forgets at most stuff if it is arbitrary.

Note that the stuff-structure-properties distinction is somewhat hierarchical in nature. Without structure, an object could not satisfy any (interesting) properties; and without stuff, an object would not be able to exhibit any (interesting) structure. Correspondingly, each of these levels of forgetfulness subsumes the one above: if a functor forgets stuff, then it might also forget structure and properties; and if it forgets structure, then it might also forget properties. To see how this classification of functors fits with the intuitive stuff-structure-properties distinction, let’s consider some examples.

First, suppose that a functor $F : C \to D$ forgets properties: i.e., that it is full and faithful, but not essentially surjective. That means that there are objects in $D$ that lie outside the image of $F$, even up to isomorphism: that is, which are not isomorphic to any object in the image of $F$. For example, let $\text{AbGrp}$ be the category of Abelian groups, i.e., groups whose group multiplication operation is commutative, with group homomorphisms as arrows. Every Abelian group is, of course, a group: so there is a functor $F : \text{AbGrp} \to \text{Grp}$, which simply maps every Abelian group to itself (or, if you prefer, to its “copy” in the category $\text{Grp}$). This functor is full and faithful: since $F$ simply embeds $\text{AbGrp}$ within $\text{Grp}$, for any Abelian groups $G$ and $H$, $\text{AbGrp}(G, H) = \text{Grp}(F(G), F(H))$. But it is not essentially surjective: no non-Abelian group is isomorphic to any Abelian group, and only Abelian groups lie in the image of $F$. Thus, $F$ forgets properties: it forgets the property of being Abelian.

Second, consider a functor $F : C \to D$ which forgets structure (and properties) but not stuff: that is, which is faithful but not full. That means that there are objects $X$ and $Y$ of $C$ such that the induced function $F : C(X, Y) \to D(F(X), F(Y))$ is injective but not surjective. So (intuitively) there are more arrows between $F(X)$ and $F(Y)$ in $D$ than there are between $X$ and $Y$ in $C$. Insofar as we are taking the arrows to represent structure-preserving maps, a pair of objects will admit more arrows by virtue of being less structured: hence, the sense in which $F$ forgets structure. As an example, consider the functor $F : \text{Grp} \to \text{Set}$ which maps any group to its underlying set, and any group homomorphism to its corresponding function. If two homomorphisms $f, g : G \to H$ are distinct from one another, then they must correspond to different functions between the underlying sets $|G|$ and $|H|$; hence, $F$ is faithful. But (in general) there are many

---

8The below sequence of examples is taken from https://ncatlab.org/nlab/show/stuff,+structure,+property.
functions between \(|G|\) and \(|H|\) that do not correspond to homomorphisms, and so \(F\) is not full. Thus, \(F\) forgets structure: it forgets group structure.

Finally, consider a functor \(F : \mathcal{C} \to \mathcal{D}\) which forgets stuff: that is, which is not faithful. Then there are objects \(X\) and \(Y\) of \(\mathcal{C}\) such that the induced function \(F : \mathcal{C}(X,Y) \to \mathcal{D}(F(X),F(Y))\) is not injective: that is, that for some pair of arrows \(f,g : X \to Y\) such that \(f \neq g\), \(F(f) \neq F(g)\). Intuitively, then, the arrows between \(F(X)\) and \(F(Y)\) are less “fine-grained” than those between \(X\) and \(Y\): there are more ways of mapping \(X\) to \(Y\) than there are ways of mapping \(F(X)\) to \(F(Y)\). Given that more stuff provides more “raw material” for a mapping, getting rid of stuff reduces the number of ways such a mapping could be performed. As a (somewhat trivial) example, consider the (unique) functor \(F : \textbf{Set} \to \textbf{1}\); this functor sends every set to 1, and every function to \(\text{Id}_1\). Thus, \(F\) forgets stuff: it forgets the stuff that makes up the sets.
12. Categories of theories

There are many things that one can do with category theory. Our interest in the topic is rather more specific: we are interested in using category-theoretic tools to illuminate the concepts of structure and equivalence that we have been discussing so far. As such, this final chapter considers how some of what we have already done can be put into a category-theoretic form. The first section considers how to apply category theory to the model-theoretic considerations of Part I, and the second looks at how we can use category theory in the context of the physical theories in Parts II and III.

12.1. Categories of Tarski-models

In Part I, we considered the concept of a class of models of a first-order theory. Now that we have category-theoretic resources to hand, we can work with the richer concept of a category of models.

Definition 33. Given a theory $T$, the category of models of $T$ is a category whose objects are models of $T$ and whose arrows are elementary embeddings.

As we discussed in Chapter 1, there are various kinds of mappings between models that one might work with in model theory. So why do we choose elementary embeddings to be the arrows in our category of models? The reason is that if we do so, then translations between theories induce functors between their categories of models:

Proposition 11. Given theories $T_1$ and $T_2$, let $\text{Mod}(T_1)$ and $\text{Mod}(T_2)$ be their categories of models. If $\tau : T_1 \rightarrow T_2$ is a translation, then $\tau^*$ is extendable to a functor, by stipulating that for any elementary embedding $h : \mathcal{A} \rightarrow \mathcal{B}$ (where $\mathcal{A}, \mathcal{B} \in \text{Mod}(T_2)$), $\tau^*(h) = h$.

Proof. Suppose that $\tau^*(h)$, i.e. $h$, is not an elementary embedding of $\tau^*(\mathcal{A})$ into $\tau^*(\mathcal{B})$: that is, that there is some $\Sigma_1$-formula $\phi(x_1, \ldots, x_n)$ and some $a_1, \ldots, a_n \in |\mathcal{A}|$ such that $\tau^*(\mathcal{A}) \models \phi[a_1, \ldots, a_n]$ but $\tau^*(\mathcal{B}) \not\models \phi[h(a_1), \ldots, h(a_n)]$. It follows that $\mathcal{A} \models \tau(\phi)[a_1, \ldots, a_n]$.

---

1The ideas in this section are covered in much greater detail in Halvorson (2019).
and $\mathfrak{B} \not \models \tau(\phi)[h(a_1), \ldots, a_n]$; so $h$ is not an elementary embedding of $\mathfrak{A}$ into $\mathfrak{B}$. It follows that given an elementary embedding $h$ in $\text{Mod}(T_2)$, $\tau^*(h)$ is an elementary embedding in $\text{Mod}(T_1)$. 

Had we taken homomorphisms or embeddings as arrows, this would not hold true, as the following examples demonstrate.

**Example 11.** Consider the following theories, $T_1$ and $T_2$. $T_1$ has signature $\{P\}$, and axioms

$$\exists x \exists y(x \neq y \land \forall z(z = x \lor z = y))$$  
(12.1a)

$$\exists x P x$$  
(12.1b)

whilst $T_2$ has signature $\{Q\}$, and axioms

$$\exists x \exists y(x \neq y \land \forall z(z = x \lor z = y))$$  
(12.2a)

$$\exists x \neg Q x$$  
(12.2b)

Figure 12.1 displays the models of these theories: the models of $T_1$ are $\mathfrak{A}_1$ and $\mathfrak{B}_1$, and the models of $T_2$ are $\mathfrak{A}_2$ and $\mathfrak{B}_2$.

The map

$$Q x \mapsto \neg P x$$  
(12.3)

is a translation from $T_2$ to $T_1$; the associated semantic map will map $\mathfrak{A}_1$ to $\mathfrak{A}_2$ and $\mathfrak{B}_1$ to $\mathfrak{B}_2$. But there is a homomorphism from $\mathfrak{A}_1$ to $\mathfrak{B}_1$ yet no homomorphisms from $\mathfrak{A}_2$ to $\mathfrak{B}_2$; so there can be no functor from the category of models of $T_1$ with homomorphisms as arrows to the category of models of $T_2$ with homomorphisms as arrows.

**Example 12.** Let $T_1$ be the theory in signature $\{P^{(1)}\}$ with the following axioms:

$$\exists x \exists y(\forall z(z = x \lor z = y))$$  
(12.4)

$$\exists x(P x \land \forall y(P y \rightarrow y = x))$$  
(12.5)

In English: there are at most two things, of which exactly one is $P$. Let $T_2$ be the theory in signature $\{R^{(2)}\}$ with the following axioms:

$$\exists x \exists y(\forall z(z = x \lor z = y))$$  
(12.6)

$$(\forall x \forall y(R x y \leftrightarrow x = y) \leftrightarrow \exists y \exists z(y \neq z))$$  
(12.7)
In English: there are at most two things, and $R$ is reflexive just in case there are two things. The models of $T_1$ and $T_2$ are depicted and labelled in Figure 12.2.

The map

$$Rxy \mapsto (x = y \land \exists z \neg Pz) \quad (12.8)$$

is a translation from $T_2$ to $T_1$. However, although there is an embedding $h$ from $\mathcal{A}_1$ to $\mathcal{B}_1$, there are no embeddings from $\mathcal{A}_2$ to $\mathcal{B}_2$; hence, there can be no functor from the category of models of $T_1$ with embeddings as arrows to the category of models of $T_2$ with embeddings as arrows.

Thus, we take the category of models of a theory to be one with elementary embeddings as arrows. We have already seen that invertible translations induce bijective maps on models; since the extension of such a map to a functor just acts as the identity on elementary embeddings, it follows more or less immediately that the functor is an isomorphism between the categories of models—and hence, that it is an equivalence.
Thus, intertranslatability entails categorical equivalence.

In the other direction, categorical equivalence does *not* entail intertranslatability, as the following example (due to Barrett and Halvorson (2016b)) demonstrates.

**Example 13.** $T_1$ is in the signature $\Sigma_1 = \{P_{0}^{(1)}, P_{1}^{(1)}, \ldots \}$, and consists of the sole axiom

$$\exists x \forall y (y = x)$$

(12.9)

$T_2$ is in the signature $\Sigma_2 = \{Q_{0}^{(1)}, Q_{1}^{(1)}, \ldots \}$, and consists of the axioms

$$\exists x \forall y (y = x)$$

(12.10)

$$\forall y(Q_{0}y \rightarrow Q_{1}y)$$

(12.11)

$$\forall y(Q_{0}y \rightarrow Q_{2}y)$$

(12.12)

$$\vdots$$

That is, both $T_1$ and $T_2$ assert that there is exactly one thing; but $T_2$ asserts that the
predicate \( Q_0 \) is such that if the unique thing satisfies the predicate \( Q_0 \), then it satisfies every other predicate \( Q_i \).

If we identify isomorphic models then the categories of both \( T_1 \) and \( T_2 \) are discrete, i.e., contain no arrows that are not identity arrows. Moreover, any model of of \( T_1 \) or \( T_2 \) can be specified by a subset of \( \mathbb{N} \): we include \( n \) in the subset just in case the model satisfies \( \exists x P_n x \) (for models of \( T_1 \)), or \( \exists x Q_n x \) (for models of \( T_2 \)). In the other direction, any subset of \( \mathbb{N} \) determines a model of \( T_1 \), whilst \( \mathbb{N} \) itself and any subset of \( \mathbb{N} \setminus \{0\} \) determines a model of \( T_2 \). There are \( \aleph_0 \) members of \( \mathbb{N} \), and \( \aleph_0 \) members of \( \mathbb{N} \setminus \{0\} \); so up to isomorphism, there are \( 2^{\aleph_0} \) models of \( T_1 \) and \( 2^{\aleph_0} \) models of \( T_2 \). Since \( \text{Mod}(T_1) \) and \( \text{Mod}(T_2) \) are discrete, any bijection between them is an equivalence of categories.

However, \( T_1 \) and \( T_2 \) are not intertranslatable. Suppose that they were, with inverse translations \( \tau: T_1 \to T_2 \) and \( \sigma: T_2 \to T_1 \). Let \( \mathcal{B} \) be the model of \( T_2 \) that corresponds to \( \mathbb{N} \), i.e. that satisfies \( \exists x Q_0 x \) for all \( i \in \mathbb{N} \). Now consider the sentence \( \exists x Q_0 x \). We know that \( \mathcal{B} \models \exists x Q_0 x \), and that (up to isomorphism) it is the only model of \( T_2 \) to do so. So \( \tau^* \mathcal{B} \models \sigma(\exists x Q_0 x) \), since \( \tau \) and \( \sigma \) are inverse to one another. Since \( \sigma(\exists x Q_0 x) \) is a \( \Sigma_1 \)-formula of finite length, there must exist some \( P_i \in \Sigma_1 \) which does not occur in it. So now let \( \mathfrak{A} \) be the model obtained from \( \tau^* \mathcal{B} \) by ‘switching’ the value of \( P_i \): if the sole object satisfies \( P_i \) in \( \tau^* \mathcal{B} \) then it does not in \( \mathfrak{A} \), and vice versa (and otherwise, \( \tau^* \mathcal{B} \) and \( \mathfrak{A} \) are identical). Since \( P_i \) does not occur in \( \sigma(\exists x Q_0 x) \), it follows that \( \mathfrak{A} \models \sigma(\exists x Q_0 x) \). So \( \sigma^* \mathfrak{A} \models \exists x Q_0 x \); and hence, \( \sigma^* \mathfrak{A} = \mathcal{B} \). But since \( \sigma^*(\tau^* \mathcal{B}) = \mathcal{B} \), it follows that \( \sigma^* \) is not injective, and hence not bijective; so \( \tau \) and \( \sigma \) cannot be a pair of inverse translations after all.

Therefore, categorical equivalence of theories is a strictly weaker notion than intertranslatability. Whether this is a feature or a bug is a thorny question: it will depend on whether the theories \( T_1 \) and \( T_2 \) in the above example ‘ought’ to be regarded as equivalent or not, and the answer to that is not obvious. In the meantime, there is interesting work to be done in further clarifying the relationship between definability or translatability on the one hand, and categorical structure on the other. For example, Hudetz (2019) outlines a way to strengthen categorical equivalence so as to make it sensitive to the definability of the models of one theory in terms of the other, and shows that this is equivalent to intertranslatability ‘up to surplus structure’ (in a sense that is made precise). On the other hand, Barrett (nd) shows that if an equivalence functor is induced from a translation, then the models of the two theories are are codeterminate with one another (although in a sense weaker than that arising from full intertranslatability).
12.2. Categories of electromagnetic models

Let us now consider categories of models in physics; in the interests of space I will only discuss electromagnetism, but the main ideas here could also be applied to Newtonian mechanics. To the theory of electromagnetic fields we associate a category $\mathcal{F}$, and to the theory of electromagnetic potentials we associate a category $\mathcal{A}$, where

- An object of $\mathcal{F}$ is a 2-form $F$ on Minkowski spacetime $M$ that satisfies Maxwell’s equations (7.12); and

- An arrow in $\mathcal{F}$ between the 2-forms $F$ and $F'$ is an isometry $\psi : M \to M$, such that $F' = \psi^* F$.

- An object in $\mathcal{A}$ is a 1-form $A$ on Minkowski spacetime that satisfies equation (8.4); and

- An arrow in $\mathcal{A}$ between the 1-forms $A$ and $A'$ consists of an isometry $\psi : M \to M$ and an exact one-form $G$ on $M$, such that

$$A' = \psi^* A + G$$

In other words, the category $\mathcal{F}$ represents field-theoretic solutions of Maxwell’s equations, with isometries regarded as isomorphisms; and the category $\mathcal{A}$ represents potential-theoretic solutions of Maxwell’s equations, with isometries and gauge transformations regarded as isomorphisms. In Chapter 8 I suggested that so long as we confined our attention to contractible spaces—such as Minkowski spacetime—there is a sense in which the theory of electromagnetic fields is equivalent to the gauge-invariant content of the theory of potentials. The following proposition, due to Weatherall (2016), can be thought of as a way of making this idea precise.

**Proposition 12** (Weatherall (2016), Proposition 5.5). The categories $\mathcal{F}$ and $\mathcal{A}$ are equivalent.

**Proof.** We define a functor $W : \mathcal{A} \to \mathcal{F}$ as follows. First, given any object $A$ in $\mathcal{A}$, we define

$$W(A) := dA$$

Second, given any arrow $(\psi, G)$ in $\mathcal{A}$, we define

$$W(\psi, G) = \psi$$
We need to verify that $W$ is indeed a functor. Given a pair of objects $A$ and $A'$ in $A$ such that

$$A' = \psi^* A + G$$  \hspace{1cm} (12.16)

then they are mapped by $W$ to, respectively

$$F = dA$$
$$F' = dA'$$
$$= d(\psi^* A + G)$$

Since $\psi$ is an isometry, $d(\psi^* A) = \psi^* (dA)$; and since $G$ is exact, $dG = 0$. Hence,

$$F' = \psi^* (dA)$$
$$= \psi^* F$$

Hence, if $(\psi, G)$ is an arrow from $A$ to $A'$, then $W(\psi, G) = \psi$ is an arrow from $W(A)$ to $W(A')$. It is immediate from the definition of $W$ that $W(\text{Id}_M, 0) = \text{Id}_M$ and that $W(\Phi \circ \Psi) = W(\Phi) \circ W(\Psi)$; so $W$ is a functor.

It remains to show that $W$ is full, faithful, and essentially surjective. First, let $F$ be any object in $\mathcal{F}$. By Stokes’ Theorem (given that $dF = 0$ is the first of Maxwell’s equations), there exists some $A$ on $M$ such that $F = dA = W(A)$. So $W$ is surjective, and hence essentially surjective.

Second, consider any pair of arrows $(\psi, G)$ and $(\chi, K)$ from $A$ to $A'$. If $W(\psi, G) = W(\chi, K)$, then $\psi = \chi$. Since $A' = \psi^* A + G = \chi^* A + K$, it follows that $\psi^* A + G = \chi^* A + K$, and hence that $G = K$. So $(\psi, G) = (\chi, K)$, and hence, $W$ is faithful.

Finally, consider any arrow $\psi: W(A) \to W(A')$ in $\mathcal{F}$. By definition, this is an isometry such that

$$\psi^* (dA) = dA'$$  \hspace{1cm} (12.17)

Consider the 1-form $G := A' - \psi^* A$. $G$ is closed, i.e. $dG = 0$:

$$dG = d(A' - \psi^* A)$$
$$= dA' - \psi^* (dA)$$
$$= 0$$

where we have used equation (12.17) and the fact that $\psi$ is an isometry. Since $M$ is
contractible, it follows that \( G \) is exact. Since
\[
A' = \psi^* A + G
\] (12.18)
it follows that \((\psi, G)\) is an arrow in \( \mathcal{A} \), and evidently \( W(\psi, G) = \psi \). Thus, \( W \) is full. 

This result demonstrates one of the most useful features of categorical equivalence: it can elicit the sense in which two theories can be regarded as equivalent even if an individual model in one theory corresponds to a set of models in the other—provided that the models in that set are equivalent to one another (formally, that they are regarded as isomorphic in the ambient category). However, there are some limitations to the above.

One obvious limitation is that it will not hold if we expand our category to include non-contractible spaces: for, as discussed in Chapter 9, there are pairs of non-gauge-equivalent potentials \( A, A' \) (over such spaces) that give rise to the same field \( F \). The functor \( W \) (or rather, the extension of \( W \) to this enlarged category) would therefore map \( A \) and \( A' \) to \( F \); and clearly, \( W_{AA'} \) would not be surjective, since there are no arrows between \( A \) and \( A' \), but there is (at least) the identity arrow between \( F \) and itself. That said, this seems like a feature rather than a bug: as we have already discussed, the fact that gauge-equivalence is a stricter criterion than field-equivalence seems to indicate that the potentials theory should not be considered equivalent to the fields theory over non-contractible spaces, even if gauge-equivalent models are taken to be physically equivalent.

A more troubling limitation is that Proposition 2 seems to depend quite sensitively on the question of how we represent gauge transformations. In the above, we took each gauge transformation to be specified by an exact 1-form \( G \); i.e. by a 1-form such that for some scalar field \( \lambda \), \( G = d\lambda \). What if we instead take a gauge transformation to be specified by the scalar field \( \lambda \) itself? In particular, recall that when we come to couple the Maxwell theory to the quantum theory of a charged particle, a gauge transformation will be specified by such a scalar field \( \lambda \)—it is just that the action of this gauge transformation on \( A \) is given by \( d\lambda \).

Thus, let us define a category \( \mathcal{A}' \). In this category, each object is a four-potential \( A \) on Minkowski spacetime (as before); but an arrow from \( A \) to \( A' \) is a pair \((\psi, \lambda)\), where \( \psi : M \to M \) is an isometry and \( \lambda : M \to \mathbb{R} \) is a smooth scalar field such that
\[
A' = \psi^* A + d\lambda
\] (12.19)

Assuming, without loss of generality, that there is no isometry \( \psi \) such that \( A' = \psi^* A \).
This is indeed a category, with composition of arrows given by \((\chi, \mu) \circ (\psi, \lambda) = (\chi \circ \psi, \lambda + \mu)\) and the identity arrow on any \(A\) given by \((\Id_M, 0)\). However, it is not equivalent to the category \(\mathcal{A}\).

**Proposition 13.** There is no full functor from \(\mathcal{A}\) to \(\mathcal{A}'\).

**Proof.** There exist solutions to Maxwell’s equations that lack any non-trivial isometries; let \(\mathbf{A}\) be such a solution. Then the only arrow from \(\mathbf{A}\) to itself in \(\mathcal{A}\) is the identity arrow, i.e. the arrow specified by \((\Id_M, 0)\). Suppose that \(J\) is a functor from \(\mathcal{A}\) to \(\mathcal{A}'\), and consider \(J(\mathbf{A})\). At a minimum, any pair of the form \((\Id_M, \lambda)\) for \(\lambda\) a constant scalar field is an arrow from \(J(\mathbf{A})\) to itself. But \(J(\Id_M, 0) = (\Id_M, 0)\) (by the fact that \(J\) is a functor); so if \(\lambda \neq 0\), then there is no arrow from \(\mathbf{A}\) to itself that \(J\) sends to \((\Id_M, \lambda)\). So \(J\) is not full.

This suggests that we may be able to think of the two categories \(\mathcal{A}\) and \(\mathcal{A}'\) as encoding different interpretations of the theory of electromagnetic potentials. The difference between these two interpretations, though, is rather subtle. The two interpretations agree on what the models of the theory are, and they agree on the relations of physical equivalence between models; what they disagree on are the relations of physical equivalence between gauge transformations (in a sense, they disagree on the relations of physical equivalence between relations of physical equivalence). According to \(\mathcal{A}'\), two gauge transformations—say, those represented by scalar fields \(\lambda\) and \(\mu\)—are physically inequivalent if \(\lambda \neq \mu\); according to \(\mathcal{A}\), if \(d\lambda = d\mu\), then we should regard these two gauge transformations as equivalent.

This indicates that we should not expect merely moving to a more abstract, category-theoretic perspective will free us from troublesome questions of interpretation. Instead, that perspective gives us new tools with which to articulate those questions. As with most things in life, the category-theoretic resources giveth, by showing how to make precise a sense in which (for example) the field-theoretic and potential-theoretic formulations of electromagnetism can be seen as equivalent (under the right circumstances); and it taketh away, by showing how to draw even more fine-grained distinctions between different potential-theoretic formulations than we could before.