3 The Category of Propositional Theories

One of the primary goals of this book is to provide a formal model of "the universe of all scientific theories." In the twentieth century, mathematics stepped up another level of abstraction, and it began to talk of structured collections of mathematical objects – e.g., the category of groups, topological spaces, manifolds, Hilbert spaces, or sets. This maneuver can be a little bit challenging for foundationally oriented thinkers, viz. philosophers, because we are now asked to consider collections that are bigger than any set. However, mathematicians know very well how to proceed in this manner without falling into contradictions (e.g., by availing themselves of Grothendieck universes).

We want to follow the lead of the mathematicians, but instead of talking about the category of groups, or manifolds, or Hilbert spaces, etc., we want to talk about the **category of all theories**. In the present chapter, we work out one special case: the category of all propositional theories. Of course, this category is too simple to serve as a good model for the category of all scientific theories. However, already for predicate logic, the category of theories becomes extremely complex, almost to the point of mathematical intractability. In subsequent chapters, we will make some headway with that case; for the remainder of this chapter, we restrict ourselves to the propositional case.

After defining the relevant category **Th** of propositional theories, we will show that **Th** is equivalent to the category **Bool** of Boolean algebras. We then prove a version of the famous Stone duality theorem, which shows that **Bool** is dual to a certain category **Stone** of topological spaces. This duality shows that each propositional theory corresponds to a unique topological space, viz. the space of its models, and each translation between theories corresponds to a continuous mapping between their spaces of models.

3.1 Basics

DEFINITION 3.1.1 We let **Th** denote the category whose objects are propositional theories and whose arrows are translations between theories. We say that two translations $f, g : T \rightrightarrows T'$ are equal, written $f \simeq g$, just in case $T' \vdash f(\phi) \leftrightarrow g(\phi)$ for every $\phi \in \mathsf{Sent}(\Sigma)$. (Note well: equality between translations is weaker than set-theoretic equality.)

DEFINITION 3.1.2 We say that a translation $f : T \to T'$ is **conservative** just in case, for any $\phi \in \mathsf{Sent}(\Sigma)$, if $T' \vdash f(\phi)$ then $T \vdash \phi$.

PROPOSITION 3.1.3 *A translation $f : T \to T'$ is conservative if and only if f is a monomorphism in the category* **Th**.

Proof Suppose first that f is conservative, and let $g, h : T'' \to T$ be translations such that $f \circ g = f \circ h$. That is, $T' \vdash fg(\phi) \leftrightarrow fh(\phi)$ for every sentence ϕ of Σ''. Since f is conservative, $T \vdash g(\phi) \leftrightarrow h(\phi)$ for every sentence ϕ of Σ''. Thus, $g = h$, and f is a monomorphism in **Th**.

Conversely, suppose that f is a monomorphism in the category **Th**. Let ϕ be a Σ sentence such that $T' \vdash f(\phi)$. Thus, $T' \vdash f(\phi) \leftrightarrow f(\psi)$, where ψ is any Σ sentence such that $T \vdash \psi$. Now let T'' be the empty theory in signature $\Sigma'' = \{p\}$. Define $g : \Sigma'' \to \mathsf{Sent}(\Sigma)$ by $g(p) = \phi$, and define $h : \Sigma'' \to \mathsf{Sent}(\Sigma)$ by $h(p) = \psi$. It's easy to see then that $f \circ g = f \circ h$. Since f is monic, $g = h$, which means that $T \vdash g(p) \leftrightarrow h(p)$. Therefore, $T \vdash \phi$, and f is conservative. □

DEFINITION 3.1.4 We say that a translation $f : T \to T'$ is **essentially surjective** just in case for any sentence ϕ of Σ', there is a sentence ψ of Σ such that $T' \vdash \phi \leftrightarrow f(\psi)$. (Sometimes we use the abbreviation "eso" for essentially surjective.)

PROPOSITION 3.1.5 *If $f : T \to T'$ is essentially surjective, then f is an epimorphism in* **Th**.

Proof Suppose that $f : T \to T'$ is eso. Let $g, h : T' \rightrightarrows T''$ such that $g \circ f = h \circ f$. Let ϕ be an arbitrary Σ' sentence. Since f is eso, there is a sentence ψ of Σ such that $T' \vdash \phi \leftrightarrow f(\psi)$. But then $T'' \vdash g(\phi) \leftrightarrow h(\phi)$. Since ϕ was arbitrary, $g = h$. Therefore, f is an epimorphism. □

What about the converse of this proposition? Are all epimorphisms in **Th** essentially surjective? The answer is yes, but the result is not easy to prove. We'll prove it later on, by means of the correspondence that we establish between theories, Boolean algebras, and Stone spaces.

PROPOSITION 3.1.6 *Let $f : T \to T'$ be a translation. If f is conservative and essentially surjective, then f is a homotopy equivalence.*

Proof Let $p \in \Sigma'$. Since f is eso, there is some $\phi_p \in \mathsf{Sent}(\Sigma)$ such that $T' \vdash p \leftrightarrow f(\phi_p)$. Define a reconstrual $g : \Sigma' \to \mathsf{Sent}(\Sigma)$ by setting $g(p) = \phi_p$. As usual, g extends naturally to a function from $\mathsf{Sent}(\Sigma')$ to $\mathsf{Sent}(\Sigma)$, and it immediately follows that $T' \vdash \psi \leftrightarrow fg(\psi)$, for every sentence ψ of Σ'.

We claim now that g is a translation from T' to T. Suppose that $T' \vdash \psi$. Since $T' \vdash \psi \leftrightarrow fg(\psi)$, it follows that $T' \vdash fg(\psi)$. Since f is conservative, $T \vdash g(\psi)$. Thus, for all sentences ψ of Σ', if $T' \vdash \psi$ then $T \vdash g(\psi)$, which means that $g : T' \to T$ is a translation. By the previous paragraph, $1_{T'} \simeq fg$.

It remains to show that $1_T \simeq gf$. Let ϕ be an arbitrary sentence of Σ. Since f is conservative, it will suffice to show that $T' \vdash f(\phi) \leftrightarrow fgf(\phi)$. But by the previous paragraph, $T' \vdash \psi \leftrightarrow fg(\psi)$ for all sentences ψ of Σ'. Therefore, $1_T \simeq gf$, and f is a homotopy equivalence. □

Before proceeding, let's remind ourselves of some of the motivations for these technical investigations.

The category **Sets** is, without a doubt, extremely useful. However, a person who is familiar with **Sets** might have developed some intuitions that could be misleading when applied to other categories. For example, in **Sets**, if there are injections $f : X \to Y$ and $g : Y \to X$, then there is a bijection between X and Y. Thus, it's tempting to think, for example, that if there are embeddings $f : T \to T'$ and $g : T' \to T$ of theories, then T and T' are equivalent. (Here an embedding between theories is a monomorphism in **Th**, i.e., a conservative translation.) Similarly, in **Sets**, if there is an injection $f : X \to Y$ and a surjection $g : X \to Y$, then there is a bijection between X and Y. However, in **Th**, the analogous result fails to hold.

TECHNICAL ASIDE 3.1.7 For those familiar with the category **Vect** of vector spaces: **Vect** is similar to **Sets** in that mutually embeddable vector spaces are isomorphic. That is, if $f : V \to W$ and $g : W \to V$ are monomorphisms (i.e., injective linear maps), then V and W have the same dimension and, hence, are isomorphic. The categories **Sets** and **Vect** share in common the feature that the objects can be classified by cardinal numbers. In the case of sets, if $|X| = |Y|$, then $X \cong Y$. In the case of vector spaces, if $\dim(V) = \dim(W)$, then $V \cong W$.

In Exercise 1.4.7, you showed that if $f : T \to T'$ is a translation, and if v is a model of T', then $v \circ f$ is a model of T. Let $M(T)$ be the set of all models of T, and define a function $f^* : M(T') \to M(T)$ by setting $f^*(v) = f \circ v$.

PROPOSITION 3.1.8 *Let $f : T \to T'$ be a translation. If $f^* : M(T') \to M(T)$ is surjective, then f is conservative.*

Proof Suppose that f^* is surjective, and suppose that ϕ is a sentence of Σ such that $T \nvdash \phi$. Then there is a $v \in M(T)$ such that $v(\phi) = 0$. (Here we have invoked the completeness theorem, but we haven't proven it yet. Note that our proof of the completeness theorem, page 79, does not cite this result or any that depend on it.) Since f^* is surjective, there is a $w \in M(T')$ such that $f^*(w) = v$. But then

$$w(f(\phi)) = f^* w(\phi) = v(\phi) = 0,$$

from which it follows that $T' \nvdash f(\phi)$. Therefore, f is conservative. □

Example 3.1.9 Let $\Sigma = \{p_0, p_1, \ldots\}$, and let T be the empty theory in Σ. Let $\Sigma' = \{q_0, q_1, \ldots\}$, and let T' be the theory with axioms $q_0 \to q_i$, for $i = 0, 1, \ldots$. We will show that there are conservative translations $f : T \to T'$ and $g : T' \to T$.

Define $f : \Sigma \to \mathsf{Sent}(\Sigma')$ by $f(p_i) = q_{i+1}$. Since T is the empty theory, f is a translation. Then for any valuation v of Σ', we have

$$f^* v(p_i) = v(f(p_i)) = v(q_{i+1}).$$

Furthermore, for any sequence of zeros and ones, there is a valuation v of Σ' that assigns that sequence to $q_1, q_2, \ldots$. Thus, f^* is surjective, and f is conservative.

Now define $g : \Sigma' \to \mathsf{Sent}(\Sigma)$ by setting $g(q_i) = p_0 \vee p_i$. Since $T \vdash p_0 \vee p_0 \to p_0 \vee p_i$, it follows that g is a translation. Furthermore, for any valuation v of Σ, we have

$$g^* v(q_i) = v(g(q_i)) = v(p_0 \vee p_i).$$

Recall that $M(T')$ splits into two parts: (1) a singleton set containing the valuation z where $z(q_i) = 1$ for all i, and (2) the infinitely many other valuations that assign 0 to q_0. Clearly, $z = g^*v$, where v is any valuation such that $v(p_0) = 1$. Furthermore, for any valuation w of Σ' such that $w(p_0) = 0$, we have $w = g^*v$, where $v(p_i) = w(q_i)$. Therefore, g^* is surjective, and g is conservative. ⌟

EXERCISE 3.1.10 In Example 3.1.9, show that f and g are not essentially surjective.

Example 3.1.11 Let T and T' be as in the previous example. Now we'll show that there are essentially surjective (eso) translations $k : T \to T'$ and $h : T' \to T$. The first is easy: the translation $k(p_i) = q_i$ is obviously eso. For the second, define $h(q_0) = \bot$, where $\bot$ is some contradiction, and define $h(q_i) = p_{i-1}$ for $i > 0$. ⌟

Let's pause to think about some of the questions we might want to ask about theories. We arrange these in roughly decreasing order of technical tractability.

1. Does **Th** have the **Cantor–Bernstein property**? That is, if there are monomorphisms $f : T \to T'$ and $g : T' \to T$, then is there an isomorphism $h : T \to T'$?
2. Is **Th** balanced, in the sense that if $f : T \to T'$ is both a monomorphism and an epimorphism, then f is an isomorphism?
3. If there is both a monomorphism $f : T \to T'$ and an epimorphism $g : T' \to T$, then are T and T' homotopy equivalent?
4. Can an arbitrary theory T be embedded into a theory T_0 that has no axioms? Quine and Goodman (1940) present a proof of this claim – and they argue that it undercuts the analytic-synthetic distinction. They are right about the technical claim (see 3.7.10), but have perhaps misconstrued its philosophical implications.
5. If theories have the same number of models, then are they equivalent? If not, then can we determine whether T and T' are equivalent by inspecting $M(T)$ and $M(T')$?
6. How many theories (up to isomorphism) are there with n models?
7. (Does supervenience imply reduction?) Suppose that the truth value of a sentence ψ **supervenes** on the truth value of some other sentences $\phi_1, \dots, \phi_n$, i.e., for any valuations v, w of the propositional constants occurring in $\phi_1, \dots, \phi_n, \psi$, if $v(\phi_i) = w(\phi_i)$, for $i = 1, \dots, n$, then $v(\psi) = w(\psi)$. Does it follow then that $\vdash \psi \leftrightarrow \theta$, where θ contains only the propositional constants that occur in $\phi_1, \dots, \phi_n$? We will return to this issue in Section 6.7.
8. Suppose that $f : T \to T'$ is conservative. Suppose also that every model of T extends uniquely to a model of T'. Does it follow that $T \cong T'$?
9. Suppose that T and T' are consistent in the sense that there is no sentence θ in $\Sigma \cap \Sigma'$ such that $T \vdash \theta$ and $T' \vdash \neg\theta$. Is there a unified theory T'' that extends both T and T'? (The answer is yes, as shown by **Robinson's theorem**.)
10. What does it mean for one theory to be **reducible** to another? Can we explicate this notion in terms of a certain sort of translation between the relevant theories? Some philosophers have claimed that the reduction relation ought to be treated

semantically, rather than syntactically. In other words, they would have us consider functions from $M(T')$ to $M(T)$, rather than translations from T to T'. In light of the Stone duality theorem proved later in the chapter, it appears that syntactic and semantic approaches are equivalent to each other.

11. Consider various formally definable notions of theoretical equivalence. What are the advantages and disadvantages of the various notions? Is homotopy equivalence too liberal? Is it too conservative?

3.2 Boolean Algebras

DEFINITION 3.2.1 A **Boolean algebra** is a set B together with a unary operation $\neg$, two binary operations $\wedge$ and $\vee$, and designated elements $0 \in B$ and $1 \in B$, which satisfy the following equations:

1. Top and Bottom
$$a \wedge 1 = a \vee 0 = a$$
2. Idempotence
$$a \wedge a = a \vee a = a$$
3. De Morgan's Rules
$$\neg(a \wedge b) = \neg a \vee \neg b, \quad \neg(a \vee b) = \neg a \wedge \neg b$$
4. Commutativity
$$a \wedge b = b \wedge a, \quad a \vee b = b \vee a$$
5. Associativity
$$(a \wedge b) \wedge c = a \wedge (b \wedge c), \quad (a \vee b) \vee c = a \vee (b \vee c)$$
6. Distribution
$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c), \quad a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$$
7. Excluded Middle
$$a \wedge \neg a = 0, \quad a \vee \neg a = 1$$

Here we are implicitly universally quantifying over a, b, c.

Example 3.2.2 Let 2 denote the Boolean algebra of subsets of a one-point set. Note that 2 looks just like the truth-value set Ω. Indeed, Ω is equipped with operations $\wedge, \vee$, and $\neg$ that make it into a Boolean algebra. ⌟

Example 3.2.3 Let $\Sigma = \{p\}$. Define an equivalence relation $\simeq$ on sentences of Σ by $\phi \simeq \psi$ just in case $\vdash \phi \leftrightarrow \psi$. If we let F denote the set of equivalence classes, then it's not hard to see that F has four elements: $0, 1, [p], [\neg p]$. Define $[\phi] \wedge [\psi] = [\phi \wedge \psi]$, where the $\wedge$ on the right is the propositional connective, and the $\wedge$ on the left is a newly defined binary function on F. Perform a similar construction for the other logical connectives. Then F is a Boolean algebra. ⌟

We now derive some basic consequences from the axioms for Boolean algebras. The first two results are called the **absorption laws**.

1. $a \wedge (a \vee b) = a$
$$a \wedge (a \vee b) = (a \vee 0) \wedge (a \vee b) = a \vee (0 \wedge b) = a \vee 0 = a.$$

2. $a \vee (a \wedge b) = a$
$$a \vee (a \wedge b) = (a \wedge 1) \vee (a \wedge b) = a \wedge (1 \vee b) = a \wedge 1 = a.$$

3. $a \vee 1 = 1$
$$a \vee 1 = a \vee (a \vee \neg a) = a \vee \neg a = 1.$$

4. $a \wedge 0 = 0$
$$a \wedge 0 = a \wedge (a \wedge \neg a) = a \wedge \neg a = 0.$$

DEFINITION 3.2.4 If B is a Boolean algebra and $a, b \in B$, we write $a \leq b$ when $a \wedge b = a$.

Since $a \wedge 1 = a$, it follows that $a \leq 1$, for all $a \in B$. Since $a \wedge 0 = 0$, it follows that $0 \leq a$, for all $a \in B$. Now we will show that $\leq$ is a partial order, i.e., reflexive, transitive, and asymmetric.

PROPOSITION 3.2.5 *The relation $\leq$ on a Boolean algebra B is a partial order.*

Proof (Reflexive) Since $a \wedge a = a$, it follows that $a \leq a$.
(Transitive) Suppose that $a \wedge b = a$ and $b \wedge c = b$. Then

$$a \wedge c = (a \wedge b) \wedge c = a \wedge (b \wedge c) = a \wedge b = a,$$

which means that $a \leq c$.
(Asymmetric) Suppose that $a \wedge b = a$ and $b \wedge a = b$. By commutativity of $\wedge$, it follows that $a = b$. □

We now show how $\leq$ interacts with $\wedge, \vee$, and $\neg$. In particular, we show that if $\leq$ is thought of as implication, then $\wedge$ behaves like conjunction, $\vee$ behaves like disjunction, $\neg$ behaves like negation, 1 behaves like a tautology, and 0 behaves like a contradiction.

PROPOSITION 3.2.6 *$c \leq a \wedge b$ iff $c \leq a$ and $c \leq b$.*

Proof Since $a \wedge (a \wedge b) = a \wedge b$, it follows that $a \wedge b \leq a$. By similar reasoning, $a \wedge b \leq b$. Thus, if $c \leq a \wedge b$, then transitivity of $\leq$ entails that both $c \leq a$ and $c \leq b$.
Now suppose that $c \leq a$ and $c \leq b$. That is, $c \wedge a = c$ and $c \wedge b = c$. Then $c \wedge (a \wedge b) = (c \wedge a) \wedge (c \wedge b) = c \wedge c = c$. Therefore, $c \leq a \wedge b$. □

Notice that $\leq$ and $\wedge$ interact precisely as implication and conjunction interact in propositional logic. The elimination rule says that $a \wedge b$ implies a and b. Hence, if c implies $a \wedge b$, then c implies a and b. The introduction rule says that a and b imply $a \wedge b$. Hence, if c implies a and b, then c implies $a \wedge b$.

PROPOSITION 3.2.7 *$a \leq c$ and $b \leq c$ iff $a \vee b \leq c$*

Proof Suppose first that $a \leq c$ and $b \leq c$. Then

$$(a \vee b) \wedge c = (a \wedge c) \vee (b \wedge c) = a \vee b.$$

Therefore, $a \vee b \leq c$.

Suppose now that $a \vee b \leq c$. By the absorption law, $a \wedge (a \vee b) = a$, which implies that $a \leq a \vee b$. By transitivity, $a \leq c$. Similarly, $b \leq a \vee b$, and by transitivity, $b \leq c$. □

Now we show that the connectives $\wedge$ and $\vee$ are monotonic.

PROPOSITION 3.2.8 *If $a \leq b$, then $a \wedge c \leq b \wedge c$, for any $c \in B$.*

Proof

$$(a \wedge c) \wedge (b \wedge c) = (a \wedge b) \wedge c = a \wedge c.$$

□

PROPOSITION 3.2.9 *If $a \leq b$, then $a \vee c \leq b \vee c$, for any $c \in B$.*

Proof

$$(a \vee c) \wedge (b \vee c) = (a \wedge b) \vee c = a \vee c.$$

□

PROPOSITION 3.2.10 *If $a \wedge b = a$ and $a \vee b = a$, then $a = b$.*

Proof $a \wedge b = a$ means that $a \leq b$. We now claim that $a \vee b = a$ iff $b \wedge a = b$ iff $b \leq a$. Indeed, if $a \vee b = a$, then

$$b \wedge a = b \wedge (a \vee b) = (0 \vee b) \wedge (a \vee b) = (0 \wedge a) \vee b = b.$$

Conversely, if $b \wedge a = b$, then

$$a \vee b = a \vee (a \wedge b) = (a \wedge 1) \vee (a \wedge b) = a \wedge (1 \vee b) = a.$$

Thus, if $a \wedge b = a$ and $a \vee b = a$, then $a \leq b$ and $b \leq a$. By asymmetry of $\leq$, it follows that $a = b$. □

We now show that $\neg a$ is the unique complement of a in B.

PROPOSITION 3.2.11 *If $a \wedge b = 0$ and $a \vee b = 1$, then $b = \neg a$.*

Proof Since $b \vee a = 1$, we have

$$b = b \vee 0 = b \vee (a \wedge \neg a) = (b \vee a) \wedge (b \vee \neg a) = b \vee \neg a.$$

Since $b \wedge a = 0$, we also have

$$b = b \wedge 1 = b \wedge (a \vee \neg a) = (b \wedge a) \vee (b \wedge \neg a) = b \wedge \neg a.$$

By the preceding proposition, $b = \neg a$. □

PROPOSITION 3.2.12 $\neg 1 = 0$.

Proof We have $1 \wedge 0 = 0$ and $1 \vee 0 = 1$. By the preceding proposition, $0 = \neg 1$. □

PROPOSITION 3.2.13 *If $a \leq b$, then $\neg b \leq \neg a$.*

Proof Suppose that $a \leq b$, which means that $a \wedge b = a$, and, equivalently, $a \vee b = b$. Thus, $\neg a \wedge \neg b = \neg(a \vee b) = \neg b$, which means that $\neg b \leq \neg a$. □

PROPOSITION 3.2.14 $\neg\neg a = a$.

Proof We have $\neg a \vee \neg\neg a = 1$ and $\neg a \wedge \neg\neg a = 1$. By Proposition 3.2.11, it follows that $\neg\neg a = a$. □

DEFINITION 3.2.15 Let A and B be Boolean algebras. A **homomorphism** is a map $\phi : A \to B$ such that $\phi(0) = 0$, $\phi(1) = 1$, and for all $a, b \in A$, $\phi(\neg a) = \neg\phi(a)$, $\phi(a \wedge b) = \phi(a) \wedge \phi(b)$ and $\phi(a \vee b) = \phi(a) \vee \phi(b)$.

It is easy to see that if $\phi : A \to B$ and $\psi : B \to C$ are homomorphisms, then $\psi \circ \phi : A \to C$ is also a homomorphism. Moreover, $1_A : A \to A$ is a homomorphism, and composition of homomorphisms is associative.

DEFINITION 3.2.16 We let **Bool** denote the category whose objects are Boolean algebras and whose arrows are homomorphisms of Boolean algebras.

Since **Bool** is a category, we have notions of **monomorphisms**, **epimorphisms**, **isomorphisms**, etc. Once again, it is easy to see that an injective homomorphism is a monomorphism and a surjective homomorphism is an epimorphism.

PROPOSITION 3.2.17 *Monomorphisms in* **Bool** *are injective.*

Proof Let $f : A \to B$ be a monomorphism, and let $a, b \in A$. Let F denote the Boolean algebra with four elements, and let p denote one of the two elements in F that is neither 0 nor 1. Define $\hat{a} : F \to A$ by $\hat{a}(p) = a$, and define $\hat{b} : F \to A$ by $\hat{b}(p) = b$. It is easy to see that $\hat{a}$ and $\hat{b}$ are uniquely defined by these conditions, and that they are Boolean homomorphisms. Suppose now that $f(a) = f(b)$. Then $f\hat{a} = f\hat{b}$, and, since f is a monomorphism, $\hat{a} = \hat{b}$, and, therefore, $a = b$. Therefore, f is injective. □

It is also true that epimorphisms in **Bool** are surjective. However, proving that fact is no easy task. We will return to it later in the chapter.

PROPOSITION 3.2.18 *If $f : A \to B$ is a homomorphism of Boolean algebras, then $a \leq b$ only if $f(a) \leq f(b)$.*

Proof $a \leq b$ means that $a \wedge b = a$. Thus,

$$f(a) \wedge f(b) = f(a \wedge b) = f(a),$$

which means that $f(a) \leq f(b)$. □

DEFINITION 3.2.19 A homomorphism $\phi : B \to 2$ is called a **state** of B.

3.3 Equivalent Categories

We now have two categories on the table: the category **Th** of theories and the category **Bool** of Boolean algebras. Our next goal is to show that these categories are **structurally identical**. But what do we mean by this? What we mean is that they are **equivalent categories**. In order to explain what that means, we need a few more definitions.

DEFINITION 3.3.1 Suppose that $\mathbf{C}$ and $\mathbf{D}$ are categories. We let $\mathbf{C}_0$ denote the objects of $\mathbf{C}$, and we let $\mathbf{C}_1$ denote the arrows of $\mathbf{C}$. A (covariant) **functor** $F : \mathbf{C} \to \mathbf{D}$ consists of a pair of maps: $F_0 : \mathbf{C}_0 \to \mathbf{D}_0$, and $F_1 : \mathbf{C}_1 \to \mathbf{D}_1$ with the following properties:

1. F_0 and F_1 are compatible in the sense that if $f : X \to Y$ in $\mathbf{C}$, then $F_1(f) : F_0(X) \to F_0(Y)$ in $\mathbf{D}$.
2. F_1 preserves identities and composition in the following sense: $F_1(1_X) = 1_{F_0(X)}$, and $F_1(g \circ f) = F_1(g) \circ F_1(f)$.

When no confusion can result, we simply use F in place of F_0 and F_1.

NOTE 3.3.2 There is also a notion of a **contravariant functor**, where F_1 reverses the direction of arrows: if $f : X \to Y$ in $\mathbf{C}$, then $F_1(f) : F_0(Y) \to F_0(X)$ in $\mathbf{D}$. Contravariant functors will be especially useful for examining the relation between a theory and its set of models. We've already seen that a translation $f : T \to T'$ induces a function $f^* : M(T') \to M(T)$. In Section 3.7, we will see that $f \mapsto f^*$ is part of a contravariant functor.

Example 3.3.3 For any category $\mathbf{C}$, there is a functor $1_{\mathbf{C}}$ that acts as the identity on both objects and arrows. That is, for any object X of $\mathbf{C}$, $1_{\mathbf{C}}(X) = X$. And for any arrow f of $\mathbf{C}$, $1_{\mathbf{C}}(f) = f$. ⌟

DEFINITION 3.3.4 Let $F : \mathbf{C} \to \mathbf{D}$ and $G : \mathbf{C} \to \mathbf{D}$ be functors. A **natural transformation** $\eta : F \Rightarrow G$ consists of a family $\{\eta_X : F(X) \to G(X) \mid X \in \mathbf{C}_0\}$ of arrows in $\mathbf{D}$, such that for any arrow $f : X \to Y$ in $\mathbf{C}$, the following diagram commutes:

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

DEFINITION 3.3.5 A natural transformation $\eta : F \Rightarrow G$ is said to be a **natural isomorphism** just in case each arrow $\eta_X : F(X) \to G(X)$ is an isomorphism. In this case, we write $F \cong G$.

DEFINITION 3.3.6 Let $F : \mathbf{C} \to \mathbf{D}$ and $G : \mathbf{D} \to \mathbf{C}$ be functors. We say that F and G are a **categorical equivalence** just in case $GF \cong 1_{\mathbf{C}}$ and $FG \cong 1_{\mathbf{D}}$.

3.4 Propositional Theories Are Boolean Algebras

In this section, we show that there is a one-to-one correspondence between theories (in propositional logic) and Boolean algebras. We first need some preliminaries.

DEFINITION 3.4.1 Let Σ be a propositional signature (i.e., a set), let B be a Boolean algebra, and let $f : \Sigma \to B$ be an arbitrary function. (Here we use $\cap, \cup$ and $-$ for the Boolean operations in order to avoid confusion with the logical connectives $\wedge, \vee$ and $\neg$.) Then f naturally extends to a map $f : \mathsf{Sent}(\Sigma) \to B$ as follows:

1. $f(\phi \wedge \psi) = f(\phi) \cap f(\psi)$
2. $f(\phi \vee \psi) = f(\phi) \cup f(\psi)$
3. $f(\neg\phi) = -f(\phi)$.

Now let T be a theory in Σ. We say that f is an **interpretation** of T in B just in case: for all sentences ϕ, if $T \vdash \phi$ then $f(\phi) = 1$.

DEFINITION 3.4.2 Let $f : T \to B$ be an interpretation. We say that

1. f is **conservative** just in case: for all sentences ϕ, if $f(\phi) = 1$ then $T \vdash \phi$.
2. f **surjective** just in case: for each $a \in B$, there is a $\phi \in \mathsf{Sent}(\Sigma)$ such that $f(\phi) = a$.

LEMMA 3.4.3 *Let $f : T \to B$ be an interpretation. Then the following are equivalent:*

1. *f is conservative.*
2. *For any $\phi, \psi \in \mathsf{Sent}(\Sigma)$, if $f(\phi) = f(\psi)$ then $T \vdash \phi \leftrightarrow \psi$.*

Proof Note first that $f(\phi) = f(\psi)$ if and only if $f(\phi \leftrightarrow \psi) = 1$. Suppose then that f is conservative. If $f(\phi) = f(\psi)$, then $f(\phi \leftrightarrow \psi) = 1$, and hence $T \vdash \phi \leftrightarrow \psi$. Suppose now that (2) holds. If $f(\phi) = 1$, then $f(\phi) = f(\phi \vee \neg\phi)$, and hence $T \vdash (\phi \vee \neg\phi) \leftrightarrow \phi$. Therefore, $T \vdash \phi$, and f is conservative. □

LEMMA 3.4.4 *If $f : T \to B$ is an interpretation, and $g : B \to A$ is a homomorphism, then $g \circ f$ is an interpretation.*

Proof This is almost obvious. □

LEMMA 3.4.5 *If $f : T \to B$ is an interpretation, and $g : T' \to T$ is a translation, then $f \circ g : T' \to B$ is an interpretation.*

Proof This is almost obvious. □

LEMMA 3.4.6 *Suppose that T is a theory, and $e : T \to B$ is a surjective interpretation. If $f, g : B \rightrightarrows A$ are homomorphisms such that $fe = ge$, then $f = g$.*

Proof Suppose that $fe = ge$, and let $a \in B$. Since e is surjective, there is a $\phi \in \mathsf{Sent}(\Sigma)$ such that $e(\phi) = a$. Thus, $f(a) = fe(\phi) = ge(\phi) = g(a)$. Since a was arbitrary, $f = g$. □

Let T' and T be theories, and let $f, g : T' \rightrightarrows T$ be translations. Recall that we defined identity between translations as follows: $f = g$ if and only if $T \vdash f(\phi) \leftrightarrow g(\phi)$ for all $\phi \in \mathsf{Sent}(\Sigma')$.

LEMMA 3.4.7 *Suppose that $m : T \to B$ is a conservative interpretation. If $f, g : T' \rightrightarrows T$ are translations such that $mf = mg$, then $f = g$.*

Proof Let $\phi \in \mathsf{Sent}(\Sigma')$, where Σ' is the signature of T'. Then $mf(\phi) = mg(\phi)$. Since m is conservative, $T \vdash f(\phi) \leftrightarrow g(\phi)$. Since this holds for all sentences, it follows that $f = g$. □

PROPOSITION 3.4.8 *For each theory T, there is a Boolean algebra $L(T)$ and a conservative, surjective interpretation $i_T : T \to L(T)$ such that for any Boolean algebra B and interpretation $f : T \to B$, there is a unique homomorphism $\overline{f} : L(T) \to B$ such that $\overline{f} i_T = f$.*

$$\begin{array}{ccc} T & \xrightarrow{i_T} & L(T) \\ & {\scriptstyle f} \searrow & \downarrow {\scriptstyle \overline{f}} \\ & & B \end{array}$$

We define an equivalence relation $\equiv$ on the sentences of Σ:

$$\phi \equiv \psi \quad \text{iff} \quad T \vDash \phi \leftrightarrow \psi,$$

and we let

$$E_\phi := \{\psi \mid \phi \equiv \psi\}.$$

Finally, let

$$L(T) := \{E_\phi \mid \phi \in \mathsf{Sent}(\Sigma)\}.$$

We now equip $L(T)$ with the structure of a Boolean algebra. To this end, we need the following facts, which correspond to easy proofs in propositional logic.

FACT 3.4.9 *If $E_\phi = E_{\phi'}$ and $E_\psi = E_{\psi'}$, then:*

1. $E_{\phi \wedge \psi} = E_{\phi' \wedge \psi'}$
2. $E_{\phi \vee \psi} = E_{\phi' \vee \psi'}$
3. $E_{\neg \phi} = E_{\neg \phi'}$.

We then define a unary operation $-$ on $L(T)$ by

$$-E_\phi := E_{\neg \phi},$$

and we define two binary operations on $L(T)$ by

$$E_\phi \cap E_\psi := E_{\phi \wedge \psi}, \qquad E_\phi \cup E_\psi := E_{\phi \vee \psi}.$$

Finally, let ϕ be an arbitrary Σ sentence, and let $0 = E_{\phi \wedge \neg \phi}$ and $1 = E_{\phi \vee \neg \phi}$. The proof that $\langle L(T), \cap, \cup, -, 0, 1 \rangle$ is a Boolean algebra requires a series of straightforward verifications. For example, let's show that $1 \cap E_\psi = E_\psi$, for all sentences ψ. Recall that $1 = E_{\phi \vee \neg\phi}$ for some arbitrarily chosen sentence ϕ. Thus,

$$1 \cap E_\psi \;=\; E_{\phi\vee\neg\phi} \cap E_\psi \;=\; E_{(\phi\vee\neg\phi)\wedge\psi}.$$

Moreover, $T \vdash \psi \leftrightarrow ((\phi \vee \neg\phi) \wedge \psi)$, from which it follows that $E_{(\phi\vee\neg\phi)\wedge\psi} = E_\psi$. Therefore, $1 \cap E_\psi = E_\psi$.

Consider now the function $i_T : \Sigma \to L(T)$ given by $i_T(\phi) = E_\phi$, and its natural extension to $\mathsf{Sent}(\Sigma)$. A quick inductive argument, using the definition of the Boolean operations on $L(T)$, shows that $i_T(\phi) = E_\phi$ for all $\phi \in \mathsf{Sent}(\Sigma)$. The following shows that i_T is a conservative interpretation of T in $L(T)$.

PROPOSITION 3.4.10 *$T \vdash \phi$ if and only if $i_T(\phi) = 1$.*

Proof $T \vdash \phi$ iff $T \vdash (\psi \vee \neg\psi) \leftrightarrow \phi$ iff $i_T(\phi) = E_\phi = E_{\psi\vee\neg\psi} = 1$. □

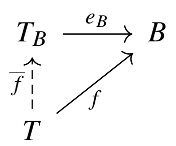
Since $i_T(\phi) = E_\phi$, the interpretation i_T is also surjective.

PROPOSITION 3.4.11 *Let B be a Boolean algebra, and let $f : T \to B$ be an interpretation. Then there is a unique homomorphism $\overline{f} : L(T) \to B$ such that $\overline{f} i_T = f$.*

Proof If $E_\phi = E_\psi$, then $T \vdash \phi \leftrightarrow \psi$, and so $f(\phi) = f(\psi)$. Thus, we may define $\overline{f}(E_\phi) = f(\phi)$. It is straightforward to verify that $\overline{f}$ is a Boolean homomorphism, and it is clearly unique. □

DEFINITION 3.4.12 The Boolean algebra $L(T)$ is called the **Lindenbaum algebra** of T.

PROPOSITION 3.4.13 *Let B be a Boolean algebra. There is a theory T_B and a conservative, surjective interpretation $e_B : T_B \to B$ such that for any theory T and interpretation $f : T \to B$, there is a unique interpretation $\overline{f} : T \to T_B$ such that $e_B \overline{f} = f$.*



Proof Let $\Sigma_B = B$ be a signature. (Recall that a propositional signature is just a set where each element represents an elementary proposition.) We define $e_B : \Sigma_B \to B$ as the identity and use the symbol e_B also for its extension to $\mathsf{Sent}(\Sigma_B)$. We define a theory T_B on Σ_B by $T_B \vdash \phi$ if and only if $e_B(\phi) = 1$. Thus, $e_B : T_B \to B$ is automatically a conservative interpretation of T_B in B.

Now let T be some theory in signature Σ, and let $f : T \to B$ be an interpretation. Since $\Sigma_B = B$, f automatically gives rise to a reconstrual $f : \Sigma \to \Sigma_B$, which we will rename $\overline{f}$ for clarity. And since e_B is just the identity on $B = \Sigma_B$, we have $f = e_B \overline{f}$.

Finally, to see that $\overline{f} : T \to T_B$ is a translation, suppose that $T \vdash \phi$. Since f is an interpretation of T_B, $f(\phi) = 1$, which means that $e_B(\overline{f}(\phi)) = 1$. Since e_B is conservative, $T_B \vdash \overline{f}(\phi)$. Therefore, $\overline{f}$ is a translation. □

We have shown that each propositional theory T corresponds to a Boolean algebra $L(T)$ and each Boolean algebra B corresponds to a propositional theory T_B. We will now show that these correspondences are functorial. First we show that a morphism $f : B \to A$ in **Bool** naturally gives rise to a morphism $T(f) : T_B \to T_A$ in **Th**. Indeed, consider the following diagram:

$$\begin{array}{ccc} T_B & \xrightarrow{T(f)} & T_A \\ \downarrow{\scriptstyle e_B} & & \downarrow{\scriptstyle e_A} \\ B & \xrightarrow{f} & A \end{array}$$

Since fe_B is an interpretation of T_B in A, Prop. 3.4.13 entails that there is a unique translation $T(f) : T_B \to T_A$ such that $e_A T(f) = fe_B$. The uniqueness clause also entails that T commutes with composition of morphisms, and maps identity morphisms to identity morphisms. Thus, T : **Bool** $\to$ **Th** is a functor.

Let's consider this translation $T(f) : T_B \to T_A$ more concretely. First of all, recall that translations from T_B to T_A are actually equivalence classes of maps from Σ_B to $\mathsf{Sent}(\Sigma_A)$. Thus, there's no sense to the question, "which function is $T(f)$?" However, there's a natural choice of a representative function. Indeed, consider f itself as a function from $\Sigma_B = B$ to $\Sigma_A = A$. Then, for $x \in \Sigma_B = B$, we have

$$(e_A \circ T(f))(x) = e_A(f(x)) = f(x) = f(e_B(x)),$$

since e_A is the identity on Σ_A, and e_B is the identity on Σ_B. In other words, $T(f)$ is the equivalence class of f itself. [But recall that translations, while initially defined on the signature Σ_B, extend naturally to all elements of $\mathsf{Sent}(\Sigma_B)$. From this point of view, $T(f)$ has a larger domain than f.]

A similar construction can be used to define the functor L : **Th** $\to$ **Bool**. In particular, let $f : T \to T'$ be a morphism in **Th**, and consider the following diagram:

$$\begin{array}{ccc} T & \xrightarrow{f} & T' \\ \downarrow{\scriptstyle i_T} & & \downarrow{\scriptstyle i_{T'}} \\ L(T) & \xrightarrow{L(f)} & L(T') \end{array}$$

Since $i_{T'}f$ is an interpretation of T in $L(T')$, Prop. 3.4.8 entails that there is a unique homomorphism $L(f) : L(T) \to L(T')$ such that $L(f)i_T = i_{T'}f$.

More explicitly,

$$L(f)(E_\phi) \;=\; L(f)(i_T(\phi)) \;=\; i_{T'}f(\phi) \;=\; E_{f(\phi)}.$$

Recall, however, that identity of arrows in **Th** is *not* identity of the corresponding functions, in the set-theoretic sense. Rather, $f \simeq g$ just in case $T' \vdash f(\phi) \leftrightarrow g(\phi)$, for all $\phi \in \mathsf{Sent}(\Sigma)$. Thus, we must verify that if $f \simeq g$ in **Th**, then $L(f) = L(g)$. Indeed, since $i_{T'}$ is an interpretation of T', we have $i_{T'}(f(\phi)) = i_{T'}(g(\phi))$; and since the diagram above commutes, $L(f) \circ i_T = L(g) \circ i_T$. Since i_T is surjective, $L(f) = L(g)$. Thus, $f \simeq g$ only if $L(f) = L(g)$. Finally, the uniqueness clause in Prop. 3.4.8

entails that L commutes with composition and maps identities to identities. Therefore, $L : \mathbf{Th} \to \mathbf{Bool}$ is a functor.

We will soon show that the functor $L : \mathbf{Th} \to \mathbf{Bool}$ is an equivalence of categories, from which it follows that L preserves all categorically definable properties. For example, a translation $f : T \to T'$ is monic if and only if $L(f) : L(T) \to L(T')$ is monic, etc. However, it may be illuminating to prove some such facts directly.

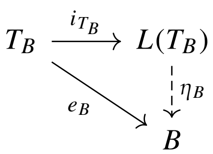
PROPOSITION 3.4.14 *Let $f : T \to T'$ be a translation. Then f is conservative if and only if $L(f)$ is injective.*

Proof Suppose first that f is conservative. Let $E_\phi, E_\psi \in L(T)$ such that $L(f)(E_\phi) = L(f)(E_\psi)$. Using the definition of $L(f)$, we have $E_{f(\phi)} = E_{f(\psi)}$, which means that $T' \vdash f(\phi) \leftrightarrow f(\psi)$. Since f is conservative, $T \vdash \phi \leftrightarrow \psi$, from which $E_\phi = E_\psi$. Therefore, $L(f)$ is injective.

Suppose now that $L(f)$ is injective. Let ϕ be a Σ sentence such that $T' \vdash f(\phi)$. Since $f(\top) = \top$, we have $T' \vdash f(\top) \leftrightarrow f(\phi)$, which means that $L(f)(E_\top) = L(f)(E_\phi)$. Since $L(f)$ is injective, $E_\top = E_\phi$, from which $T \vdash \phi$. Therefore, f is conservative. □

PROPOSITION 3.4.15 *For any Boolean algebra B, there is a natural isomorphism $\eta_B : B \to L(T_B)$.*

Proof Let $e_B : T_B \to B$ be the interpretation from Prop. 3.4.13, and let $i_{T_B} : T_B \to L(T_B)$ be the interpretation from Prop. 3.4.8. Consider the following diagram:



By Prop. 3.4.8, there is a unique homomorphism $\eta_B : L(T_B) \to B$ such that $e_B = \eta_B i_{T_B}$. Since e_B is the identity on Σ_B,

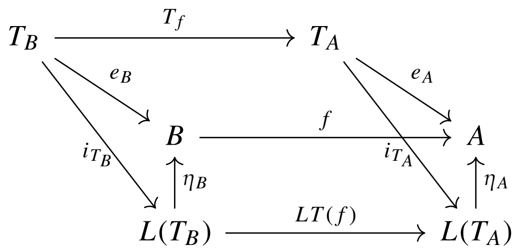
$$\eta_B(E_x) = \eta_B i_{T_B}(x) = e_B(x) = x,$$

for any $x \in B$. Thus, if η_B has an inverse, it must be given by the map $x \mapsto E_x$. We claim that this map is a Boolean homomorphism. To see this, recall that $\Sigma_B = B$. Moreover, for $x, y \in B$, the Boolean meet $x \cap y$ is again an element of B, hence an element of the signature Σ_B. By the definition of T_B, we have $T_B \vdash (x \cap y) \leftrightarrow (x \wedge y)$, where the $\wedge$ symbol on the right is conjunction in $\mathsf{Sent}(\Sigma_B)$. Thus,

$$E_{x \cap y} = E_{x \wedge y} = E_x \cap E_y.$$

A similar argument shows that $E_{-x} = -E_x$. Therefore, $x \mapsto E_x$ is a Boolean homomorphism, and η_B is an isomorphism.

It remains to show that η_B is natural in B. Consider the following diagram:



The top square commutes by the definition of the functor T. The triangles on the left and right commute by the definition of η. And the outmost square commutes by the definition of the functor L. Thus we have

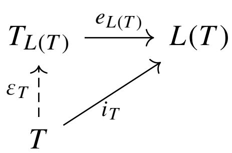
$$\begin{aligned} f \circ \eta_B \circ i_{T_B} &= f \circ e_B \\ &= e_A \circ T_f \\ &= \eta_A \circ i_{T_A} \circ T_f \\ &= \eta_A \circ LT(f) \circ i_{T_B}. \end{aligned}$$

Since i_{T_B} is surjective, it follows that $f \circ \eta_B = \eta_A \circ LT(f)$, and, therefore, η is a natural transformation. □

DISCUSSION 3.4.16 Consider the algebra $L(T_B)$, which we have just proved is isomorphic to B. This result is hardly surprising. For any $x, y \in \Sigma_B$, we have $T_B \vdash x \leftrightarrow y$ if and only if $x = e_B(x) = e_B(y) = y$. Thus, the equivalence class E_x contains x and no other element from Σ_B. (That's why $\eta_B(E_x) = x$ makes sense.) We also know that for every $\phi \in \mathsf{Sent}(\Sigma_B)$, there is an $x \in \Sigma_B = B$ such that $T_B \vdash x \leftrightarrow \phi$. In particular, $T_B \vdash e_B(\phi) \leftrightarrow \phi$. Thus, $E_\phi = E_x$, and there is a natural bijection between elements of $L(T_B)$ and elements of B.

PROPOSITION 3.4.17 *For any theory T, there is a natural isomorphism $\varepsilon_T : T \to T_{L(T)}$.*

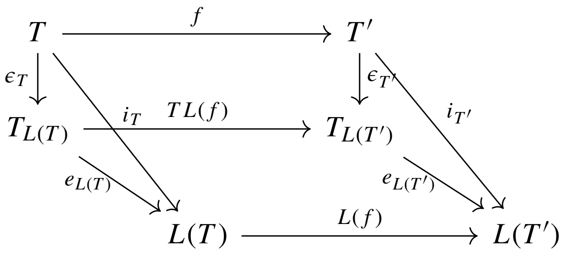
Proof Consider the following diagram:



By Prop. 3.4.13, there is a unique interpretation $\varepsilon_T : T \to T_{L(T)}$ such that $e_{L(T)}\varepsilon_T = i_T$. We claim that ε_T is an isomorphism. To see that ϵ_T is conservative, suppose that $T_{L(T)} \vdash \epsilon_T(\phi)$. Since $e_{L(T)}$ is an interpretation, $e_{L(T)}\epsilon_T(\phi) = 1$ and hence $i_T(\phi) = 1$. Since i_T is conservative, $T \vdash \phi$. Therefore ϵ_T is conservative.

To see that ϵ_T is essentially surjective, suppose that $\psi \in \mathsf{Sent}(\Sigma_{L(T)})$. Since i_T is surjective, there is a $\phi \in \mathsf{Sent}(\Sigma)$ such that $i_T(\phi) = e_{L(T)}(\psi)$. Thus, $e_{L(T)}(\epsilon_T(\phi)) = e_{L(T)}(\psi)$. Since $e_{L(T)}$ is conservative, $T_{L(T)} \vdash \epsilon_T(\phi) \leftrightarrow \psi$. Therefore, ϵ_T is essentially surjective.

It remains to show that ϵ_T is natural in T. Consider the following diagram:



The triangles on the left and the right commute by the definition of ϵ. The top square commutes by the definition of L, and the bottom square commutes by the definition of T. Thus, we have

$$\begin{aligned} e_{L(T')} \circ \epsilon_{T'} \circ f &= i_{T'} \circ f \\ &= L(f) \circ i_T \\ &= L(f) \circ e_{L(T)} \circ \epsilon_T \\ &= e_{L(T')} \circ TL(f) \circ \epsilon_T . \end{aligned}$$

Since $e_{L(T')}$ is conservative, $\epsilon_{T'} \circ f = TL(f) \circ \epsilon_T$. Therefore, ϵ_T is natural in T. □

DISCUSSION 3.4.18 Recall that ϵ_T doesn't denote a unique function; it denotes an equivalence class of functions. One representative of this equivalence class is the function $\epsilon_T : \Sigma \to \Sigma_{L(T)}$ given by $\epsilon_T(p) = E_p$. In this case, a straightforward inductive argument shows that $T_{L(T)} \vdash E_\phi \leftrightarrow \epsilon_T(\phi)$, for all $\phi \in \mathsf{Sent}(\Sigma)$.

We know that ϵ_T has an inverse, which itself is an equivalence class of functions from $\Sigma_{L(T)}$ to $\mathsf{Sent}(\Sigma)$. We can define a representative f of this equivalence class by choosing, for each $E \in \Sigma_{L(T)} = L(T)$, some $\phi \in E$, and setting $f(E) = \phi$. Another straightforward argument shows that if we made a different set of choices, the resulting function f' would be equivalent to f – i.e., it would correspond to the same translation from $T_{L(T)}$ to T.

Since there are natural isomorphisms $\epsilon : 1_{\mathbf{Th}} \Rightarrow TL$ and $\eta : 1_{\mathbf{Bool}} \Rightarrow LT$, we have the following result:

Lindenbaum Theorem

The categories **Th** and **Bool** are equivalent.

3.5 Boolean Algebras Again

The Lindenbaum theorem would deliver everything we wanted – if we had a perfectly clear understanding of the category **Bool**. However, there remain questions about **Bool**.

For example, are all epimorphisms in **Bool** surjections? In order to shed even more light on **Bool**, and hence on **Th**, we will show that **Bool** is dual to a certain category of topological spaces. This famous result is called the **Stone duality theorem**. But in order to prove it, we need to collect a few more facts about Boolean algebras.

DEFINITION 3.5.1 Let B be a Boolean algebra. A subset $F \subseteq B$ is said to be a **filter** just in case

1. If $a, b \in F$, then $a \wedge b \in F$.
2. If $a \in F$ and $a \leq b$, then $b \in F$.

If, in addition, $F \neq B$, then we say that F is a **proper filter**. We say that F is an **ultrafilter** just in case F is maximal among proper filters – i.e., if $F \subseteq F'$ where F' is a proper filter, then $F = F'$.

DISCUSSION 3.5.2 Consider the Boolean algebra B as a theory. Then a filter $F \subseteq B$ can be thought of as supplying an update of information. The first condition says that if we learn a and b, then we've learned $a \wedge b$. The second condition says that if we learn a, and $a \leq b$, then we've learned b. In particular, an ultrafilter supplies maximal information.

EXERCISE 3.5.3 Let F be a filter. Show that F is proper if and only if $0 \notin F$.

DEFINITION 3.5.4 Let $F \subseteq B$ be a filter, and let $a \in B$. We say that a is **compatible** with F just in case $a \wedge x \neq 0$ for all $x \in F$.

LEMMA 3.5.5 *Let $F \subseteq B$ be a proper filter, and let $a \in B$. Then either a or $\neg a$ is compatible with F.*

Proof Suppose for reductio ad absurdum that neither a nor $\neg a$ is compatible with F. That is, there is an $x \in F$ such that $x \wedge a = 0$, and there is a $y \in F$ such that $y \wedge \neg a = 0$. Then

$$x \wedge y = (x \wedge y) \wedge (a \vee \neg a) = (x \wedge y \wedge a) \vee (x \wedge y \wedge \neg a) = 0.$$

Since $x, y \in F$, it follows that $0 = x \wedge y \in F$, contradicting the assumption that F is proper. Therefore, either a or $\neg a$ is compatible with F. □

PROPOSITION 3.5.6 *Let F be a proper filter on B. Then the following are equivalent:*

1. *F is an ultrafilter.*
2. *For all $a \in B$, either $a \in F$ or $\neg a \in F$.*
3. *For all $a, b \in B$, if $a \vee b \in F$, then either $a \in F$ or $b \in F$.*

Proof ($1 \Rightarrow 2$) Suppose that F is an ultrafilter. By Lemma 3.5.5, either a or $\neg a$ is compatible with F. Suppose first that a is compatible with F. Then the set

$$F' = \{y : x \wedge a \leq y, \text{ some } x \in F\},$$

is a proper filter that contains F and a. Since F is an ultrafilter, $F' = F$, and hence $a \in F$. By symmetry, if $\neg a$ is compatible with F, then $\neg a \in F$.

(2 $\Rightarrow$ 3) Suppose that $a \vee b \in F$. By 2, either $a \in F$ or $\neg a \in F$. If $\neg a \in F$, then $\neg a \wedge (a \vee b) \in F$. But $\neg a \wedge (a \vee b) \leq b$, and so $b \in F$.

(3 $\Rightarrow$ 1) Suppose that F' is a filter that contains F, and let $a \in F' - F$. Since $a \vee \neg a = 1 \in F$, it follows from (3) that $\neg a \in F$. But then $0 = a \wedge \neg a \in F'$; that is, $F' = B$. Therefore F is an ultrafilter. □

PROPOSITION 3.5.7 *There is a bijective correspondence between ultrafilters in B and homomorphisms from B into 2. In particular, for any homomorphism $f : B \to 2$, the subset $f^{-1}(1)$ is an ultrafilter in B.*

Proof Let U be an ultrafilter on B. Define $f : B \to 2$ by setting $f(a) = 1$ iff $a \in U$. Then

$$\begin{aligned} f(a \wedge b) = 1 \quad &\text{iff} \quad a \wedge b \in U \\ &\text{iff} \quad a \in U \text{ and } b \in U \\ &\text{iff} \quad f(a) = 1 \text{ and } f(b) = 1. \end{aligned}$$

Furthermore,

$$\begin{aligned} f(\neg a) = 1 \quad &\text{iff} \quad \neg a \in U \\ &\text{iff} \quad a \notin U \\ &\text{iff} \quad f(a) = 0. \end{aligned}$$

Therefore, f is a homomorphism.

Now suppose that $f : B \to 2$ is a homomorphism, and let $U = f^{-1}(1)$. Since $f(a) = 1$ and $f(b) = 1$ only if $f(a \wedge b) = 1$, it follows that U is closed under conjunction. Since $a \leq b$ only if $f(a) \leq f(b)$, it follows that U is closed under implication. Finally, since $f(a) = 0$ iff $f(\neg a) = 1$, it follows that $a \notin U$ iff $\neg a \in U$. □

DEFINITION 3.5.8 For $a, b \in B$, define

$$a \to b \ := \ \neg a \vee b,$$

and define

$$a \leftrightarrow b \ := \ (a \to b) \wedge (b \to a).$$

It's straightforward to check that $\to$ behaves like the conditional from propositional logic. The next lemma gives a Boolean algebra version of modus ponens.

LEMMA 3.5.9 *Let F be a filter. If $a \to b \in F$ and $a \in F$, then $b \in F$.*

Proof Suppose that $\neg a \vee b = a \to b \in F$ and $a \in F$. We then compute

$$b = b \vee 0 = b \vee (a \wedge \neg a) = (a \vee b) \wedge (\neg a \vee b).$$

Since $a \in F$ and $a \leq a \vee b$, we have $a \vee b \in F$. Since F is a filter, $b \in F$. □

EXERCISE 3.5.10

1. Let B be a Boolean algebra, and let $a, b, c \in B$. Show that the following hold:
   (a) $(a \to b) = 1$ iff $a \leq b$
   (b) $(a \wedge b) \leq c$ iff $a \leq (b \to c)$
   (c) $a \wedge (a \to b) \leq b$
   (d) $(a \leftrightarrow b) = (b \leftrightarrow a)$
   (e) $(a \leftrightarrow a) = 1$
   (f) $(a \leftrightarrow 1) = a$
2. Let $\mathscr{P}N$ be the powerset of the natural numbers, and let $\mathscr{U}$ be an ultrafilter on $\mathscr{P}N$. Show that if $\mathscr{U}$ contains a finite set F, then $\mathscr{U}$ contains a singleton set.

DEFINITION 3.5.11 Let B be a Boolean algebra, and let R be an equivalence relation on the underlying set of B. We say that R is a **congruence** just in case R is compatible with the operations on B in the following sense: if aRa' and bRb', then $(a \wedge b)R(a' \wedge b')$, and $(a \vee b)R(a' \vee b')$, and $(\neg a)R(\neg a')$.

In a category **C** with limits (products, equalizers, pullbacks, etc.), it's possible to formulate the notion of an equivalence relation in **C**. Thus, in **Bool**, an equivalence relation R on B is a subalgebra R of $B \times B$ that satisfies the appropriate analogues of reflexivity, symmetry, and transitivity. Since R is a subalgebra of $B \times B$, it follows in particular that if $\langle a, b \rangle \in R$, and $\langle a', b' \rangle \in R$, then $\langle a \wedge a', b \wedge b' \rangle \in R$. Continuing this reasoning, it's not difficult to see that congruences, as defined earlier, are precisely the equivalence relations in the category **Bool** of Boolean algebras. Thus, in the remainder of this chapter, when we speak of an equivalence relation on a Boolean algebra B, we mean an equivalence relation in **Bool** – in other words, a congruence. (To be clear, not every equivalence relation on the set B is an equivalence relation on the Boolean algebra B.)

Now suppose that **C** is a category in which equivalence relations are definable, and let $p_0, p_1 : R \rightrightarrows B$ be an equivalence relation. (Here p_0 and p_1 are the projections of R, considered as a subobject of $B \times B$.) Then we can ask, do these two maps p_0 and p_1 have a coequalizer? That is, is there an object B/R, and a map $q : B \to B/R$, with the relevant universal property? In the case of **Bool**, a coequalizer can be constructed directly. We merely note that the Boolean operations on B can be used to induce Boolean operations on the set B/R of equivalence classes.

DEFINITION 3.5.12 (Quotient algebra) Suppose that R is an equivalence relation on B. For each $a \in B$, let E_a denote its equivalence class, and let $B/R = \{E_a \mid a \in B\}$. We then define $E_a \wedge E_b = E_{a \wedge b}$, and similarly for $E_a \vee E_b$ and $\neg E_a$. Since R is a congruence (i.e., an equivalence relation on **Bool**), these operations are well defined. It then follows immediately that B/R is a Boolean algebra, and the quotient map $q : B \to B/R$ is a surjective Boolean homomorphism.

LEMMA 3.5.13 *Let $R \subseteq B \times B$ be an equivalence relation. Then $q : B \to B/R$ is the coequalizer of the projection maps $p_0 : R \to B$ and $p_1 : R \to B$. In particular, q is a regular epimorphism.*

Proof It is obvious that $qp_0 = qp_1$. Now suppose that A is another Boolean algebra and $f : B \to A$ such that $fp_0 = fp_1$. Define $g : B/R \to A$ by setting $g(E_x) = f(x)$. Since $fp_0 = fp_1$, g is well defined. Furthermore,

$$g(E_x \wedge E_y) = g(E_{x \wedge y}) = f(x \wedge y) = f(x) \wedge f(y) = g(E_x) \wedge g(E_y).$$

Similarly, $g(\neg E_x) = \neg g(E_x)$. Therefore, g is a Boolean homomorphism. Since q is an epimorphism, g is the unique homomorphism such that $gq = f$. Therefore, $q : B \to B/R$ is the coequalizer of p_0 and p_1. □

The category **Bool** has further useful structure: there is a one-to-one correspondence between equivalence relations and filters.

LEMMA 3.5.14 *Suppose that $R \subseteq B \times B$ is an equivalence relation. Let $F = \{a \in B \mid aR1\}$. Then F is a filter, and $R = \{\langle a,b \rangle \in B \times B \mid a \leftrightarrow b \in F\}$.*

Proof Suppose that $a, b \in F$. That is, $aR1$ and $bR1$. Since R is a congruence, $(a \wedge b)R(1 \wedge 1)$ and, therefore, $(a \wedge b)R1$. That is, $a \wedge b \in F$. Now suppose that x is an arbitrary element of B such that $a \leq x$. That is, $x \vee a = x$. Since R is a congruence, $(x \vee a)R(x \vee 1)$ and so $(x \vee a)R1$, from which it follows that $xR1$. Therefore, $x \in F$, and F is a filter.

Now suppose that aRb. Since R is reflexive, $(a \vee \neg a)R1$, and, thus, $(b \vee \neg a)R1$. Similarly, $(a \vee \neg b)R1$, and, therefore, $(a \leftrightarrow b)R1$. That is, $a \leftrightarrow b \in F$. □

LEMMA 3.5.15 *Suppose that F is a filter on B. Let $R = \{\langle a,b \rangle \in B \times B \mid a \leftrightarrow b \in F\}$. Then R is an equivalence relation, and $F = \{a \in B \mid aR1\}$.*

Proof Showing that R is an equivalence relation requires several straightforward verifications. For example, $a \leftrightarrow a = 1$, and $1 \in F$; therefore, aRa. We leave the remaining verifications to the reader.

Now suppose that $a \in F$. Since $a = (a \leftrightarrow 1)$, it follows that $a \leftrightarrow 1 \in F$, which means that $aR1$. □

DEFINITION 3.5.16 (Quotient algebra) Let F be a filter on B. Given the correspondence between filters and equivalence relations, we write B/F for the corresponding algebra of equivalence classes.

PROPOSITION 3.5.17 *Let F be a proper filter on B. Then B/F is a two-element Boolean algebra if and only if F is an ultrafilter.*

Proof Suppose first that $B/F \cong 2$. That is, for any $a \in B$, either $a \leftrightarrow 1 \in F$ or $a \leftrightarrow 0 \in F$. But $a \leftrightarrow 1 = a$ and $a \leftrightarrow 0 = \neg a$. Therefore, either $a \in F$ or $\neg a \in F$, and F is an ultrafilter.

Suppose now that F is an ultrafilter. Then for any $a \in B$, either $a \in F$ or $\neg a \in F$. In the former case, $a \leftrightarrow 1 \in F$. In the latter case, $a \leftrightarrow 0 \in F$. Therefore, $B/F \cong 2$. □

EXERCISE 3.5.18 (This exercise presupposes knowledge of measure theory.) Let Σ be the Boolean algebra of Borel subsets of $[0,1]$, and let μ be Lebesgue measure on $[0,1]$. Let $\mathscr{F} = \{S \in \Sigma \mid \mu(S) = 1\}$. Show that $\mathscr{F}$ is a filter, and describe the equivalence relation on Σ corresponding to $\mathscr{F}$.

According to our motivating analogy, a Boolean algebra B is like a theory, and a homorphism $\phi : B \to 2$ is like a model of this theory. We say that the algebra B is **syntactically consistent** just in case $0 \neq 1$. (In fact, we defined Boolean algebras so as to require syntactic consistency.) We say that the algebra B is **semantically consistent** just in case there is a homomorphism $\phi : B \to 2$. Then semantic consistency clearly implies syntactic consistency. But does syntactic consistency imply semantic consistency?

It's at this point that we have to invoke a powerful theorem – or, more accurately, a powerful set-theoretic axiom. In short, if we use the axiom of choice, or some equivalent such as Zorn's lemma, then we can prove that every syntactically consistent Boolean algebra is semantically consistent. However, we do not actually need the full power of the Axiom of Choice. As set-theorists know, the Boolean ultrafilter axiom (UF for short) is strictly weaker than the Axiom of Choice.

PROPOSITION 3.5.19 *The following are equivalent:*

1. ***Boolean ultrafilter axiom (UF):*** *For any Boolean algebra B, there is a homomorphism $f : B \to 2$.*
2. *For any Boolean algebra B, and proper filter $F \subseteq B$, there is a homomorphism $f : B \to 2$ such that $f(a) = 1$ when $a \in F$.*
3. *For any Boolean algebra B, if $a,b \in B$ such that $a \neq b$, then there is a homomorphism $f : B \to 2$ such that $f(a) \neq f(b)$.*
4. *For any Boolean algebra B, if $\phi(a) = 1$ for all $\phi : B \to 2$, then $a = 1$.*
5. *For any two Boolean algebras A, B, and homomorphisms $f, g : A \rightrightarrows B$, if $\phi f = \phi g$ for all $\phi : B \to 2$, then $f = g$.*

Proof ($1 \Rightarrow 2$) Suppose that F is a proper filter in B. Then there is a homomorphism $q : B \to B/F$ such that $q(a) = 1$ for all $a \in F$. By UF, there is a homomorphism $\phi : B/F \to 2$. Therefore, $\phi \circ q : B \to 2$ is a homomorphism such that $(\phi \circ q)(a) = 1$ for all $a \in F$.

($1 \Rightarrow 3$) Suppose that $a, b \in B$ with $a \neq b$. Then either $\neg a \wedge b \neq 0$ or $a \wedge \neg b \neq 0$. Without loss of generality, we assume that $\neg a \wedge b \neq 0$. In this case, the filter F generated by $\neg a \wedge b$ is proper. By UF, there is a homomorphism $\phi : B \to 2$ such that $\phi(x) = 1$ when $x \in F$. In particular, $\phi(\neg a \wedge b) = 1$. But then $\phi(a) = 0$ and $\phi(b) = 1$.

($2 \Rightarrow 4$) Suppose that $\phi(a) = 1$ for all $\phi : B \to 2$. Now let F be the filter generated by $\neg a$. If F is proper, then by (2), there is a $\phi : B \to 2$ such that $\phi(\neg a) = 1$, a contradiction. Thus, $F = B$, which implies that $\neg a = 0$ and $a = 1$.

($4 \Rightarrow 5$) Let $f, g : A \to B$ be homomorphisms, and suppose that for all $\phi : B \to 2$, $\phi f = \phi g$. That is, for each $a \in A$, $\phi(f(a)) = \phi(g(a))$. But then $\phi(f(a) \leftrightarrow g(a)) = 1$ for all $\phi : B \to 2$. By (4), $f(a) \leftrightarrow g(a) = 1$ and, therefore, $f(a) = g(a)$.

($5 \Rightarrow 3$) Let B be a Boolean algebra, and $a, b \in B$. Suppose that $\phi(a) = \phi(b)$ for all $\phi : B \to 2$. Let F be the four element Boolean algebra, with generator p. Then there is a homomorphism $\hat{a} : F \to B$ such that $\hat{a}(p) = a$, and a homomorphism $\hat{b} : F \to B$ such that $\hat{b}(p) = b$. Thus, $\phi\hat{a} = \phi\hat{b}$ for all $\phi : B \to 2$. By (5), $\hat{a} = \hat{b}$, and therefore $a = b$.

(3 ⇒ 1) Let B be an arbitrary Boolean algebra. Since $0 \neq 1$, (3) implies that there is a homomorphism $\phi : B \to 2$. □

We are finally in a position to prove the completeness of the propositional calculus. The following result assumes the Boolean ultrafilter axiom (UF).

Completeness Theorem

If $T \vDash \phi$, then $T \vdash \phi$.

Proof Suppose that $T \nvdash \phi$. Then in the Lindenbaum algebra $L(T)$, we have $E_\phi \neq 1$. In this case, there is a homomorphism $h : L(T) \to 2$ such that $h(E_\phi) = 0$. Hence, $h \circ i_T$ is a model of T such that $(h \circ i_T)(\phi) = h(E_\phi) = 0$. Therefore, $T \nvDash \phi$. □

EXERCISE 3.5.20 Let $\mathscr{P}N$ be the powerset of the natural numbers. We say that a subset E of N is **cofinite** just in case $N \backslash E$ is finite. Let $\mathscr{F} \subseteq \mathscr{P}N$ be the set of cofinite subsets of N. Show that $\mathscr{F}$ is a filter, and show that there are infinitely many ultrafilters containing $\mathscr{F}$.

3.6 Stone Spaces

If we're going to undertake an exact study of "possible worlds," then we need to make a proposal about what structure this space carries. But what do I mean here by "structure"? Isn't the collection of possible worlds just a bare set? Let me give you a couple of reasons why it's better to think of possible worlds as forming a **topological space**.

Suppose that there are infinitely many possible worlds, which we represent by elements of a set X. As philosophers are wont to do, we then represent **propositions** by subsets of X. But should we think that all $2^{|X|}$ subsets of X correspond to genuine propositions? What would warrant such a claim?

There is another reason to worry about this approach. For a person with training in set theory, it is not difficult to build a collection $C_1, C_2, \dots$ of subsets of X with the following features: (1) each C_i is nonempty, (2) $C_{i+1} \subseteq C_i$ for all i, and (3) $\bigcap_i C_i$ is empty. Intuitively speaking, $\{C_i \mid i \in \mathbb{N}\}$ is a family of propositions that are individually consistent (since nonempty) and that are becoming more and more specific, and yet there is no world in X that makes all C_i true. Why not? It seems that X is missing some worlds! Indeed, here's a description of a new world w that does not belong to X: for each proposition ϕ, let ϕ be true in w if and only if $\phi \cap C_i$ is nonempty for all i. It's not difficult to see that w is, in fact, a truth valuation on the set of all propositions – i.e., it is a possible world. But w is not represented by a point in X. What we have here is a mismatch between the set X of worlds and the set of propositions describing these worlds.

The idea behind logical topology is that not all subsets of X correspond to propositions. A designation of a topology on X is tantamount to saying which subsets of X correspond to propositions. However, the original motivation for the study of topology comes from geometry (and analysis), not from logic. Recall high school mathematics, where you learned that a continuous function is one where you don't have to lift your pencil from the paper in order to draw the graph. If your high school class was really good, or if you studied calculus in college, then you will have learned that there is a more rigorous definition of a continuous function – a definition involving epsilons and deltas. In the early twentieth century, it was realized that the essence of continuity is even more abstract than epsilons and deltas would suggest: all we need is a notion of nearness of points, which we can capture in terms of a notion of a neighborhood of a point. The idea then is that a function $f : X \to Y$ is continuous at a point x just in case for any neighborhood V of $f(x)$, there is some neighborhood U of x such that $f(U) \subseteq V$. Intuitively speaking, f preserves closeness of points.

Notice, however, that if X is an arbitrary set, then it's not obvious what "closeness" means. To be able to talk about closeness of points in X, we need specify which subsets of X count as the neighborhoods of points. Thus, a **topology** on X is a set of subsets of X that satisfies certain conditions.

DEFINITION 3.6.1 A **topological space** is a set X and a family $\mathscr{F}$ of subsets of X satisfying the following conditions:

1. $\emptyset \in \mathscr{F}$ and $X \in \mathscr{F}$.
2. If $U, V \in \mathscr{F}$ then $U \cap V \in \mathscr{F}$.
3. If $\mathscr{F}_0$ is a subfamily of $\mathscr{F}$, then $\bigcup_{U \in \mathscr{F}_0} U \in \mathscr{F}$.

The sets in $\mathscr{F}$ are called **open subsets** of the space $(X, \mathscr{F})$. If $p \in U$ with U an open subset, we say that U is a **neighborhood** of p.

There are many familiar examples of topological spaces. In many cases, however, we only know the open sets indirectly, by means of certain nice open sets. For example, in the case of the real numbers, not every open subset is an interval. However, every open subset is a union of intervals. In that case, we call the open intervals in $\mathbb{R}$ a **basis** for the topology.

PROPOSITION 3.6.2 *Let $\mathscr{B}$ be a family of subsets of X with the property that if $U, V \in \mathscr{B}$, then $U \cap V \in \mathscr{B}$. Then there is a unique smallest topology $\mathscr{F}$ on X containing $\mathscr{B}$.*

Proof Let $\mathscr{F}$ be the collection obtained by taking all unions of sets in $\mathscr{B}$, and then taking finite intersections of the resulting collection. Clearly $\mathscr{F}$ is a topology on X, and any topology on X containing $\mathscr{B}$ also contains $\mathscr{F}$. □

DEFINITION 3.6.3 If $\mathscr{B}$ is a family of subsets of X that is closed under intersection, and if $\mathscr{F}$ is the topology generated by $\mathscr{B}$, then we say that $\mathscr{B}$ is a **basis** for $\mathscr{F}$.

PROPOSITION 3.6.4 *Let $(X, \mathscr{F})$ be a topological space. Let $\mathscr{F}_0$ be a subfamily of $\mathscr{F}$ with the following properties: (1) $\mathscr{F}_0$ is closed under finite intersections, and (2) for* *each $x \in X$ and $U \in \mathscr{F}_0$ with $x \in U$, there is a $V \in \mathscr{F}_0$ such that $x \in V \subseteq U$. Then $\mathscr{F}_0$ is a basis for the topology $\mathscr{F}$.*

Proof We need only show that each $U \in \mathscr{F}$ is a union of elements in $\mathscr{F}_0$. And that follows immediately from the fact that if $x \in U$, then there is $V \in \mathscr{F}_0$ with $x \in V \subseteq U$. □

DEFINITION 3.6.5 Let X be a topological space. A subset C of X is called **closed** just in case $C = X \backslash U$ for some open subset U of X. The intersection of closed sets is closed. Hence, for each subset E of X, there is a unique smallest closed set $\overline{E}$ containing E, namely the intersection of all closed supersets of E. We call $\overline{E}$ the **closure** of E.

PROPOSITION 3.6.6 *Let $p \in X$ and let $S \subseteq X$. Then $p \in \overline{S}$ if and only if every open neighborhood U of p has nonempty intersection with S.*

Proof Exercise. □

DEFINITION 3.6.7 Let S be a subset of X. We say that S is **dense** in X just in case $\overline{S} = X$.

DEFINITION 3.6.8 Let $E \subseteq X$. We say that p is a **limit point** of E just in case for each open neighborhood U of p, $U \cap E$ contains some point besides p. We let E' denote the set of all limit points of E.

LEMMA 3.6.9 $E' \subseteq \overline{E}$.

Proof Let $p \in E'$, and let C be a closed set containing E. If $p \in X \backslash C$, then p is contained in an open set that has empty intersection with E. Thus, $p \in C$. Since C was an arbitrary closed superset of E, it follows that $p \in \overline{E}$. □

PROPOSITION 3.6.10 $\overline{E} = E \cup E'$.

Proof The previous lemma gives $E' \subseteq \overline{E}$. Thus, $E \cup E' \subseteq \overline{E}$.

Suppose now that $p \notin E$ and $p \notin E'$. Then there is an open neighborhood U of p such that $U \cap E$ is empty. Then $E \subseteq X \backslash U$, and since $X \backslash U$ is closed, $\overline{E} \subseteq X \backslash U$. Therefore, $p \notin \overline{E}$. □

DEFINITION 3.6.11 A topological space X is said to be

- T_1, or **Frechet**, just in case all singleton subsets are closed.
- T_2, or **Hausdorff**, just in case, for any $x, y \in X$, if $x \neq y$, then there are disjoint open neighborhoods of x and y.
- T_3, or **regular**, just in case for each $x \in X$, and for each closed $C \subseteq X$ such that $x \notin C$, there are open neighborhoods U of x, and V of C, such that $U \cap V = \emptyset$.
- T_4, or **normal**, just in case any two disjoint closed subsets of X can be separated by disjoint open sets.

Clearly we have the implications

$$(T_1 + T_4) \Rightarrow (T_1 + T_3) \Rightarrow T_2 \Rightarrow T_1.$$

A discrete space satisfies all of the separation axioms. A nontrivial indiscrete space satisfies none of the separation axioms. A useful heuristic here is that the stronger the separation axiom, the closer the space is to discrete. In this book, most of the spaces we consider are very close to discrete (which means that all subsets are open).

EXERCISE 3.6.12

1. Show that X is regular iff for each $x \in X$ and open neighborhood U of x, there is an open neighborhood V of x such that $\overline{V} \subseteq U$.
2. Show that if $E \subseteq F$, then $\overline{E} \subseteq \overline{F}$.
3. Show that $\overline{\overline{E}} = \overline{E}$.
4. Show that the intersection of two topologies is a topology.
5. Show that the infinite distributive law holds:

$$U \cap \left(\bigcup_{i \in I} V_i \right) = \bigcup_{i \in I} (U \cap V_i).$$

DEFINITION 3.6.13 Let $S \subseteq X$. A family $\mathscr{C}$ of open subsets of X is said to **cover** S just in case $S \subseteq \bigcup_{U \in \mathscr{C}} U$. We say that S is **compact** just in case for every open cover $\mathscr{C}$ of S, there is a finite subcollection $\mathscr{C}_0$ of $\mathscr{C}$ that also covers S. We say that the space X is compact just in case it's compact as a subset of itself.

DEFINITION 3.6.14 A collection $\mathscr{C}$ of subsets of X is said to satisfy the **finite intersection property** if for every finite subcollection $C_1, \ldots, C_n$ of $\mathscr{C}$, the intersection $C_1 \cap \cdots \cap C_n$ is nonempty.

DISCUSSION 3.6.15 Suppose that X is the space of possible worlds, so that we can think of subsets of X as propositions. If $A \cap B$ is nonempty, then the propositions A and B are consistent – i.e., there is a world in which they are both true. Thus, a collection $\mathscr{C}$ of propositions has the finite intersection property just in case it is finitely consistent.

Recall that compactness of propositional logic states that if a set $\mathscr{C}$ of propositions is finitely consistent, then $\mathscr{C}$ is consistent. The terminology here is no accident; a topological space is compact just in case finite consistency entails consistency.

PROPOSITION 3.6.16 *A space X is compact if and only if for every collection $\mathscr{C}$ of closed subsets of X, if $\mathscr{C}$ satisfies the finite intersection property, then $\bigcap \mathscr{C}$ is nonempty.*

Proof ($\Rightarrow$) Assume first that X is compact, and let $\mathscr{C}$ be a family of closed subsets of X. We will show that if $\mathscr{C}$ satisfies the finite intersection property, then the intersection of all sets in $\mathscr{C}$ is nonempty. Assume the negation of the consequent, i.e., that $\bigcap_{C \in \mathscr{C}} C$ is empty. Let $\mathscr{C}' = \{C' : C \in \mathscr{C}\}$, where $C' = X \backslash C$ is the complement of C in X. (Warning: this notation can be confusing. Previously we used E' to denote the set of limit points of E. This C' has nothing to do with limit points.) Each C' is open, and

$$\left(\bigcup_{C \in \mathscr{C}} C' \right)' = \bigcap_{C \in \mathscr{C}} C,$$

which is empty. It follows then that $\mathscr{C}'$ is an open cover of X. Since X is compact, there is a finite subcover $\mathscr{C}_0'$ of $\mathscr{C}'$. If we let $\mathscr{C}_0$ be the complements of sets in $\mathscr{C}_0'$, then $\mathscr{C}_0$ is a finite collection of sets in $\mathscr{C}$ whose intersection is empty. Therefore, $\mathscr{C}$ does not satisfy the finite intersection property.

($\Leftarrow$) Assume now that X is not compact. In particular, suppose that $\mathscr{U}$ is an open cover with no finite subcover. Let $\mathscr{C} = \{X \backslash U \mid U \in \mathscr{U}\}$. For any finite subcollection $X \backslash U_1, \dots, X \backslash U_n$ of $\mathscr{C}$, we have

$$U_1 \cup \dots \cup U_n \neq X,$$

and hence

$$(X \backslash U_1) \cap \dots \cap (X \backslash U_n) \neq \emptyset.$$

Thus, $\mathscr{C}$ has the fip. Nonetheless, since $\mathscr{U}$ covers X, the intersection of all sets in $\mathscr{C}$ is empty. □

PROPOSITION 3.6.17 *In a compact space, closed subsets are compact.*

Proof Let $\mathscr{C}$ be an open cover of S, and consider the cover $\mathscr{C}' = \mathscr{C} \cup \{X \backslash S\}$ of X. Since X is compact, there is a finite subcover $\mathscr{C}_0$ of $\mathscr{C}'$. Removing $X \backslash S$ from $\mathscr{C}_0$ gives a finite subcover of the original cover $\mathscr{C}$ of S. □

PROPOSITION 3.6.18 *Suppose that X is compact, and let U be an open set in X. Let $\{F_i\}_{i \in I}$ be a family of closed subsets of X such that $\bigcap_{i \in I} F_i \subseteq U$. Then there is a finite subset J of I such that $\bigcap_{i \in J} F_i \subseteq U$.*

Proof Let $C = X \backslash U$, which is closed. Thus, the hypotheses of the proposition say that the family $\mathscr{C} := \{C\} \cup \{F_i : i \in I\}$ has empty intersection. Since X is compact, $\mathscr{C}$ also fails to have the finite intersection property. That is, there are $i_1, \dots, i_k \in I$ such that $C \cap F_{i_1} \cap \dots \cap F_{i_k} = \emptyset$. Therefore, $F_{i_1} \cap \dots \cap F_{i_k} \subseteq U$. □

PROPOSITION 3.6.19 *If X is compact Hausdorff, then X is regular.*

Proof Let $x \in X$, and let $C \subseteq X$ be closed. For each $y \in C$, let U_y be an open neighborhood of x, and V_y an open neighborhood of y such that $U_y \cap V_y = \emptyset$. The V_y form an open cover of C. Since C is closed and X is compact, C is compact. Hence, there is a finite subcollection $V_{y_1}, \dots, V_{y_n}$ that cover C. But then $U = \cap_{i=1}^n U_{y_i}$ is an open neighborhood of x, and $V = \cup_{i=1}^n V_{y_i}$ is an open neighborhood of C, such that $U \cap V = \emptyset$. Therefore, X is regular. □

PROPOSITION 3.6.20 *In Hausdorff spaces, compact subsets are closed.*

Proof Let p be a point of X that is not in K. Since X is Hausdorff, for each $x \in K$, there are open neighborhoods U_x of x and V_x of p such that $U_x \cap V_x = \emptyset$. The family $\{U_x : x \in K\}$ covers K. Since K is compact, it is covered by a finite subcollection $U_{x_1}, \dots, U_{x_n}$. But then $\cap_{i=1}^n V_{x_i}$ is an open neighborhood of p that is disjoint from K. It follows that $X \backslash K$ is open, and K is closed. □

DEFINITION 3.6.21 Let X, Y be topological spaces. A function $f : X \to Y$ is said to be **continuous** just in case for each open subset U of Y, $f^{-1}(U)$ is an open subset of X.

Example 3.6.22 Let $f : \mathbb{R} \to \mathbb{R}$ be the function that is constantly 0 on $(-\infty, 0)$, and 1 on $[0, \infty)$. Then f is not continuous: $f^{-1}(\frac{1}{2}, \frac{3}{2}) = [0, \infty)$, which is not open. ⌟

In the exercises, you will show that a function f is continuous if and only if $f^{-1}(C)$ is closed whenever C is closed. Thus, in particular, if C is a clopen subset of Y, then $f^{-1}(C)$ is a clopen subset of X.

PROPOSITION 3.6.23 *Let* **Top** *consist of the class of topological spaces and continuous maps between them. For* $X \xrightarrow{f} Y \xrightarrow{g} Z$*, define* $g \circ f$ *to be the composition of* g *and* f*. Then* **Top** *is a category.*

Proof It needs to be confirmed that if f and g are continuous, then $g \circ f$ is continuous. We leave this to the exercises. Since composition is associative, **Top** is a category. □

PROPOSITION 3.6.24 *Suppose that* $f : X \to Y$ *is continuous. If* K *is compact in* X*, then* $f(K)$ *is compact in* Y*.*

Proof Let $\mathscr{G}$ be a collection of open subsets of Y that covers $f(K)$. Let

$$\mathscr{G}' = \{f^{-1}(U) : U \in \mathscr{G}\}.$$

When $\mathscr{G}'$ is an open cover of K. Since K is compact, $\mathscr{G}'$ has a finite subcover $f^{-1}(U_1), \ldots, f^{-1}(U_n)$. But then $U_1, \ldots, U_n$ is a finite subcover of $\mathscr{G}$. □

We remind the reader of the category theoretic definitions:

- f is a **monomorphism** just in case $fh = fk$ implies $h = k$.
- f is an **epimorphism** just in case $hf = kf$ implies $h = k$.
- f is an **isomorphism** just in case there is a $g : Y \to X$ such that $gf = 1_X$ and $fg = 1_Y$.

For historical reasons, isomorphisms in **Top** are usually called **homeomorphisms**. It is easy to show that a continuous map $f : X \to Y$ is monic if and only if f is injective. It is also true that $f : X \to Y$ is epi if and only if f is surjective (but the proof is somewhat subtle). In contrast, a continuous bijection is not necessarily an isomorphism in **Top**. For example, if we let X be a two-element set with the discrete topology, and if we let Y be a two-element set with the indiscrete topology, then any bijection $f : X \to Y$ is continuous but is not an isomorphism.

EXERCISE 3.6.25

1. Show that if f and g are continuous, then $g \circ f$ is continuous.
2. Suppose that $f : X \to Y$ is a surjection. Show that if E is dense in X, then $f(E)$ is dense in Y.
3. Show that $f : X \to Y$ is continuous if and only if $f^{-1}(C)$ is closed whenever C is closed.
4. Let Y be a Hausdorff space, and let $f, g : X \to Y$ be continuous. Show that if f and g agree on a dense subset of X, then $f = g$.

EXERCISE 3.6.26 Show that $f^{-1}(V) \subseteq U$ if and only if $V \subseteq Y \backslash f(X \backslash U)$.

DEFINITION 3.6.27 A continuous mapping $f : X \to Y$ is said to be **closed** just in case for every closed set $C \subseteq X$, the image $f(C)$ is closed in Y. Similarly, $f : X \to Y$ is said to be **open** just in case for every open set $U \subseteq X$, the image $f(U)$ is open in Y.

PROPOSITION 3.6.28 *Let $f : X \to Y$ be continuous. Then the following are equivalent.*

1. *f is closed.*
2. *For every open set $U \subseteq X$, the set $\{y \in Y \mid f^{-1}\{y\} \subseteq U\}$ is open.*
3. *For every $y \in Y$, and every neighborhood U of $f^{-1}\{y\}$, there is a neighborhood V of y such that $f^{-1}(V) \subseteq U$.*

Proof (2 $\Leftrightarrow$ 3) The equivalence of (2) and (3) is straightforward, and we leave its proof as an exercise.

(3 $\Rightarrow$ 1) Suppose that f satisfies condition (3), and let C be a closed subset of X. To show that $f(C)$ is closed, assume that $y \in Y \backslash f(C)$. Then $f^{-1}\{y\} \subseteq X \backslash C$. Since $X \backslash C$ is open, there is a neighborhood V of y such that $f^{-1}(V) \subseteq U$. Then

$$V \subseteq Y \backslash f(X \backslash U) = Y \backslash f(C).$$

Since y was an arbitrary element of $Y \backslash f(C)$, it follows that $Y \backslash f(C)$ is open, and $f(C)$ is closed.

(1 $\Rightarrow$ 3) Suppose that f is closed. Let $y \in Y$, and let U be a neighborhood of $f^{-1}\{y\}$. Then $X \backslash U$ is closed, and $f(X \backslash U)$ is also closed. Let $V = Y \backslash f(X \backslash U)$. Then V is an open neighborhood of y and $f^{-1}(V) \subseteq U$. □

PROPOSITION 3.6.29 *Suppose that X and Y are compact Hausdorff. If $f : X \to Y$ is continuous, then f is a closed map.*

Proof Let B be a closed subset of X. By Proposition 3.6.17, B is compact. By Proposition 3.6.24, $f(B)$ is compact. And by Proposition 3.6.20, $f(B)$ is closed. Therefore, f is a closed map. □

PROPOSITION 3.6.30 *Suppose that X and Y are compact Hausdorff. If $f : X \to Y$ is a continuous bijection, then f is an isomorphism.*

Proof Let $f : X \to Y$ be a continuous bijection. Thus, there is function $g : Y \to X$ such that $gf = 1_X$ and $fg = 1_Y$. We will show that g is continuous. By Proposition 3.6.29, f is closed. Moreover, for any closed subset B of X, we have $g^{-1}(B) = f(B)$. Thus, g^{-1} preserves closed subsets, and hence g is continuous. □

DEFINITION 3.6.31 A topological space X is said to be **totally separated** if for any $x, y \in X$, if $x \neq y$ then there is a closed and open (clopen) subset of X containing x but not y.

DEFINITION 3.6.32 We say that X is a **Stone space** if X is compact and totally separated. We let **Stone** denote the full subcategory of **Top** consisting of Stone spaces.

To say that **Stone** is a full subcategory means that the arrows between two Stone spaces X and Y are just the arrows between X and Y considered as topological spaces – i.e., continuous functions.

NOTE 3.6.33 Let E be a clopen subset of X. Then there is a continuous function $f : X \to \{0,1\}$ such that $f(x) = 1$ for $x \in E$, and $f(x) = 0$ for $x \in X \backslash E$. Here we are considering $\{0,1\}$ with the discrete topology.

PROPOSITION 3.6.34 *Let X and Y be Stone spaces. If $f : X \to Y$ is an epimorphism, then f is surjective.*

Proof Suppose that f is not surjective. Since X is compact, the image $f(X)$ is compact in Y, hence closed. Since f is not surjective, there is a $y \in Y \backslash f(X)$. Since Y is a regular space, there is a clopen neighborhood U of y such that $U \cap f(X) = \emptyset$. Define $g : Y \to \{0,1\}$ to be constantly 0. Define $h : Y \to \{0,1\}$ to be 1 on U, and 0 on $Y \backslash U$. Then $g \circ f = h \circ f$, but $g \neq h$. Therefore, f is not an epimorphism. □

PROPOSITION 3.6.35 *Let X and Y be Stone spaces. If $f : X \to Y$ is both a monomorphism and an epimorphism, then f is an isomorphism.*

Proof By Proposition 3.6.34, f is surjective. Therefore, f is a continuous bijection. By Proposition 3.6.30, f is an isomorphism. □

3.7 Stone Duality

In this section, we show that the category **Bool** is dual to the category **Stone** of Stone spaces. To say that categories are "dual" means that the first is equivalent to the mirror image of the second.

DEFINITION 3.7.1 We say that categories **C** and **D** are **dual** just in case there are contravariant functors $F : \mathbf{C} \to \mathbf{D}$ and $G : \mathbf{D} \to \mathbf{C}$ such that $GF \cong 1_{\mathbf{C}}$ and $FG \cong 1_{\mathbf{D}}$. To see that this definition makes sense, note that if F and G are contravariant functors, then GF and FG are covariant functors. If **C** and **D** are dual, we write $\mathbf{C} \cong \mathbf{D}^{op}$, to indicate that **C** is equivalent to the opposite category of **D** – i.e., the category that has the same objects as **D**, but arrows running in the opposite direction.

The Functor from Bool to Stone

We now define a contravariant functor $S : \mathbf{Bool} \to \mathbf{Stone}$. For reasons that will become clear later, the functor S is sometimes called the **semantic functor**.

Consider the set $\hom(B,2)$ of two-valued homomorphisms of the Boolean algebra B. For each $a \in B$, define

$$C_a = \{\phi \in \hom(B,2) \mid \phi(a) = 1\}.$$

Clearly, the family $\{C_a \mid a \in B\}$ forms a basis for a topology on $\hom(B,2)$. We let $S(B)$ denote the resulting topological space. Note that $S(B)$ has a basis of clopen sets. Thus, if $S(B)$ is compact, then $S(B)$ is a Stone space.

LEMMA 3.7.2 *If B is a Boolean algebra, then $S(B)$ is a Stone space.*

Proof Let $\mathscr{B} = \{C_a \mid a \in B\}$ denote the chosen basis for the topology on $S(B)$. To show that $S(B)$ is compact, it will suffice to show that for any subfamily $\mathscr{C}$ of $\mathscr{B}$, if $\mathscr{C}$ has the finite intersection property, then $\bigcap \mathscr{C}$ is nonempty. Now let F be the set of $b \in B$ such that

$$C_{a_1} \cap \cdots \cap C_{a_n} \subseteq C_b,$$

for some $C_{a_1}, \ldots, C_{a_n} \in \mathscr{C}$. Since $\mathscr{C}$ has the finite intersection property, F is a filter in B. Thus, UF entails that F is contained in an ultrafilter U. This ultrafilter U corresponds to a $\phi : B \to 2$, and we have $\phi(a) = 1$ whenever $C_a \in \mathscr{C}$. In other words, $\phi \in C_a$, whenever $C_a \in \mathscr{C}$. Therefore, $\bigcap \mathscr{C}$ is nonempty, and $S(B)$ is compact. □

Let $f : A \to B$ be a homomorphism, and let $S(f) : S(B) \to S(A)$ be given by $S(f) = \hom(f, 2)$; that is,

$$S(f)(\phi) = \phi \circ f, \qquad \forall \phi \in S(B).$$

We claim now that $S(f)$ is a continuous map. Indeed, for any basic open subset C_a of $S(A)$, we have

$$S(f)^{-1}(C_a) = \{\phi \in S(B) \mid \phi(f(a)) = 1\} = C_{f(a)}. \tag{3.1}$$

It is straightforward to verify that $S(1_A) = 1_{S(A)}$, and that $S(g \circ f) = S(f) \circ S(f)$. Therefore, $S : \mathbf{Bool} \to \mathbf{Stone}$ is a contravariant functor.

The Functor from Stone to Bool

Let X be a Stone space. Then the set $K(X)$ of clopen subsets of X is a Boolean algebra, and is a basis for the topology on X. We now show that K is the object part of a contravariant functor $K : \mathbf{Stone} \to \mathbf{Bool}$. For reasons that will become clear later, K is sometimes called the **syntactic functor**.

Indeed, if X, Y are Stone spaces, and $f : X \to Y$ is continuous, then for each clopen subset U of Y, $f^{-1}(U)$ is a clopen subset of X. Moreover, f^{-1} preserves union, intersection, and complement of subsets; thus $f^{-1} : K(Y) \to K(X)$ is a Boolean homomorphism. We define the mapping K on arrows by $K(f) = f^{-1}$. Obviously, $K(1_X) = 1_{K(X)}$, and $K(g \circ f) = K(f) \circ K(g)$. Therefore, K is a contravariant functor.

Now we will show that KS is naturally isomorphic to the identity on **Bool**, and SK is naturally isomorphic to the identity on **Stone**. For each Boolean algebra B, define $\eta_B : B \to KS(B)$ by

$$\eta_B(a) = C_a = \{\phi \in S(B) \mid \phi(a) = 1\}.$$

LEMMA 3.7.3 *The map $\eta_B : B \to KS(B)$ is an isomorphism of Boolean algebras.*

Proof We first verify that $a \mapsto C_a$ is a Boolean homomorphism. For $a, b \in B$, we have

$$\begin{aligned} C_{a\wedge b} &= \{\phi \mid \phi(a\wedge b)=1\} \\ &= \{\phi \mid \phi(a)=1 \text{ and } \phi(b)=1\} \\ &= C_a \wedge C_b. \end{aligned}$$

A similar calculation shows that $C_{\neg a} = X \backslash C_a$. Therefore, $a \mapsto C_a$ is a Boolean homomorphism.

To show that $a \mapsto C_a$ is injective, it will suffice to show that $C_a = \emptyset$ only if $a = 0$. In other words, it will suffice to show that for each $a \in B$, if $a \neq 0$ then there is some $\phi : B \to 2$ such that $\phi(a) = 1$. Thus, the result follows from UF.

Finally, to see that η_B is surjective, let U be a clopen subset of $S(B)$. Since U is open, $U = \bigcup_{a\in I} C_a$, for some subset I of B. Since U is closed in the compact space $G(B)$, it follows that U is compact. Thus, there is a finite subset F of B such that $U = \bigcup_{a\in F} C_a$. And since $a \mapsto C_a$ is a Boolean homomorphism, $\bigcup_{a\in F} C_a = C_b$, where $b = \bigvee_{a\in F} a$. Therefore, η_B is surjective. □

LEMMA 3.7.4 *The family of maps* $\{\eta_A : A \to KS(A)\}$ *is natural in* A.

Proof Suppose that A and B are Boolean algebras and that $f : A \to B$ is a Boolean homomorphism. Consider the following diagram:

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ \downarrow{\scriptstyle \eta_A} & & \downarrow{\scriptstyle \eta_B} \\ KS(A) & \xrightarrow{KS(f)} & KS(B) \end{array}$$

For $a \in A$, we have $\eta_B(f(a)) = C_{f(a)}$, and $\eta_A(a) = C_a$. Furthermore,

$$KS(f)(C_a) \;=\; S(f)^{-1}(C_a) \;=\; C_{f(a)},$$

by Eqn. 3.1. Therefore, the diagram commutes, and η is a natural transformation. □

Now we define a natural isomorphism $\theta : 1_{\mathbf{S}} \Rightarrow SK$. For a Stone space X, $K(X)$ is the Boolean algebra of clopen subsets of X, and $SK(X)$ is the Stone space of $K(X)$. For each point $\phi \in X$, let $\hat{\phi} : K(X) \to 2$ be defined by

$$\hat{\phi}(C) = \begin{cases} 1 & \phi \in C, \\ 0 & \phi \notin C. \end{cases}$$

It's straightforward to verify that $\hat{\phi}$ is a Boolean homomorphism. We define $\theta_X : X \to SK(X)$ by $\theta_X(\phi) = \hat{\phi}$.

LEMMA 3.7.5 *The map* $\theta_X : X \to SK(X)$ *is a homeomorphism of Stone spaces.*

Proof It will suffice to show that θ_X is bijective and continuous. (Do you remember why? Hint: Stone spaces are compact Hausdorff.) To see that θ_X is injective, suppose that ϕ and ψ are distinct elements of X. Since X is a Stone space, there is a clopen set U of X such that $\phi \in U$ and $\psi \notin U$. But then $\hat{\phi} \neq \hat{\psi}$. Thus, θ_X is injective.

To see that θ_X is surjective, let $h : K(X) \to 2$ be a Boolean homomorphism. Let

$$\mathscr{C} \;=\; \{C \in K(X) \mid h(C) = 1\}.$$

In particular $X \in \mathscr{C}$; and since h is a homomorphism, $\mathscr{C}$ has the finite intersection property. Since X is compact, $\bigcap \mathscr{C}$ is nonempty. Let ϕ be a point in $\bigcap \mathscr{C}$. Then for any $C \in K(X)$, if $h(C) = 1$, then $C \in \mathscr{C}$ and $\phi \in C$, from which it follows that $\hat{\phi}(C) = 1$. Similarly, if $h(C) = 0$ then $X \backslash C \in \mathscr{C}$, and $\hat{\phi}(C) = 0$. Thus, $\theta_X(\phi) = \hat{\phi} = h$, and θ_X is surjective.

To see that θ_X is continuous, note that each basic open subset of $SK(X)$ is of the form

$$\hat{C} \;=\; \{h : K(X) \to 2 \mid h(C) = 1\},$$

for some $C \in K(X)$. Moreover, for any $\phi \in X$, we have $\hat{\phi} \in \hat{C}$ iff $\hat{\phi}(C) = 1$ iff $\phi \in C$. Therefore,

$$\theta_X^{-1}(\hat{C}) \;=\; \{\phi \in X \mid \hat{\phi}(C) = 1\} \;=\; C.$$

Therefore, θ_X is continuous. □

LEMMA 3.7.6 *The family of maps $\{\theta_X : X \to SK(X)\}$ is natural in X.*

Proof Let X, Y be Stone spaces, and let $f : X \to Y$ be continuous. Consider the diagram:

$$\begin{array}{ccc} X & \xrightarrow{\;f\;} & Y \\ \downarrow{\scriptstyle \theta_X} & & \downarrow{\scriptstyle \theta_Y} \\ SK(X) & \xrightarrow{SK(f)} & SK(Y) \end{array}$$

For arbitrary $\phi \in X$, we have $(\theta_Y \circ f)(\phi) = \widehat{f(\phi)}$. Furthermore.

$$SK(f) \;=\; \hom(K(f), 2) \;=\; \hom(f^{-1}, 2),$$

In other words, for a homomorphism $h : K(X) \to 2$, we have

$$SK(f)(h) \;=\; h \circ f^{-1}.$$

In particular, $SK(f)(\hat{\phi}) = \hat{\phi} \circ f^{-1}$. For any $C \in K(Y)$, we have

$$(\hat{\phi} \circ f^{-1})(C) \;=\; \begin{cases} 1 & f(\phi) \in C, \\ 0 & f(\phi) \notin C. \end{cases}$$

That is, $\hat{\phi} \circ f^{-1} = \widehat{f(\phi)}$. Therefore, the diagram commutes, and θ is a natural isomorphism. □

This completes the proof that K and S are quasi-inverse, and yields the famous theorem:

Stone Duality Theorem

The categories **Stone** and **Bool** are dual to each other. In particular, any Boolean algebra B is isomorphic to the field of clopen subsets of its state space $S(B)$.

PROPOSITION 3.7.7 *Let $A \subseteq B$, and let $a \in B$. Then the following are equivalent:*

1. *For any states f and g of B, if $f|_A = g|_A$ then $f(a) = g(a)$.*
2. *If h is a state of A, then any two extensions of h to B agree on a.*
3. *$a \in A$.*

Proof Since every state of A can be extended to a state of B, (1) and (2) are obviously equivalent. Furthermore, (3) obviously implies (1). Thus, we only need to show that (1) implies (3).

Let $m : A \to B$ be the inclusion of A in B, and let $s : S(B) \to S(A)$ be the corresponding surjection of states. We need to show that $C_a = s^{-1}(U)$ for some clopen subset U of $S(A)$.

By (1), for any $x \in S(A)$, either $s^{-1}\{x\} \subseteq C_a$ or $s^{-1}\{x\} \subseteq C_{\neg a}$. By Proposition 3.6.29, s is a closed map. Since C_a is open, Proposition 3.6.28 entails that the sets,

$$U = \{x \in S(B) \mid s^{-1}\{x\} \subseteq C_a\}, \quad \text{and} \quad V = \{x \in S(B) \mid s^{-1}\{x\} \subseteq C_{\neg a}\},$$

are open. Since $U = S(A) \backslash V$, it follows that U is clopen. Finally, it's clear that $s^{-1}(U) = C_a$. □

PROPOSITION 3.7.8 *In* **Bool**, *epimorphisms are surjective.*

Proof Suppose that $f : A \to B$ is not surjective. Then $f(A)$ is a proper subalgebra of B. By Proposition 3.7.7, there are states g, h of B such that $g \neq h$, but $g|_{f(A)} = h|_{f(A)}$. In other words, $g \circ f = h \circ f$, and f is not an epimorphism. □

Combining the previous two theorems, we have the following equivalences:

$$\mathbf{Th} \cong \mathbf{Bool} \cong \mathbf{Stone}^{op}.$$

We will now exploit these equivalences to explore the structure of the category of theories.

PROPOSITION 3.7.9 *Let T be a propositional theory in a countable signature. Then there is a conservative translation $f : T \to T_0$, where T_0 is an empty theory – i.e., a theory with no axioms.*

Proof After proving the above equivalences, we have several ways of seeing why this result is true. In terms of Boolean algebras, the proposition says that every countable Boolean algebra is embeddable into the free Boolean algebra on a countable number of generators (i.e., the Boolean algebra of clopen subsets of the Cantor space). That well-known result follows from the fact that Boolean algebras are always generated by their finite subalgebras. (In categorical terms, every Boolean algebra is a filtered colimit of finite Boolean algebras.)

In terms of Stone spaces, the proposition says that for every separable Stone space Y, there is a continuous surjection $p : X \to Y$, where X is the Cantor space. That fact is well known to topologists. One interesting proof uses the fact that a Stone space Y is *profinite* – i.e., Y is a limit of finite Hausdorff (hence discrete) spaces. One then shows that the Cantor space X has enough surjections onto discrete spaces and lifts these up to a surjection $p : X \to Y$. See, for example, Ribes and Zalesskii (2000). Or, for a more direct argument: each clopen subset U of Y corresponds to a continuous map $p_U : Y \to \{0, 1\}$. There are countably many such clopen subsets of Y. Since $X \simeq \prod_{i \in \mathbb{N}} \{0, 1\}$, these p_U induce a continuous function $p : X \to Y$. Moreover, since every point $y \in Y$ has a neighborhood basis of clopen sets, p is surjective. □

DISCUSSION 3.7.10 (Quine on eliminating posulates) It's no surprise that one can be charitable to a fault. Suppose that I am a theist, and you are an extremely charitable atheist. You are so charitable that you want to affirm the things I say. Here's how you can do it: when I say "God," assume that I really mean "kittens." Then when I say "God exists," you can interpret me to be saying "kittens exist." Then you can smile and say "I completely agree!"

Proposition 3.7.9 provides a general recipe for charitable interpretation. Imagine that I accept a theory T, which might be controversial. Imagine that you, on the other hand, like to play it safe: you only accept tautologies, viz. empty theory T_0. The previous proposition shows that there is a conservative translation $f : T \to T_0$. In other words, you can reinterpret my sentences in such a way that everything I say comes out as true by your lights – i.e., true by logic alone.

Since we're dealing merely with propositional logic, this result might not seem very provocative. However, a directly analogous result – proven by Quine and Goodman (1940); Quine (1964) – was thought to refute the analytic–synthetic distinction that was central to the logical positivist program. Quine's argument runs as follows: suppose that T is intended to represent a contingently true theory, such as (presumably) quantum mechanics or evolutionary biology. By making a series of clever definitions, the sentences of T can be reconstrued as tautologies. That is, any contingently true theory T can be reconstrued so that all of its claims come out as true by definition.

What we see here is an early instance of a strategy that Quine was to use again and again throughout his philosophical career. There is a supposedly important distinction in a theory T. Quine shows that this distinction doesn't survive translation of T into some other theory T_0. This result, Quine claims, shows that the distinction must be rejected.

Whether or not Quine's strategy is generally good, we should be a bit suspicious in the present case. The translation $f : T \to T_0$ is not an *equivalence* of theories – i.e., it does not show that T is equivalent to T_0. Since f is conservative, it does show a sense in which T is *embeddable in* or *reducible to* T_0. But we are left wondering: why should the existence of a formal relation $f : T \to T_0$ undercut the importance of the distinctions that are made within T?

If Proposition 3.7.9 was surprising, then the following result is even more surprising:

PROPOSITION 3.7.11 *Let T be a consistent propositional theory in a countably infinite signature. If T has a finite number of axioms, then T is equivalent to the empty theory T_0.*

Sketch of proof Suppose that T has a finite number of axioms. Without loss of generality, we assume that T has a single axiom ϕ. Let X be the Cantor space – i.e., the Stone space of the empty theory T_0. Let $U_\phi \subseteq X$ be the clopen subset of all models in which

ϕ is true. Then U_ϕ is homeomorphic to the Stone space of T. Assume for the moment any nonempty clopen subset of the Cantor space is homeomorphic to the Cantor space. In that case, U_ϕ is homeomorphic to the Cantor space X; and by Stone duality, T is equivalent to T_0.

We now argue that nonempty clopen subset of the Cantor space is homeomorphic to the Cantor space. (This result admits of several proofs, some more topologically illuminating than the one we give here.) We begin by arguing that if ϕ is a conjunction of literals (atomic or negated atomic sentences), then U_ϕ is homeomorphic to the Cantor space. Indeed, there is a direct proof that the theory $\{\phi\}$ is equivalent to the empty theory; hence, by Stone duality, U_ϕ is homeomorphic to X. Now, an arbitrary clopen subset U of X has the form U_ϕ for some sentence ϕ. We may rewrite ϕ in disjunctive normal form – i.e., as a finite disjunction of conjunctions of literals. Thus, U_ϕ is a disjoint union of $U_{\phi_1}, U_{\phi_2}, \ldots, U_{\phi_n}$. By the previous argument, each U_{ϕ_i} is homeomorphic to X, and a disjoint union of copies of X is also homeomorphic to X. □

The previous proposition might suggest that the notion of equivalence we have adopted (Definition 1.4.6) is too *liberal* – i.e., that it counts too many theories as equivalent. If you think that's the case, we enjoin you to propose another criterion and explore its consequences.

DISCUSSION 3.7.12 The Stone duality theorem suggests that accepting a theory T involves accepting some claims about nearness/similarity relations among possible worlds. One theory T leads to a particular topological structure on the set of possible worlds, and another theory T' leads to a different topological structure on the set of possible worlds. That fact applies not just to propositional theories, but also to real-life scientific theories. For example, when one accepts the general theory of relativity, one doesn't simply believe that our universe is isomorphic to one of its models. Rather, one believes that the situtation we find ourselves in is one among many other situations that obey the laws of this theory. Moreover, some such situations are more similar than others. See Fletcher (2016) for an extended discussion of this example.

3.8 Notes

We have given only the most cursory introduction to the rich mathematical fields of Boolean algebras, topology, and the interactions between them. There is much more to be learned and many good books on these topics. Some of our favorites are the following:

- For more on Boolean algebras, see Sikorski (1969); Dwinger (1971); Koppelberg (1989); Givant and Halmos (2008); Monk (2014),
- There are many good books on topology. We learned originally from Munkres (2000), and our favorites include Engelking (1989) and Willard (1970). The latter is notable for its presentation of the ultrafilter approach to convergence.

- Stone spaces, being a particular kind of topological space, are sometimes mentioned in books about topology. But for a more systematic treatment of Stone spaces, you'll need to consult other resources. For a fully general and categorical treatment of Stone duality, see Johnstone (1986). For briefer and more pedestrian treatments, see Bell and Machover (1977); Halmos and Givant (1998); Cori and Lascar (2000). For a proof that Stone spaces are profinite, see Ribes and Zalesskii (2000).

4 Syntactic Metalogic

First-order logic plays a starring role in our best account of the structure of human knowledge. There is reason to believe that first-order logic is fully sufficient to encode *all* deductively valid reasoning. It was discovered in the early twentieth century that first-order logic is powerful enough to axiomatize many of the theories that mathematicians use, such as number theory, group theory, ring theory, field theory, etc. And although there are other mathematical theories that are overtly second-order (e.g., the theory of topological spaces quantifies over subsets, and not just individual points), nonetheless first-order logic can be used to axiomatize set theory, and any second-order theory can be formalized within set theory. Thus, first-order logic provides an expansive framework in which much, if not all, deductively valid human reasoning can be represented.

In this chapter, we will study the properties of first-order logic, the theories that can be formulated within it, and the relations that hold between them. Let's begin from the concrete – with examples of some theories that can be regimented in first-order logic.

4.1 Regimenting Theories

Example 4.1.1 (The theory of partial orders) We suppose that there is a relation, which we'll denote by $\leq$, and we then proceed to lay down some postulates for this relation. In particular:

- Postulate 1: The relation $\leq$ is reflexive in the sense that it holds between anything and itself. For example, if we were working with numbers, we could write $2 \leq 2$, or, more generally, we could write $n \leq n$ for any n. For this last phrase, we have a shorthand: we abbreviate it by $\forall n(n \leq n)$, which can be read out as "for all n, $n \leq n$." The symbol $\forall$ is called the **universal quantifier**.
- Postulate 2: The relation $\leq$ is transitive in the sense that if $x \leq y$ and $y \leq z$, then $x \leq z$. Again, we can abbreviate this last sentence as

$$\forall x \forall y \forall z((x \leq y \wedge y \leq z) \rightarrow x \leq z),$$

  which can be read as, "for all x, for all y, and for all z, if ..."
- Postulate 3: The relation $\leq$ is antisymmetric in the sense that if $x \leq y$ and $y \leq x$, then $x = y$. This postulate can be formalized as

$$\forall x \forall y((x \leq y \wedge y \leq x) \rightarrow x = y).$$

In these previous postulates, we see the same logical connectives that we used in propositional logic, such as $\wedge$ and $\rightarrow$. But now these connectives might hold between things that are not themselves sentences. For example, $x \leq y$ is not itself a sentence, because x and y aren't names of things. We say that x and y are **variables**, that $\leq$ is a **relation symbol**, and that $x \leq y$ is a **formula**. Finally, the familiar symbol $=$ is also a relation symbol. ⌟

We've described just the barest of bones of the theory of a partial order. There are a couple of further things that we would definitely like to be able to do with this theory. First, we would like to be able to derive consequences from the postulates – i.e., we would like to derive theorems from the axioms. In order to do so, we will need to specify the **rules of derivation** for first-order logic. We will do that later in this chapter. We would also like to be able to identify mathematical structures that exemplify the axioms of partial order. To that end, we devote the following chapter to the **semantics**, or **model theory**, of first-order logic.

Example 4.1.2 (The theory of a linear order) Take the axioms of the theory of a partial order, and then add the following axiom:

$$\forall x \forall y((x \leq y) \vee (y \leq x)).$$

This axiom says that any two distinct things stand in the relation $\leq$. In other words, the elements from the domain form a total order. There are further specifications that we could then add to the theory of a linear order. For example, we could add an axiom saying that the linear order has endpoints. Alternatively, we could add an axiom saying that the linear order does not have endpoints. (Note, incidentally, that since either one of those axioms could be added, the original theory of linear orders is not complete – i.e., it leaves at least one sentence undecided.) We could also add an axiom saying that the linear order is dense, i.e., that between any two elements there is yet another element. ⌟

Example 4.1.3 (The theory of an equivalence relation) Let R be a binary relation symbol. The following axioms give the theory of an equivalence relation:

reflexive	$\vdash R(x,x)$
symmetric	$\vdash R(x,y) \rightarrow R(y,x)$
transitive	$\vdash (R(x,y) \wedge R(y,z)) \rightarrow R(x,z)$

Here when we write an open formula, such as $R(x,x)$, we mean to implicitly quantify universally over the free variables. That is, $\vdash R(x,x)$ is shorthand for $\vdash \forall x\, R(x,x)$. ⌟

Example 4.1.4 (The theory of abelian groups) We're all familiar with number systems such as the integers, the rational numbers, and the real numbers. What do these number systems have in common? One common structure between them is that they have a binary relation $+$ and a neutral element 0, and each number has a unique inverse. We also notice that the binary relation $+$ is associative in the sense that $x+(y+z) = (x+y)+z$, for all x, y, z. We can formalize this last statement as

$$\forall x \forall y \forall z(x + (y + z) = (x + y) + z).$$

In many familiar cases, the operation $+$ is also commutative; that is,

$$\forall x \forall y(x + y = y + x).$$

Bringing these postulates together, we have the theory of abelian groups. Notice that in this case, we've enlarged our vocabulary to include a symbol $+$ and a symbol 0. The symbol $+$ is not exactly a relation symbol, but instead is a function symbol. Intuitively speaking, given any names n and m of numbers, $n + m$ also names a number. Similarly, 0 is taken to be the name of some specific number, and in this sense it differs from a variable. ⌟

Example 4.1.5 (Boolean algebra) Suppose that $+$ and $\cdot$ are binary function symbols, and that 0 and 1 are constant symbols. If you look back at our discussion of Boolean algebras (Section 3.2), you'll see that each of the axioms amounts to a first-order sentence, where we use the $+$ symbol instead of the $\vee$ symbol, and the $\cdot$ symbol instead of the $\wedge$ symbol (since those symbols are already being used as our logical connectives). The theory of Boolean algebras is an example of an **algebraic theory**, which means that it can be axiomatized using only function symbols and equations. ⌟

Example 4.1.6 (Arithmetic) It's possible to formulate a first-order theory of arithmetic, e.g., Peano arithmetic. For this, we could use a signature Σ with constant symbols 0 and 1, and binary function symbols $+$ and $\cdot$. ⌟

Example 4.1.7 (Set theory) It's possible to formulate a first-order theory of sets, e.g., Zermelo–Fraenkel set theory. For this, we could use a signature Σ with a single relation symbol $\in$. However, for the elementary theory of the category of sets (ETCS), as we developed in Chapter 2, it would be more natural to use the framework of many-sorted logic, having one sort for sets and another sort for functions between sets. For more on many-sorted logic, see Chapter 5. ⌟

Example 4.1.8 (Mereology) There are various ways to formulate a first-order theory of mereology. Most presentations begin with a relation symbol $pt(x, y)$ to indicate that x is a part of y. Then we add some axioms that look a lot like the axioms for the less-than relation $<$ for a finite Boolean algebra. ⌟

4.2 Logical Grammar

Abstracting from the previous examples and many others like them throughout mathematics, we now define the language of first-order logic as follows.

DEFINITION 4.2.1 The **logical vocabulary** consists of the symbols:

$$\bot \; \forall \; \exists \; \wedge \; \vee \; \neg \; \rightarrow \; (\;)$$

The symbol $\bot$ will serve as a propositional constant. The final two symbols here, the parentheses, are simply punctuation symbols that will allow us to keep track of groupings of the other symbols.

Please note that we intentionally excluded the equality symbol $=$ from the list of logical vocabulary. Several philosophers in the twentieth century discussed the question of whether the axioms for equality were analytic truths or whether they should be considered to form a specific, contingent theory. We will not enter into the philosophical discussion at this point, but it will help us to separate out the theory of equality from the remaining content of our logical system. We will also take a more careful approach to variables by treating them as part of a theory's nonlogical vocabulary. Our reason for doing so will become clear when we discuss the notion of translations between theories.

DEFINITION 4.2.2 A **signature** Σ consists of

1. A countably infinite collection of **variables**.
2. A collection of **relation symbols**, each of which is assigned a natural number called its **arity**. A 0-ary relation symbol is called a **propositional constant**.
3. A collection of **function symbols**, each of which is assigned a natural number called its arity. A 0-ary function symbol is called a **constant symbol**.

DISCUSSION 4.2.3 Some logicians use the name **similarity type** as a synonym for **signature**. There is also a tendency among philosophers to think of a signature as the vocabulary for an **uninterpreted language**. The idea here is that the elements of the signature are symbols that receive meaning by means of a semantic interpretation. Nonetheless, we should be careful with this kind of usage, which might suggest that formal languages lie on the "mind side" of the mind–world divide, and that an interpretation relates a mental object to an object in the world. In fact, formal languages, sentences, and theories are all *mathematical objects* – of precisely the same ontological kind as the models that interpret them. We discuss this issue further in the next chapter.

Although a list of variables is technically part of a signature, we will frequently omit mention of the variables and defer to using the standard list $x, y, x_1, x_2, \ldots$ Only in cases where we are comparing two theories will we need to carefully distinguish their variables from each other.

Example 4.2.4 Every propositional signature is a special case of a signature in the sense just defined. ⌟

Example 4.2.5 For the theory of abelian groups, we used a signature Σ that has a binary function symbol $+$ and a constant symbol 0. Some other presentations of the theory of abelian groups use a signature Σ' that also has a unary function symbol "$-$" for the inverse of an element. Still other presentations of the theory use a signature that doesn't have the constant symbol 0. We will soon see that there is a sense in which these different theories all deserve to be called *the* theory of abelian groups. ⌟

DISCUSSION 4.2.6 Let Σ be the signature consisting of a binary relation symbol r, and let Σ' be the signature consisting of a binary relation symbol R. Are these signatures the same or different? That depends on what implicit background conventions that we adopt – in particular, whether our specification of a signature is case sensitive or not. In fact, we could adopt a convention that was even stricter in how it individuates signatures.

For example, let Σ'' be the signature consisting of a binary relation symbol r. One could say that Σ'' is a different signature from Σ because the r in Σ'' occurs at a different location on the page than the r that occurs in Σ. Of course, we would typically assume that $\Sigma'' = \Sigma$, but such a claim depends on an implicit background assumption that there is a single letterform of which the two occurences of r are instances.

We will generally leave these implicit background assumptions unmentioned. Indeed, to make these background assumptions explicit, we would have to rely on further implicit background assumptions, and we would never make progress in our study of first-order logic.

Let Σ be a fixed signature. We first define the sets of Σ-terms and Σ-formulas.

DEFINITION 4.2.7 We simultaneously define the set of **Σ-terms**, and the set $FV(t)$ of **free variables** of a Σ-term t as follows:

1. If x is a variable of Σ, then x is a Σ-term and $FV(x) = \{x\}$.
2. If f is a function symbol of Σ, and $t_1, \dots, t_n$ are Σ-terms, then $f(t_1, \dots, t_n)$ is a Σ-term and

$$FV(f(t_1, \dots, t_n)) \ := \ FV(t_1) \cup \cdots \cup FV(t_n).$$

DEFINITION 4.2.8 We simultaneously define the set of **Σ-formulas** and the set $FV(\phi)$ of free variables of each Σ-formula ϕ as follows:

1. $\perp$ is a formula and $FV(\perp) = \emptyset$.
2. If r is an n-ary relation symbol in Σ, and $t_1, \dots, t_n$ are terms, then $r(t_1, \dots, t_n)$ is a formula and

$$FV(r(t_1, \dots, t_n)) \ := \ FV(t_1) \cup \cdots \cup FV(t_n).$$

3. If ϕ and ψ are formulas, then $\phi \wedge \psi$ is a formula with $FV(\phi \wedge \psi) = FV(\phi) \cup FV(\psi)$. Similarly for the other Boolean connectives $\neg, \vee, \rightarrow$.
4. If ϕ is a formula, then so is $\exists x\phi$ and $FV(\exists x\phi) = FV(\phi)\backslash\{x\}$. Similarly for $\forall x\phi$.

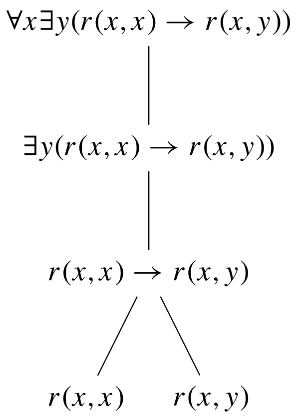
A formula ϕ is called **closed**, or a **sentence**, if $FV(\phi) = \emptyset$.

A more fully precise definition of Σ-formulas would take into account the precise location of parentheses. For example, we would want to say that $(\phi \wedge \psi)$ is a Σ-formula when ϕ and ψ are Σ-formulas. Nonetheless, we will continue to allow ourselves to omit parentheses when no confusion is likely to result.

NOTE 4.2.9 Our definition of the set of formulas allows for redundant quantification. For example, the string $\exists x \forall x(x = x)$ is a well-formed formula according to our definition. This formula results from applying the quantifier $\exists x$ to the sentence $\forall x(x = x)$. We will have to be careful in our definition of derivation rules, and semantic rules, to take the case of empty quantification into account.

DEFINITION 4.2.10 The **elementary formulas** are those of the form $r(t_1, \dots, t_n)$, i.e., formulas that involve no Boolean connectives or quantifiers.

It is helpful to think of formulas in terms of their **parse trees**. For example, the formula $\forall x \exists y (r(x,x) \to r(x,y))$ has the following parse tree:



The bottom nodes must each be elementary formulas, i.e., either $\bot$ or a relation symbol followed by the appropriate number of terms. Each parent–child relationship in the tree corresponds to one of the Boolean connectives or to one of the quantifiers.

Formulas stand in one-to-one correspondence with parse trees: each well-formed tree ends with a specific formula, and no other tree yields the same formula. Using the identity of formulas and parse trees, we can easily define a few further helpful notions:

DEFINITION 4.2.11 Let ϕ be a Σ-formula. The family of **subformulas** of ϕ consists of all those formulas that occur at some node in its parse tree.

DEFINITION 4.2.12 If a quantifier $\exists x$ occurs in the formula ϕ, then the **scope** of that occurrence is the formula that occurs at the immediately previous node in the parse tree.

For example, in the formula $\forall x \exists y (r(x,x) \to r(x,y))$, the scope of $\exists y$ is the formula $r(x,x) \to r(x,y)$. In contrast, in the formula $\forall x (r(x,x) \to \exists y r(x,y))$, the scope of $\exists y$ is the formula $r(x,y)$.

We can now make the notion of free and bound variables even more precise. In particular, each individual occurrence of a variable in ϕ is either free or bound. For example, in the formula $p(x) \wedge \exists x p(x)$, x occurs freely in the first subformula and bound in the second subformula.

DEFINITION 4.2.13 (Free and bound occurrences) An occurence of a variable x in ϕ is **bound** just in case that occurrence is within the scope of either $\forall x$ or $\exists x$. Otherwise that occurence of x is free.

We could now perform a sanity check to make sure that our two notions of bound/free variables coincide with each other.

FACT 4.2.14 *A variable x is free in ϕ (in the sense of the definition of Σ-formulas) if and only if there is a free occurence of x in ϕ (in the sense that this occurence does not lie in the scope of any corresponding quantifier).*

It is also sometimes necessary to distinguish particular occurrences of a subformula of a formula and to define the **depth** at which such an instance occurs.

DEFINITION 4.2.15 Let ψ be a node in the parse tree of ϕ. The **depth** of ψ is the number of steps from ψ to the root node. We say that ψ is a **proper subformula** of ϕ if ψ occurs with depth greater than 0.

The parse trees of formulas are finite by definition. Therefore, the depth of every occurrence of a subformula of ϕ is some finite number.

There are a number of other properties of formulas that are definable in purely syntactic terms. For example, we could define the **length** of a formula. We could then note that the connectives take formulas of a certain length and combine them to create formulas of a certain greater length.

EXERCISE 4.2.16 Show that no Σ-formula can occur as a proper subformula of itself.

We now define a substitution operation $\phi \mapsto \phi[t/x]$ on formulas, where t is a fixed term and x is a fixed variable. The intention here is that $\phi[t/x]$ results from replacing all free occurrences of x in ϕ with t. We first define a corresponding operation on terms.

DEFINITION 4.2.17 Let t be a fixed term, and let x be a fixed variable. We define the operation $s \mapsto s[t/x]$, where s is an arbitrary term, as follows:

1. If s is a variable, then $s[t/x] \equiv s$ when $s \not\equiv x$, and $s[t/x] \equiv t$ when $s \equiv x$. (Here $\equiv$ means literal identity of strings of symbols.)
2. Suppose that $s \equiv f(t_1, \dots, t_n)$, where f is a function symbol and $t_1, \dots, t_n$ are terms. Then we define

$$s[t/x] \;\equiv\; f(t_1[t/x], \dots, t_n[t/x]).$$

   This includes the special case where f is a 0-ary function symbol, where $f[t/x] \equiv f$.

DEFINITION 4.2.18 Let t be a fixed term, and let x be a fixed variable. We define the operation $\phi \mapsto \phi[t/x]$, for ϕ an arbitrary formula, as follows:

1. For the proposition $\bot$, let $\bot[t/x] := \bot$.
2. For an elementary formula $r(t_1, \dots, t_n)$, let

$$r(t_1, \dots, t_n)[t/x] \;:=\; r(t_1[t/x], \dots, t_n[t/x]).$$

3. For a Boolean combination $\phi \wedge \psi$, let

$$(\phi \wedge \psi)[t/x] \;:=\; \phi[t/x] \wedge \psi[t/x],$$

   and similarly for the other Boolean connectives.
4. For an existentially quantified formula $\exists y\phi$, let

$$(\exists y\phi)[t/x] := \begin{cases} \exists y(\phi[t/x]) & \text{if } x \not\equiv y, \\ \exists y\phi & \text{if } x \equiv y. \end{cases}$$

5. For a universally quantifier formula $\forall y\phi$, let

$$(\forall y\phi)[t/x] := \begin{cases} \forall y(\phi[t/x]) & \text{if } x \not\equiv y, \\ \forall y\phi & \text{if } x \equiv y. \end{cases}$$

PROPOSITION 4.2.19 *For any formula ϕ, the variable x is not free in $\phi[y/x]$.*

Proof We first show that $x \notin FV(t[y/x])$ for any term t. That result follows by a simple induction on the construction of terms.

Now let ϕ be an elementary formula. That is, $\phi = r(t_1, \ldots, t_n)$. Then we have

$$\begin{aligned} FV(\phi[y/x]) &= FV(r(t_1, \ldots, t_n)[y/x]) \\ &= FV(r(t_1[y/x], \ldots, t_n[y/x])) \\ &= FV(t_1[y/x]) \cup \cdots \cup FV(t_n[y/x]). \end{aligned}$$

Since $x \notin FV(t_i[y/x])$, for $i = 1, \ldots, n$, it follows that $x \notin FV(\phi[y/x])$.

The argument for the Boolean connectives is trivial, so we turn to the argument for the quantifiers. Suppose that the result is true for ϕ. We need to show that it's also true for $\exists v\phi$. Suppose first that $v \equiv x$. In this case, we have

$$(\exists v\phi)[y/x] = (\exists x\phi)[y/x] = \exists x\phi.$$

Since $x \notin FV(\exists x\phi)$, it follows that $x \notin FV((\exists v\phi)[y/x])$. Suppose now that $v \not\equiv x$. In this case, we have

$$(\exists v\phi)[y/x] = \exists v(\phi[y/x]).$$

Since $x \notin FV(\phi[y/x])$, it follows then that $x \notin FV((\exists v\phi)[y/x])$. The argument is analogous for the quantifier $\forall v$. Therefore, for any formula ϕ, the variable x is not free in $\phi[y/x]$. □

4.3 Deduction Rules

We suppose again that Σ is a fixed signature. The goal now is to define a relation $\Gamma \vdash \phi$ of derivability, where Γ is a finite sequence of Σ-formulas and ϕ is a Σ-formula. Our derivation rules come in three groupings: rules for the Boolean connectives, rules for the $\perp$ symbol, and rules for the quantifiers.

Boolean Connectives

We carry over all of the rules for the Boolean connectives from propositional logic (see Section 1.2). These rules require no special handling of variables. For example, the following is a valid instance of $\wedge$-elim:

$$\frac{\Gamma \vdash \phi(x) \wedge \psi(y)}{\Gamma \vdash \phi(x)}.$$

Falsum

We intend for the propositional constant $\perp$ to serve as shorthand for "the false." To this end, we define its introduction and elimination rules as follows.

$\perp$ **intro**	$\dfrac{\Gamma \vdash \phi \wedge \neg\phi}{\Gamma \vdash \perp}$	$\perp$ **elim**	$\dfrac{\Gamma \vdash \perp}{\Gamma \vdash \phi}$

Quantifiers

In order to formulate good derivation rules for the quantifiers, we have to make a couple of strategic choices. In actual mathematical practice, mathematicians simply introduce new vocabulary whenever they need it. In some cases, new vocabulary is introduced by way of definition – for example, when a mathematician says something like, "we say that a number x is *prime* just in case ..." where the words following the dots refer to previously understood mathematical concepts. In other cases, the newly introduced vocabulary is really just newly introduced notation – for example, when a mathematician says something like, "let n be a natural number." In this latter case, the letter "n" wasn't a part of the original vocabulary of the theory of arithmetic, and was introduced as a matter of notational convenience.

Nonetheless, for our purposes it will be most convenient to have a fixed vocabulary Σ for a theory. But this means that if Σ has no constant symbols, then we might have trouble making use of the quantifier introduction and elimination rules. For example, imagine trying to derive a theorem in the theory of Boolean algebras if you weren't permitted to say, "let a be an arbitrary element of the Boolean algebra B." In order to simulate mathematics' free use of new notation, we'll simply be a bit more liberal in the way that we allow free variables to be used. To this end, we define the following notion.

DEFINITION 4.3.1 We say that t **is free for** x **in** ϕ just in case one of the following conditions holds:

1. ϕ is atomic, or
2. ϕ is a Boolean combination of formulas, in each of which t is free for x, or
3. $\phi := \exists y\psi$, and $y \notin FV(t)$, and t is free for x in ψ, where $x \neq y$.

Intuitively speaking, t is free for x in ϕ just in case substituting t in for x in ϕ does not result in any of the variables in t being captured by quantifiers. For example, in the formula $p(x)$, the variable y is free for x (since y is free in $p(y)$). In contrast, in the formula $\exists yp(x)$, the variable y is not free for x (since y is not free in $\exists yp(y)$). We will need this notion in order to coordinate our intro and elim rules for the quantifiers. For example, the rule of $\forall$-elim should say something like: $\forall x\phi(x) \vdash \phi(y)$. However, if this rule were not restricted in some way, then it would yield

$$\forall x\exists y(x \neq y) \ \vdash \ \exists y(y \neq y),$$

which is intuitively invalid.

$\forall$ **intro**	$\dfrac{\Gamma \vdash \phi}{\Gamma \vdash \forall x\phi}$	where x is not free in Γ.

$\forall$ **elim**	$\dfrac{\Gamma \vdash \forall x\phi}{\Gamma \vdash \phi[t/x]}$	where t is free for x.

The ∀-intro rule is easy to apply, for we only need to check that the variable x doesn't occur in the assumptions Γ from which ϕ is derived. Note that application of the ∀-intro rule can result in empty quantification; for example, $\forall x \forall x p(x)$ follows from $\forall x p(x)$.

To understand the restrictions on ∀-elim, note that it does license

$$\forall x\, r(x,x) \;\vdash\; r(y,y),$$

since $r(x,x)[y/x] \equiv r(y,y)$. In contrast, ∀-elim does not license

$$\forall x\, r(x,x) \;\vdash\; r(x,y),$$

since it is not the case that $r(x,x)[y/x] \equiv r(x,y)$. Similarly, ∀-elim does not license

$$\forall x \exists y\, r(x,y) \;\vdash\; \exists y\, r(y,y),$$

since y is not free for x in $\exists y\, r(x,y)$. Finally, ∀-elim permits universal quantifiers to be peeled off when they don't bind any variables. For example, $\forall x\, p \vdash p$ is licensed by ∀-elim.

Now we turn to the rules for the existential quantifier. First we state the rules in all their sequential glory:

∃ intro	$\dfrac{\Gamma \vdash \phi[t/x]}{\Gamma \vdash \exists x \phi}$	provided t is free for x in ϕ.

∃ elim	$\dfrac{\Gamma, \phi \vdash \psi}{\Gamma, \exists x \phi \vdash \psi}$	provided x is not free in ψ or Γ.

If we omit the use of auxiliary assumptions, we can rewrite the ∃ rules as follows:

∃ intro	$\dfrac{\vdash \phi[t/x]}{\vdash \exists x \phi}$	provided t is free for x in ϕ.

∃ elim	$\dfrac{\phi \vdash \psi}{\exists x \phi \vdash \psi}$	provided x is not free in ψ.

Again, let's look at some examples to illustrate the restrictions. First, in the case of the ∃-intro rule, suppose that there were no restriction on the term t. Let ϕ be the formula $\forall y\, r(x,y)$, and let t be the variable y, in which case $\phi[t/x] \equiv \forall y\, r(y,y)$. Then the ∃-in rule would yield

$$\forall y\, r(y,y) \;\vdash\; \exists x \forall y\, r(x,y),$$

which is intuitively invalid. (Consider, for example, the case where r is the relation $\leq$ on integers.) The problem, of course, is that the variable y is captured by the quantifier $\forall y$ when substituted into ϕ. Similarly, in the case of the ∃-elim rule, if there were no restriction on the variable x, then we could derive ϕ from $\exists x \phi$, and then, using ∀-intro, we could derive $\exists x \phi \vdash \forall x \phi$.

Structural Rules

In any proof system, there are some more or less tacit rules that arise from how the system is set up. For example, when someone learns natural deduction – e.g., via the system presented in Lemmon's *Beginning Logic* – then she will tacitly assume that she's allowed to absorb dependencies – e.g., if $\phi, \phi \vdash \psi$ then $\phi \vdash \psi$. These more or less tacit rules are called **structural rules** of the system – and there is a lot of interesting research on logical systems that drop one or more of these structural rules (see Restall, 2002). In this book, we stay within the confines of classical first-order logic; and we will not need to be explicit about the structural rules, except for the rule of **cut**, which allows sequents to be combined. Loosely speaking, cut says that if you have sequents $\Gamma \vdash \phi$ and $\Delta, \phi \vdash \psi$, then you may derive the sequent $\Gamma, \Delta \vdash \psi$.

As was the case with propositional logic, we will not specify a canonical way of writing predicate logic proofs. After all, our goal here is not to teach you the art of logical deduction; rather, our goal is to reflect on the relations between theories in formal logic.

Equality

As we mentioned before, there's something of a philosophical debate about whether the equality symbol $=$ should be considered as part of the logical or the nonlogical vocabulary of a theory. We don't want to get tangled up in that argument, but we do wish to point out how the axioms for equality compare to the axioms for a generic equivalence relation.

It is typical to write two axioms for equality, an introduction and an elimination rule. Equality introduction permits $\vdash t = t$ with any term t. Equality elimination permits

$$\frac{t = s \quad \phi[t/s]}{\phi}$$

so long as t is free for s in ϕ. Note that equality elimination allows us to replace single instances of a term. For example, if we let ϕ be the formula $r(s,t)$, then $\phi[t/s]$ is the formula $r(t,t)$. Hence, from $t = s$ and $r(t,t)$, equality elimination permits us to derive $r(s,t)$.

From the equality axioms, we can easily show that it's an equivalence relation. The introduction rule shows that it's reflexive. For symmetry, we let ϕ be the formula $y = x$, in which case $\phi[x/y]$ is the formula $x = x$. Thus, we have

$$\frac{x = y \quad x = x}{y = x}.$$

For transitivity, let ϕ be the formula $x = z$, in which case $\phi[y/x]$ is the formula $y = z$. Thus, we have

$$\frac{y = x \quad y = z}{x = z}.$$

This completes the list of the proof rules for our system of first-order logic, i.e., our definition of the relation $\vdash$. Before proceeding to investigate the properties of this relation, let's see a couple of examples of informal proofs.

Example 4.3.2 Let's show that $\exists x(\phi(x) \wedge \psi(x)) \vdash \exists x \phi(x)$. First note that $\phi(x) \wedge \psi(x) \vdash \phi(x)$ from $\wedge$-elim. Then $\phi(x) \wedge \psi(x) \vdash \exists x \phi(x)$ from $\exists$-intro. Finally, since $\exists x \phi(x)$ contains no free occurrences of x, we have $\exists x(\phi(x) \wedge \psi(x)) \vdash \exists x \phi(x)$. ⌟

Example 4.3.3 Of course, we should have $\forall x \phi \vdash \forall y(\phi[y/x])$, so long as y is free for x in ϕ. Using the rules we have, we can derive this result in two steps. First, we have $\forall x \phi(x) \vdash \phi[y/x]$ from $\forall$-elim, and then $\phi[y/x] \vdash \forall y \phi[y/x]$ by $\forall$-intro. We only need to verify that x is not free in $\phi[y/x]$. This can be shown by a simple inductive argument. ⌟

Recall that propositional logic is compositional in the following sense: Suppose that ϕ is a formula, and ψ is a subformula of ϕ. Let ϕ' denote the result of replacing ψ in ϕ with another formula ψ' where $\vdash \psi \leftrightarrow \psi'$. Then $\vdash \phi \leftrightarrow \phi'$. That result is fairly easy to prove by induction on the construction of proofs. It also follows from the truth-functionality of the Boolean connectives, by means of the completeness theorem. In this section, we are going to prove an analogous result for predicate logic. To simplify notation, we introduce the following.

DEFINITION 4.3.4 For formulas ϕ and ψ, we say that ϕ and ψ are **logically equivalent**, written $\phi \simeq \psi$, just in case both $\phi \vdash \psi$ and $\psi \vdash \phi$.

It is not hard to show that $\simeq$ is an equivalence relation on the set of formulas. Note that formulas ϕ and ψ can be equivalent in this sense even if they don't share all free variables in common – as long as the nonmatching variables occur vacuously. For example, $p(x)$ is equivalent to $p(x) \wedge (y = y)$, and it's also equivalent to $p(x) \vee (y \neq y)$. (The issue here has nothing in particular to do with the equality relation. The variable y also occurs vacuously in $p(y) \vee \neg p(y)$.) In contrast, the formulas $p(x)$ and $p(y)$ are not equivalent, since it's not universally valid that $\vdash \forall x \forall y(p(x) \leftrightarrow p(y))$.

LEMMA 4.3.5 *The relation $\simeq$ is compatible with the Boolean connectives in the following sense: if $\phi \simeq \phi'$ and $\psi \simeq \psi'$, then $(\phi \wedge \psi) \simeq (\phi' \wedge \psi')$, and similarly for the other Boolean connectives.*

The proof of this lemma is a fairly simple application of the introduction and elimination rules for the connectives. To complete the proof of the replacement theorem, we need one more lemma.

LEMMA 4.3.6 *If $\phi \simeq \psi$ then $\exists x \phi \simeq \exists x \psi$.*

Proof Suppose that $\phi \simeq \psi$, which means that $\phi \vdash \psi$ and $\psi \vdash \phi$. We're now going to show that $\exists x \phi \vdash \exists x \psi$. By $\exists$-in we have $\psi \vdash \exists x \psi$, hence by cut we have $\phi \vdash \exists x \psi$. Since x does not occur free in $\exists x \psi$, we have $\exists \phi \vdash \exists x \psi$ by $\exists$-out. □

THEOREM 4.3.7 (Replacement) *Suppose that ϕ is a formula in which ψ occurs as a subformula, and ϕ' is the result of replacing ψ with ψ'. If $\psi \simeq \psi'$ then $\phi \simeq \phi'$.*

In most presentations of the predicate calculus (i.e., the definition of the relation $\vdash$) the two central results are the soundness and completeness theorems. Intuitively speaking, the soundness theorem shows that the definition doesn't overgenerate, and the completeness theorem shows that it doesn't undergenerate. However, in fact, these results show something quite different – they show that the definition of $\vdash$ matches the definition of another relation $\vDash$. We will discuss this other relation $\vDash$ in Chapter 6, where we will also prove the traditional soundness and completeness theorems. In the remainder of this section, we show that the predicate calculus is consistent in the following purely syntactic sense.

DEFINITION 4.3.8 We say that the relation $\vdash$ is **consistent** just in case there is some formula ϕ that is not provable. Similarly, we say that a theory T is **consistent** just in case there is a formula ϕ such that $T \nvdash \phi$.

Note that the definition of consistency for $\vdash$ presupposes a fixed background signature Σ.

PROPOSITION 4.3.9 *A theory T is consistent iff $T \nvdash \bot$.*

Proof If T is inconsistent, then $T \vdash \phi$ for all formulas ϕ. In particular, $T \vdash \bot$. Conversely, if $T \vdash \bot$, then RA and DN yield $T \vdash \phi$ for any formula ϕ. □

THEOREM 4.3.10 *The predicate calculus is consistent.*

Proof Let Σ be a fixed predicate logic signature, and let Σ' be a propositional signature whose cardinality is greater than or equal to that of Σ. We will use the symbol $\vdash^*$ to denote derivability in the propositional calculus. Define a map $\phi \mapsto \phi^*$ from the formulas of Σ to the formulas of Σ' as follows:

- $\bot^* = \bot$.
- For any terms $t_1, \dots, t_n$, $(p_i(t_1, \dots, t_{n_i}))^* = q_i$.
- $(\phi \wedge \psi)^* = \phi^* \wedge \psi^*$, and similarly for the other Boolean connectives.
- $(\forall x \phi)^* = \phi^*$ and $(\exists x \phi)^* = \phi^*$.

We now use induction on the definition of $\vdash$ to show that if $\Gamma \vdash \phi$, then $\Gamma^* \vdash^* \phi^*$. We will provide a few representative steps, and leave it to the reader to supply the others.

- The base case, rule of assumptions, is trivial.
- Consider the case of $\wedge$-out. Suppose that $\Gamma \vdash \phi$ follows from $\Gamma \vdash \phi \wedge \psi$ by $\wedge$-out. By the inductive hypothesis, $\Gamma^* \vdash^* (\phi \wedge \psi)^*$. Using the definition of $(\phi \wedge \psi)^*$, it follows that $\Gamma^* \vdash^* \phi^* \wedge \psi^*$. Hence, by $\wedge$-out, we have $\Gamma^* \vdash^* \phi^*$.
- Consider the case of $\forall$-intro. That is, suppose that $\Gamma \vdash \forall x \phi$ is derived from $\Gamma \vdash \phi$ using $\forall$-in. In this case, the induction hypothesis tells us that $\Gamma^* \vdash^* \phi^*$. And since $(\forall x \phi)^* = \phi^*$, we have $\Gamma^* \vdash^* (\forall x \phi)^*$.

Completing the previous steps shows that if $\Gamma \vdash \phi$, then $\Gamma^* \vdash^* \phi^*$. Since the propositional calculus is consistent, $\nvdash^* \bot$ and, therefore, $\nvdash \bot$. □

DISCUSSION 4.3.11 Notice that the previous proof does not use the fact that our ∀-intro rule demands that x not occur free in Γ. Thus, this proof also shows the consistency of a proof system with an *unrestricted* ∀-intro rule.

But an unrestricted ∀-intro rule would nonetheless severely restrict the expressive power of our logic. Indeed, it would license

$$x \neq y \vdash \forall y(x \neq y) \vdash x \neq x,$$

the last of which contradicts the axioms for equality. Thus, an unrestricted ∀-intro would make $\forall x \forall y(x = y)$ a tautology.

4.4 Empirical Theories

Here we use the phrase "empirical theory" or "scientific theory," to mean a theory that one intends to describe the physical world. You know many examples of such theories: Newtonian mechanics, Einstein's general theory of relativity, quantum mechanics, evolutionary biology, the phlogiston theory of combustion, etc. You may also know many examples of theories from pure mathematics, such as set theory, group theory, ring theory, topology, and the theory of smooth manifolds. Intuitively, empirical theories differ in some important way from pure mathematical theories. We stress "intuitively" here because Quine brought into question the idea that there is a principled distinction between two types of theories. For the time being, we won't engage directly with Quine's more philosophical arguments against this distinction. Instead, we will turn back the clock to the time when Rudolf Carnap, among others, hoped that formal logic might illuminate the structure of scientific theories.

Rudolf Carnap was the primary advocate of the idea that philosophers ought to pursue a *syntactic* analysis of scientific theories. The story is typically told as follows: Carnap sought to construct a theory *of* scientific theories. Moreover, following in the footsteps of Bertrand Russell and Gottlob Frege, Carnap believed that philosophy had no business directly engaging in empirical questions. As Russell (1914b) had argued, philosophers ought to leave empirical questions to the empirical sciences. Thus, Carnap thought that a good philosophical theory of scientific theories ought to restrict itself to the purely formal aspects of those theories. In particular, the "metascientist" – i.e., the philosopher of science – ought to make use only of syntactic concepts.

Carnap begins his *Wissenschaftslogik* program in earnest in his first major book, *Logische Aufbau der Welt*. Already here we see the emphasis on "explication" – i.e., taking an intuitive concept and providing a precise formal counterpart. Carnap's paradigms of explication are those from nineteenth- and early-twentieth-century mathematics – explications of concepts such as "infinity" and "continuous function" and "open subset." Nonetheless, in the *Aufbau*, Carnap hasn't yet found his primary tool of analysis. That would only come from the development of logical metatheory in the 1920s. Carnap was working at the time in Vienna, among the other members of the infamous Vienna Circle. One of the youngest members of the circle was Kurt Gödel, whose 1929 PhD dissertation contained the first proof of the completeness of the predicate calculus. Thus, logical metatheory – or metamathematics – was in the air in Vienna, and Carnap was to try his hand at applying an analogous methodology to the empirical sciences. As the goal of metamathematics is to provide a rigorous theory *about* mathematics, Carnap wished to create a rigorous theory *about* the empirical sciences.

By the mid 1930s, Carnap had found his vision. In *Die Logische Syntax der Sprache*, Carnap states that his goal is to formalize scientific theories in the same way that Russell and Whitehead had formalized arithmetic – but with one important addition. With a theory of pure mathematics, the job is done once the relevant primitive concepts and axioms have been written down. However, empirical theories are, by their nature, "world directed" – i.e., they try to say something about concrete realities. Thus, an adequate analysis of a scientific theory cannot rest content with explaining that theory's formal structure. This analysis must also say something about how the theory gains its *empirical content*.

The task of explaining how a theory gains empirical content was to occupy Carnap for most of the remainder of his career. In fact, it became the stone on which the entire logical positivist movement stumbled. But we've gotten ahead of ourselves. We need first to see how Carnap proposed to analyze the structure of empirical theories.

What then is a theory? From the point of view of first-order logic, a theory T is specified by a signature Σ, and a set of axioms in that signature. Amazingly, many of the theories of pure mathematics can be described in terms of this simple schema. If, however, we intend for our theory T to describe concrete reality, what more do we need to add? Carnap's first proposal was a blunt instrument: he suggests identifying the empirical content of a theory by means of a division of that theory's vocabulary into two parts:

> The total language of science, L, is considered as consisting of two parts, the observation language L_O and the theoretical language L_T . . . Let the observation vocabulary V_O be the class of the descriptive constants of L_O . . . The terms of V_O are predicates designating observable properties of events or things (e.g., "blue," "hot," "large," etc.) or observable relations between them (e.g., "x is warmer than y," "x is contiguous to y," etc.). (Carnap, 1956, pp. 40-41)

Let's rewrite all of this in a better notation: the language of science consists of all the formulas built on some particular signature Σ, where Σ has a subset $O \subseteq \Sigma$ of observation vocabulary. The idea here is that terms in O have ostensive definitions – e.g., O might contain predicates such as "x is red" or "x is to the left of y." The elements of $\Sigma \backslash O$ are theoretical vocabulary, which need not have any direct empirical meaning. For example, $\Sigma \backslash O$ might contain predicates such as "x is a force." Thus, Carnap hopes to isolate empirical content by means of specifying a preferred subvocabulary of the language of science.

Before proceeding, note that Carnap – in this 1956 article – explicitly states that "for each language part the admitted types of variables are specified." That phrase was completely ignored by Carnap's subsequent critics, as we will soon see. And why did they ignore it? The reason, we suspect, is that they had been convinced by Quine that the notion of "types of variables" couldn't possibly make any difference in any philosophical debate. Well, Quine wasn't exactly right about that, as we discuss in

Section 5.3. However, at present, our goal is to see Carnap through the eyes of his critics, and according to these critics, Carnap's proposal amounts to saying the following.

DEFINITION 4.4.1 A formula ϕ of Σ is an **observation sentence** (alternatively, **protocol sentence**) just in case no symbol in ϕ comes from $\Sigma\backslash O$. If T is a theory in Σ, then we let $T|_O$ denote all the consequences of T in the sublanguage based on O.

In the light of these definitions, Carnap's proposal would amount to saying that the **empirical content** of a theory T is $T|_O$. Indeed, that's precisely what people took him to be saying – and they judged him accordingly. In fact, one of the standard "challenges for scientific realism" was to point out that the empirical subtheory $T|_O$ has the same empirical content as the original theory T. Thus, every nontrivial theory T has an empirically equivalent rival!

DEFINITION 4.4.2 Let T_1 and T_2 be theories in Σ. Then T_1 and T_2 have the same empirical content – i.e., are **empirically equivalent** – just in case $T_1|_O = T_2|_O$.

This definition fits right in with the picture that the logical positivists treat sentences as synonymous whenever those sentences have the same empirical content. Indeed, many people take the positivists to be saying that two scientific theories T_1 and T_2 should be considered equivalent *tout court* if they have the same observational consequences.

Before we go on to consider the criticisms that were brought against Carnap's picture of empirical content, let's ask ourselves what purpose the picture was supposed to serve. In other words, what questions was Carnap trying to answer by means of this proposal? In fact, it seems that Carnap was trying to answer several questions simultaneously. First, Carnap, along with many other logical positivists, was concerned with epistemological questions, such as, "am I justified in believing theory T?" Apropos of this question, the goal of isolating empirical content is to make some headway on understanding how it is that we can be warranted in believing a theory. To be clear, it's not only empiricists who should want to understand how we can use evidence to regulate our belief in a theory. That's a problem for anyone who thinks that we can learn from experience – and that's everybody besides the most extreme rationalists.

Nonetheless, there were some logical positivists – and perhaps sometimes Carnap himself – who thought that the empirical content of a theory provides the *only* route to justifying belief in that theory. For that kind of radical empiricist, isolating empirical content takes on an additional negative role: showing which parts of a theory do *not* contribute to our reasons for believing (or accepting) it.

It is sometimes forgotten, however, that epistemology was not the only reason that Carnap wanted to isolate empirical content. In fact, there are good reasons to think that epistemology wasn't even the primary reason that Carnap wanted to isolate empirical content. To the contrary, Carnap – who was, by training, a neo-Kantian – was concerned with how the abstract, highly mathematical theories of physics function in making assertions about the world. To understand this, we have to remember that Carnap was vividly aware of the upheaval caused by the discovery, in the mid-nineteenth century, of non-Euclidean geometries. One result of this upheaval was that mathematical formalism became *detached* from the empirical world, and the words that occur in it were *de-interpreted*. For example, in pre-nineteenth-century geometry, mathematicians were wont to think that a word such as "line" refers to those things in physical reality that are, in fact, lines. But insofar as the word "line" occurs in pure geometry, it has no reference at all – it is merely a symbol in a formal calculus.

Given the flight of pure mathematics away from empirical reality, the task of the mathematized empirical sciences is to tie mathematics back down. In other words, the task of the mathematical physicist is to take the uninterpreted symbols of pure mathematics and to endow them with empirical significance. It is precisely this methodological maneuver – peculiar to the new physics – that drives Carnap's desire to analyze the notion of the empirical content of a theory.

In the middle of the twentieth century, analytic philosophy moved west – from Vienna and Berlin to Oxford, Cambridge (in both old and New England), and then to Princeton, Pittsburgh, UCLA, etc. As analytic philosophy moved west, the focus on narrowly epistemological questions increased. It's no surprise, then, that Carnap's critics – first Quine, then Putnam, etc. – read him as attempting first and foremost to develop an empiricist epistemology. And their criticisms are directed almost exclusively at these aspects of his view. In fact, philosophers have been so focused with epistemological questions that they seem to have forgotten the puzzle that Carnap faced, and that we still face today: how do the sciences use abstract mathematical structures to represent concrete empirical reality?

In any case, we turn now to the criticisms of Carnap's account of the empirical content of a theory T as its restriction $T|_O$ to consequences in the observation subvocabulary O of Σ. Doubtless, all these criticisms descend, in one sense or other, from Quine's master criticism in "Two Dogmas of Empiricism" (Quine, 1951b). Here Quine's target is ostensibly statements, rather than theories. He argues that it makes no sense to talk about a statement's admitting of confirming or infirming (i.e., disconfirming) instances, at least when that statement is taken in isolation. While Quine doesn't apply his moral to the theories of the empirical sciences, it is only natural to transfer his conclusions to that case: it doesn't make sense to talk about the empirical content of a theory T.

To get an explicit statement of this criticism of Carnap's point of view, we have to wait a decade – for Putnam's paper "What Theories Are Not" (1962). Here Putnam claims that the attempt to select a subset $O \subseteq \Sigma$ of observation vocabulary is "completely broken-backed." His argument focuses on showing the incoherence of the notion of an observation term. To this end, he assumes that

> If $P(x)$ is an observation predicate, then it is never the case that $P(t)$, where t is a theoretical entity.

Putnam then simply enumerates examples where observation predicates have been applied to theoretical entities, e.g., Newton speaking of "red corpuscles."

For the sake of argument, let's assume that Putnam is correct that scientific theories sometimes use a single term in both observational and theoretical roles. Already that would pose a challenge to the adequacy of Carnap's account. Carnap assumes that among the terms of a mature scientific theory, there are some that are simply not used in observation reports – except possibly when a scientist is speaking loosely, e.g., if she says, “I saw an electron in the cloud chamber.” Nonetheless, even if Putnam is right about that, his argument equivocates between formal and material modes of speech. On the one hand, Putnam speaks of observation predicates (formal mode); on the other hand, Putnam speaks of unobservable entities (material mode). Putnam’s worry seems to be that some confusion might result if the philosopher of science classifies $P(x)$ as an observation predicate and then a scientist attributes $P(x)$ to a theoretical entity. Or perhaps the problem is that we cannot divide the vocabulary of Σ because we need to use predicates together with terms even when they would lie on opposite sides of the divide?

The anti-Carnap sentiment must have been in the air, for in the very same year, Maxwell (1962) also argued for the incoherence of the distinction between theoretical and observational terms. What’s more, Maxwell explicitly claims that, in absence of this distinction, the only rational attitude toward a successful scientific theory is *full belief* – i.e., one must be a **scientific realist**.

Putnam and Maxwell seem to have convinced an entire generation of philosophers that Carnap’s approach cannot be salvaged. In fact, the conclusion seems to have been that *nothing* of Carnap’s approach could be salvaged, save the tendency to invoke results from mathematical logic. By the 1970s, there was no longer any serious debate about these issues. Instead, we find postmortem reflections on the “received view of scientific theories,” as philosophers rushed headlong in the direction of Quinean holistic realism about everything (science, math, metaphysics).

4.5 Translation

Almost every discussion in twentieth-century philosophy of science has something or other to do with relations between theories. For example, philosophers of science have shown great interest in the notion that one theory is **reducible** to another. Similarly, several philosophical discussions pivot on the notion of a **conservative extension** of a theory. For example, Hartry Field (1980) aims to show that standard physical theories are conservative extensions over their “purely nominalistic parts” – hoping to undercut Quine’s claim that belief in the existence of mathematical entities is demanded by belief in our best scientific theories.

We turn now to the task of explicating relations bewteen theories – i.e., giving a mathematically precise account of what these relations can be. One of the main questions considered in this book is

When are two theories T and T' the *same*, or *equivalent*?

Perhaps the answer seems clear: if a theory is a set of sentences, then two theories are the same if the corresponding sets of sentences are literally identical. However, there are numerous problems with that idea. First, that idea is not as clear as it might seem. When are two sets of sentences the same? What if the first set of sentences occurs in a book in the Princeton University library and the second set of sentences occurs on a chalkboard in Munich? Why would we say that those are the *same* sentences, when they occur in different spacetime locations?

Of course, the standard philosophical response to this worry is to shift focus from sentences to propositions – those abstract objects that are supposed to be expressed by concrete sentence tokens. Let's be completely clear: while we have no problems with abstract entities such as propositions, they won't help us make any progress deciding when sentences are synonymous, or when theories are equivalent. In other words, to say that sentences are synonymous if they express the same proposition may be *true*, but it is not an *explication* of synonymy. For the purposes of this book, we will set aside appeals to propositions or other such Platonic entities. We wish, instead, to provide clear and explicit definitions of equivalence (and other relations between theories) that could be applied to concrete cases (such as the debate whether Lagrangian and Hamiltonian mechanics are equivalent).

Let's suppose then that T and T' are first-order theories in a common signature Σ. Then we have the following obvious explication of equivalence:

DEFINITION 4.5.1 Let T and T' be theories in signature Σ. We say that T and T' are **logically equivalent** just in case $\text{Cn}(T) = \text{Cn}(T')$.

However, there are a couple of reasons why logical equivalence may not be a perfect explication of the notion of theoretical equivalence. First, there are cases of theories in the same signature that are the same "up to relabelling," but are not logically equivalent. For example, in the propositional signature $\Sigma = \{p, q\}$, let $T = \{p\}$ and let $T' = \{q\}$. Certainly there is one sense in which T and T' are different theories, since they disagree on which of the two propositional constants p and q should be affirmed. Nonetheless, there is another sense in which T' could be considered as a mere relabelling of T. At least structurally speaking, these two theories appear to be the same: they both have two propositional constants, and they assert precisely one of these two.

A second reason to worry about logical equivalence is that it cannot detect sameness of theories written in different signatures.

Example 4.5.2 Consider two signatures $\Sigma = \{f\}$ and $\Sigma' = \{\mathfrak{f}\}$, where both f and $\mathfrak{f}$ are one-place function symbols. (If you can't see the difference, the second f is written in Fraktur font. That raises an interesting question about whether f and $\mathfrak{f}$ are really the same letter or not.) Now let T be the theory with axiom $(f(x) = f(y)) \to (x = y)$, and let T' be the theory with axiom $(\mathfrak{f}(x) = \mathfrak{f}(y)) \to (x = y)$. Being written in different signatures, these two theories cannot be logically equivalent. But come now! Surely this is just a matter of different notations. Can't we write the *same* theory in different notation? ⌟

The problems with the previous example might be chalked up to needing a better criterion of sameness of signatures. However, that response won't help with the following sort of example.

Example 4.5.3 There are some theories that are intuitively equivalent, but not logically equivalent. In this example, we discuss two different formulations of the mathematical theory of groups.

Let $\Sigma_1 = \{\cdot, e\}$ be a signature where $\cdot$ is a binary function symbol and e is a constant symbol. Let T_1 be the following Σ_1-theory:

$$\big\{\forall x \forall y \forall z \big((x \cdot y) \cdot z = x \cdot (y \cdot z)\big), \forall x (x \cdot e = x \wedge e \cdot x = x), \\ \forall x \exists y (x \cdot y = e \wedge y \cdot x = e)\big\}$$

Now let $\Sigma_2 = \{\cdot, ^{-1}\}$, where $\cdot$ is again a binary function symbol and $^{-1}$ is a unary function symbol. Let T_2 be the following Σ_2-theory:

$$\big\{\forall x \forall y \forall z \big((x \cdot y) \cdot z = x \cdot (y \cdot z)\big), \\ \exists x \forall y \big(y \cdot x = y \wedge x \cdot y = y \wedge y \cdot y^{-1} = x \wedge y^{-1} \cdot y = x\big)\big\}$$

If you open one textbook of group theory, you might find the axiomatization T_1. If you open another textbook of group theory, you might find the axiomatization T_2. And yet, the authors believe themselves to be talking about the same theory. How can this be so? The theories T_1 and T_2 are written in different signatures, and so are not even candidates for logical equivalence. ⌟

DEFINITION 4.5.4 If Σ_1 and Σ_2 are signatures, we call a map from elements of the signature Σ_1 to Σ_2-formulas a **reconstrual** $F : \Sigma_1 \to \Sigma_2$ if it satisfies the following three conditions.

- For every n-ary predicate symbol $p \in \Sigma_1$, $Fp(\vec{x})$ is a Σ_2-formula with n free variables.
- For every n-ary function symbol $f \in \Sigma_1$, $Ff(\vec{x}, y)$ is a Σ_2-formula with $n + 1$ free variables.
- For every constant symbol $c \in \Sigma_1$, $Fc(y)$ is a Σ_2-formula with one free variable.

One can think of the Σ_2-formula $Fp(\vec{x})$ as a "translation" of the Σ_1-formula $p(\vec{x})$ into the signature Σ_2. Similarly, $Ff(\vec{x}, y)$ and $Fc(y)$ can be thought of as translations of the Σ_1-formulas $f(\vec{x}) = y$ and $c = y$, respectively.

A reconstrual $F : \Sigma_1 \to \Sigma_2$ naturally induces a map from Σ_1-formulas to Σ_2-formulas. In order to describe this map, we first need to describe the map that F induces from Σ_1-terms to Σ_2-formulas. Let $t(\vec{x})$ be a Σ_1-term. We define the Σ_2-formula $Ft(\vec{x}, y)$ recursively as follows.

- If t is the variable x_i then $Ft(x_i, y)$ is the Σ_2-formula $x_i = y$.
- If t is the constant symbol $c \in \Sigma_1$ then $Ft(y)$ is the Σ_2-formula $Fc(y)$.
- Suppose that t is the term $f(t_1(\vec{x}), \ldots, t_k(\vec{x}))$ and that each of the Σ_2-formulas $Ft_i(\vec{x}, y)$ have been defined. Then we define $Ft(x_1, \ldots x_n, y)$ to be the Σ_2-formula

$$\exists z_1 \ldots \exists z_k \big(Ft_1(\vec{x}, z_1) \wedge \ldots \wedge Ft_k(\vec{x}, z_k) \wedge Ff(z_1, \ldots, z_k, y)\big).$$

We use this map from Σ_1-terms to Σ_2-formulas to describe how F maps Σ_1-formulas to Σ_2-formulas. Let $\phi(\vec{x})$ be a Σ_1-formula. We define the Σ_2-formula $F\phi(\vec{x})$ recursively as follows.

- If $\phi(\vec{x})$ is the Σ_1-atom $s(\vec{x}) = t(\vec{x})$, where s and t are Σ_1-terms, then $F\phi(\vec{x})$ is the Σ_2-formula
$$\exists z\big(Ft(\vec{x}, z) \wedge Fs(\vec{x}, z)\big).$$
- If $\phi(\vec{x})$ is the Σ_1-atom $p(t_1(\vec{x}), \dots, t_k(\vec{x}))$, with $p \in \Sigma_1$ a k-ary predicate symbol, then $F\phi(\vec{x})$ is the Σ_2-formula
$$\exists z_1 \dots \exists z_k\big(Ft_1(\vec{x}, z_1) \wedge \dots \wedge Ft_k(\vec{x}, z_k) \wedge Fp(z_1, \dots, z_k)\big).$$
- The definition of the Σ_2-formula $F\phi$ extends to all Σ_1-formulas ϕ in the now familiar manner.

In this way, a reconstrual $F : \Sigma_1 \to \Sigma_2$ gives rise to a map between Σ_1-formulas and Σ_2-formulas.

DEFINITION 4.5.5 We call a reconstrual $F : \Sigma_1 \to \Sigma_2$ a **translation** of a Σ_1-theory T_1 into a Σ_2-theory T_2 if $T_1 \vdash \phi$ implies that $T_2 \vdash F\phi$ for all Σ_1-sentences ϕ. We will use the notation $F : T_1 \to T_2$ to denote a translation of T_1 into T_2.

Example 4.5.6 Let Σ be the empty signature, let T_1 be the theory in Σ that says "there are at least n things," and let T_2 be the theory in Σ that says "there are exactly n things." Since Σ is empty, there is precisely one reconstrual $F : \Sigma \to \Sigma$, namely the identity reconstrual. This reconstrual is a translation from T_1 to T_2, but it is *not* a translation from T_2 to T_1. ⌟

DISCUSSION 4.5.7 A translation $F : T_1 \to T_2$ is in some ways quite rigid – e.g., it must preserve all numerical claims. To see this, observe first that for any atomic formula $x = y$, we have $F(x = y) \equiv (x = y)$. Moreover, since F preserves the Boolean connectives and quantifiers, it follows that F preserves all statements of the form, "there are at least n things," and "there are at most n things," and "there are exactly n things."

The notion of a translation between signatures gives us a particularly nice way to understand the notion of **substitution**. Informally speaking, we perform a substitution on a formula ϕ by replacing a predicate symbol p (or a function symbol f, or a constant symbol c) in ϕ uniformly with some other formula θ that has the same free variables. Intuitively speaking, since the validity of an argument depends only on form, such a substitution should map valid arguments to valid arguments. In short, if $\phi \vdash \psi$, and if ϕ^* and ψ^* are the result of a uniform substitution, then we should also have $\phi^* \vdash \psi^*$.

The notion of substitution is, in fact, a special case of the notion of a reconstrual. In the working example, we define a reconstrual $F : \Sigma \to \Sigma$ by setting $Fp = \theta$, and $Fs = s$ for every other symbol. Then to substitute θ for p in ϕ is simply to apply the function F to ϕ, yielding the formula $F\phi$. We might hope then to show that

$$\phi \vdash \psi \quad \Longrightarrow \quad F\phi \vdash F\psi.$$

But this isn't quite right yet. For example, suppose that F is a reconstrual that maps the constant symbol c to the formula $\theta(y)$. In this case, $\vdash \exists! y(y = c)$, but it is not necessarily the case that $\vdash \exists!\, y\psi(y)$. To deal with this sort of case, we need to introduce the notion of admissibility conditions for function and constant symbols.

Suppose then that $F : \Sigma \to \Sigma'$ is a reconstrual. If $f \in \Sigma$ is an n-ary function symbol, then $Ff(\vec{x}, y)$ is a $(n+1)$-ary formula in Σ'. The **admissibility condition** for $Ff(\vec{x}, y)$ is simply the sentence

$$\forall\vec{x}\exists!\, y\, Ff(\vec{x}, y),$$

which says that Ff is a functional relation. In the case that f is a constant symbol (i.e., a 0-ary function symbol), Ff is a formula $\psi(y)$, and its admissibility condition is the formula $\exists! y\psi(y)$.

DEFINITION 4.5.8 If $* : \Sigma \to \Sigma'$ is a reconstrual, then we let Δ be the set of Σ'-formulas giving the admissibility conditions for all function symbols in Σ.

Thus, the correct version of the substitution theorem can be stated as follows:

$$\phi \vdash \psi \implies \Delta, \phi^* \vdash \psi^*,$$

where Δ are the admissibility conditions for the reconstrual $*$. To prove this, we show first that for any term t of Σ, Δ implies that t^* is a functional relation.

LEMMA 4.5.9 *Let $* : \Sigma \to \Sigma'$ be a reconstrual, and let Δ be the admissibility conditions for the function symbols in Σ. Then for any term t of Σ, $\Delta \vdash \exists! y\, t^*(\vec{x}, y)$.*

Proof We prove it by induction on the construction of t. The case where t is a variable is trivial. Suppose then that $t \equiv f(t_1, \ldots, t_n)$, where the result already holds for $t_1, \ldots, t_n$. In this case, $f(t_1, \ldots, t_n)^*$ is the relation

$$\exists z_1 \cdots \exists z_n (t_1^*(\vec{x}, z_1) \wedge \cdots \wedge t_n^*(\vec{x}, z_n) \wedge f^*(z_1, \ldots, z_n, y)).$$

Fix the n-tuple $\vec{x}$. We need to show that there is at least one y that stands in this relation, and that there is only one such y. For the former, since t_i^* is functional, there is a z_i such that $t_i^*(\vec{x}, z_i)$. Moroever, $\Delta \vdash \exists y f(z_1, \ldots, z_n, y)$. Thus, we've established the existence of at least one such y. For uniqueness, we first use the fact that if $t_i^*(\vec{x}, z_i)$ and $t^*(\vec{x}, z_i')$, then $z_i = z_i'$. Then we use the fact that

$$\Delta \vdash (f^*(z_1, \ldots, z_n, y) \wedge f^*(z_1, \ldots, z_n, y')) \to y = y'.$$

This establishes that $f(t_1, \ldots, t_n)^*$ is a functional relation. □

The next two lemmas show that, modulo these admissibility conditions, reconstruals preserve the validity of the intro and elim rules for equality.

LEMMA 4.5.10 *Let $* : \Sigma \to \Sigma'$ be a reconstrual, and let Δ be the admissibility conditions for the function symbols in Σ. Then for any term t of Σ, $\Delta \vdash (t = t)^*$.*

Proof Here $(t = t)^*$ is the formula $\exists y(t^*(\vec{x}, y) \wedge t^*(\vec{x}, y))$, which is equivalent to $\exists y\, t^*(\vec{x}, y)$. Thus, the result follows immediately from the fact that Δ entails that t^* is a functional relation. □

LEMMA 4.5.11 *Let* $* : \Sigma \to \Sigma'$ *be a reconstrual, and let* Δ *be the admissibility conditions for the function symbols in* Σ. *Then* $\Delta, \phi(s)^*, (s = t)^* \vdash \phi(t)^*$.

Proof Recall that $\phi(s)^*$ is the formula $\exists y(s^*(\vec{x}, y) \wedge \phi^*(y))$, and similarly for $\phi(t)^*$. Thus, we need to show that

$$\Delta, \exists y(s^*(\vec{x}, y) \wedge \phi^*(y)), \exists z(s^*(\vec{x}, z) \wedge t^*(\vec{x}, z)) \vdash \exists w(t^*(\vec{x}, w) \wedge \phi^*(w)).$$

The key fact, again, is that Δ implies that t^* and s^* are functional relations. We can then argue intuitively: holding $\vec{x}$ fixed, from $\exists y(s^*(\vec{x}, y) \wedge \phi^*(y))$ and $\exists z(s^*(\vec{x}, z) \wedge t^*(\vec{x}, z))$, we are able to conclude this y and z are the same and, thus, that $\exists w(t^*(\vec{x}, w) \wedge \phi^*(w))$. □

Recall that a reconstrual $* : \Sigma \to \Sigma'$ maps a formula such as $t(\vec{x}) = y$ to a formula $t^*(\vec{x}, y)$. Here the variable y is chosen arbitrarily, but in such a manner as not to conflict with any variables already in use. Intuitively, this y is the only new free variable in t^*. We now validate that intuition.

LEMMA 4.5.12 *Let* $* : \Sigma \to \Sigma'$ *be a reconstrual. Then for each term* t *of* Σ, $FV(t^*) = FV(t) \cup \{y\}$.

Proof Base cases: If t is a variable x, then $t^*(x, y) \equiv (x = y)$. In this case, $FV(t^*(x, y)) = FV(t) \cup \{y\}$. If t is a constant symbol c, then t^* is the formula $c = y$, and $FV(t^*) = FV(c) \cup \{y\}$.

Inductive case: Suppose that the result holds for $t_1, \ldots, t_n$. Then

$$\begin{aligned} FV(f(t_1, \ldots, t_n)) &= FV(t_1) \cup \cdots \cup FV(t_n) \\ &= FV(t_1^*) \cup \cdots FV(t_n^*). \end{aligned}$$

Recall that $f(t_1, \ldots, t_n)^*$ is defined as the formula

$$\exists z_1 \cdots \exists z_n \, (t_1(\vec{x}, z_1) \wedge \cdots t_n(\vec{x}, z_n) \wedge f(z_1, \ldots, z_n, y)),$$

from which it can easily be seen that

$$FV(f(t_1, \ldots t_n)^*) = FV(t_1^*) \cup \cdots \cup FV(t_n^*) \cup \{y\}.$$

□

LEMMA 4.5.13 *Let* $* : \Sigma \to \Sigma'$ *be a reconstrual. Then for each formula* ϕ *of* Σ, $FV(\phi^*) = FV(\phi)$.

Proof We prove this by induction on the construction of ϕ. Base case: Suppose that ϕ is the formula $s = t$, where s and t are terms. Then ϕ^* is the formula

$$\exists y(s^*(\vec{x}, y) \wedge t^*(\vec{x}, y)).$$

By the previous lemma, y is the only free variable in s^* that doesn't occur in s, and similarly for t^* and t. Therefore, $FV(\phi^*) = FV(\phi)$.

Base case: Suppose that ϕ is the formula $p(t_1, \ldots, t_n)$, where p is a relation symbol, and $t_1, \ldots, t_n$ are terms. Here the free variables in ϕ are just all those free in the terms t_i. Moreover, $p(t_1, \ldots, t_n)^*$ is the formula that says: there are $z_1, \ldots, z_n$ such that $t_i^*(\vec{x}, z_i)$

and $p^*(z_1, \dots, z_n)$. By the previous lemma, $FV(t_i^*(\vec{x}, z_i)) = FV(t_i) \cup \{z_i\}$. Therefore, $p(t_1, \dots, t_n)^*$ has the same free variables as $p(t_1, \dots, t_n)$.

Inductive cases: The cases for the Boolean connectives are easy, and are left to the reader. Let's just check the case of the universal quantifier. Suppose that the result is true for ϕ. Then

$$FV((\forall x\phi)^*) = FV(\forall x\phi^*) = FV(\phi^*)\backslash\{x\} = FV(\phi)\backslash\{x\} = FV(\forall x\phi).$$

□

THEOREM 4.5.14 (Substitution) *Let $* : \Sigma \to \Sigma'$ be a reconstrual, and let Δ be the admissibility conditions for function symbols in Σ. Then for any formulas ϕ and ψ of Σ, if $\phi \vdash \psi$, then $\Delta, \phi^* \vdash \psi^*$.*

Proof We prove this by induction on the definition of the relation $\vdash$. The base case is the rule of assumptions: show that when $\phi \vdash \phi$, then also $\Delta, \phi^* \vdash \phi^*$. However, the latter follows immediately by the rule of assumptions, plus monotonicity of $\vdash$.

The clauses for the Boolean connectives follow immediately from the fact that $*$ is compositional. We will now look at the clause for $\forall$-intro. Suppose that $\Gamma \vdash \forall x\phi$ results from $\Gamma \vdash \phi$, where x is not free in Γ. Assume that $\Delta, \Gamma^* \vdash \phi^*$. By Lemma 4.5.13, x is not free in Γ^*. Moreover, since the admissibility conditions are sentences, x is not free in Δ. Therefore, $\Delta, \Gamma^* \vdash \forall x\phi^*$, and since $(\forall x\phi)^* \equiv \forall x\phi^*$, it follows that $\Delta, \Gamma^* \vdash (\forall x\phi)^*$. □

We began this section with a discussion of various relations bewteen theories that have been interesting to philosophers of science – e.g., equivalence, reducibility, conservative extension. We now look at how such relations might be represented as certain kinds of translations between theories. We begin by considering the proposal that theories are **equivalent** just in case they are **intertranslatable**. The key here is in specifying what is meant by "intertranslatable." Do we only require a pair of translations $F : T \to T'$ and $G : T' \to T$, with no particular relation between F and G? Or do we require more? The following condition requires that F and G are inverses, relative to the notion of "sameness" of formulas internal to the theories T and T'. To be more specific, two formulas ϕ and ϕ' are the "same" relative to theory T just in case $T \vdash \phi \leftrightarrow \phi'$.

DEFINITION 4.5.15 Let T be a Σ-theory and T' a Σ'-theory. Then T and T' are said to be **strongly intertranslatable** or **homotopy equivalent** if there are translations $F : T \to T'$ and $G : T' \to T$ such that

$$T \vdash \phi \leftrightarrow GF\phi \qquad \text{and} \qquad T' \vdash \psi \leftrightarrow FG\psi, \tag{4.1}$$

for every Σ-formula ϕ and every Σ'-formula ψ.

The conditions (4.1) can be thought of as requiring the translations $F : T \to T'$ and $G : T' \to T$ to be "almost inverse" to one another. Note, however, that F and G need not be literal inverses. The Σ-formula $GF\phi$ is not required to be *equal* to the Σ-formula ϕ. Rather, these two formulas are merely required to be *equivalent* according to the theory T.

DISCUSSION 4.5.16 Let T and T' be theories in a common signature Σ. It should be fairly obvious that if T and T' are logically equivalent, then they are homotopy equivalent. Indeed, it suffices to let both F and G be the identity reconstrual on the signature Σ.

It should also be obvious that not all homotopy equivalent theories are logically equivalent – not even theories in the same signature. For example, let $\Sigma = \{p, q\}$ be a propositional signature, let T be the theory with axiom p, and let T' be the theory with axiom q. Obviously T and T' are not logically equivalent. However, the reconstrual $Fp = q$ shows that T and T' are homotopy equivalent.

DISCUSSION 4.5.17 One might legitimately wonder: what's the motivation for the definition of homotopy equivalence, with a word "homotopy" that is not in most philosophers' active vocabulary? One might also wonder, more generally: what is the right method for deciding on an account of equivalence? What is at stake, and how do we choose between various proposed explications? Are we supposed to have strong intuitions about what "equivalent" really means? Or, at the opposite extreme, is the definition of "equivalent" merely a convention to be judged by its utility?

These are difficult philosophical questions that we won't try to answer here (but see Chapter 8). However, there is both a historical and a mathematical motivation for the definition of homotopy equivalence. The historical motivation for this definition is its appearance in various works of logic and philosophy of science beginning in the 1950s. As for mathematical motivation, the phrase "homotopy equivalence" originally comes from topology, where it denotes a kind of "sameness" that is weaker than the notion of a homeomorphism. Interestingly, the idea of weakening isomorphism is also particularly helpful in category theory. In category theory, the natural notion of "sameness" of categories is not isomorphism, but categorical equivalence. Recall that two categories $\mathbf{C}$ and $\mathbf{D}$ are equivalent just in case there is a pair of functors $F : \mathbf{C} \to \mathbf{D}$ and $G : \mathbf{D} \to \mathbf{C}$ such that both FG and GF are naturally isomorphic to the respective identity functors (see 3.3.6). In the case of Makkai and Reyes' logical categories, the natural notion of equivalence is simply categorical equivalence (see Makkai and Reyes, 1977).

At this point, we will set aside further discussion of equivalence until Section 4.6. We will now turn to some other relations between theories.

Suppose that T is a theory in signature Σ, and T' is a theory in signature Σ', where $\Sigma \subseteq \Sigma'$. We say that T' is an **extension** of T just in case: if $T \vdash \phi$, then $T' \vdash \phi$, for all sentences ϕ of Σ. An extension of of a theory amounts to the addition of new concepts – i.e., new vocabulary – to a theory. (Here we are using the word "concepts" in a nontechnical sense. To be technically precise, a conservative extension of a theory results when new symbols are added to the signature Σ.) In the development of the sciences, there can be a variety of reasons for adding new concepts; e.g., we might use them as a convenient shorthand for old concepts, or we might feel the new to expand our conceptual repertoire. For example, many physicists would say that the concept of a "quantum state" is a genuinely novel addition to the stock of concepts used in classical physics.

One very interesting question in philosophy of science is whether there are any sorts of "rules" or "guidelines" for the expansion of our conceptual repertoire. Must the new concepts be connected to the old ones? And if so, what sorts of connections should we hope for them to have?

A **conservative extension** of a theory T adds new concepts, but without in any way changing the logical relations between old concepts. In the case of formal theories, there are two extreme cases of a conservative extension: (1) the new vocabulary is shorthand for old vocabulary, and (2) the new vocabulary is unrelated to the old vocabulary. The paradigm example of the former (new vocabulary as shorthand) is a definitional extension of a theory, as discussed in the preceding paragraphs. As an example of the latter (new vocabulary unrelated), consider first the theory T in the empty signature Σ which says that "there are exactly two things." Now let T' have the same axiom as T, but let it be formulated in a signature $\Sigma' = \{p\}$, where p is a unary predicate symbol. Let $F : T \to T'$ be the translation given by the inclusion $\Sigma \subseteq \Sigma'$. It's clear, then, that F is a conservative translation.

The example we have just given is somewhat atypical, since the new theory T' says nothing about the new vocabulary $\Sigma' \backslash \Sigma$. However, the point would be unchanged if, for example, we equipped T' with the axiom $\forall x p(x)$.

DEFINITION 4.5.18 A translation $F : T \to T'$ is said to be **conservative** just in case: if $T' \vdash F\phi$, then $T \vdash \phi$.

The idea behind this definition of a conservative extension is that the target theory T' adds no new claims that can be formulated in the language Σ of the original theory. In other words, if T' says that some relation holds between sentences $F\phi_1$ and $F\phi_2$, then T already asserts that this relation holds. Of course, T' might say nontrivial things in its new vocabulary, i.e., in those Σ'-sentences that are not in the image of the mapping F.

Example 4.5.19 There are many intuitive examples of conservative extensions in mathematics. For example, the theory of the integers is a conservative extension of the theory of natural numbers, and the theory of complex numbers is a conservative extension of the theory of real numbers. ⌟

EXERCISE 4.5.20 Suppose that $F : T \to T'$ is conservative. Show that if T is consistent, then T' is consistent.

EXERCISE 4.5.21 Suppose that T is a consistent and complete theory in signature Σ. Let $\Sigma \subseteq \Sigma'$, and let T' be a consistent theory in Σ'. Show that if T' is an extension of T, then T' is a conservative extension of T.

EXERCISE 4.5.22 Let T be the theory from the previous example, and let $\Sigma' = \{p\}$, where p is a unary predicate symbol. Which theories in Σ' are extensions of T? Which of these extensions is conservative? More difficult: classify all extensions of T in the language Σ', up to homotopy equivalence. (In other words, consider two extensions to be the same if they are homotopy equivalent. Hint: consider the question, "how many p are there?")

A conservative translation $F : T \to T'$ is like a monomorphism from T to T'. Thus, we might also be interested in a dual sort of notion – something like an epimorphism from T to T'. As with propositional theories, it works well to consider a notion of surjectivity up to logical equivalence. Borrowing terminology from category theory, we call this notion "essential surjectivity."

DEFINITION 4.5.23 Let $F : T \to T'$ be a translation between theories. We say that F is **essentially surjective** (abbreviated **eso**) just in case for each Σ'-formula ψ, there is a Σ-formula ϕ such that $T \vdash \psi \leftrightarrow F\phi$.

The idea behind an essentially surjective translation is that the domain T is ideologically as rich as the codomain theory T'. We are using "ideology" here in the sense of Quine, i.e., the language in which a theory is formulated. If $F : T \to T'$ is essentially surjective, then (up to logical equivalence), T can express all of the concepts that T' can express.

A paradigm example of an essentially surjective translation is a "specialization" of a theory, i.e., where we add some new axioms, but without adding any new vocabulary. Indeed, suppose that T is a theory in Σ, and that T' results from adding some axioms to T. Then the identity reconstrual $I : \Sigma \to \Sigma$ yields an essentially surjective translation $I : T \to T'$. Of course, there are other sorts of essentially surjective translations.

EXERCISE 4.5.24 Let $\Sigma = \{p\}$, where p is a unary predicate, and let $\Sigma' = \{r\}$, where r is a binary relation. Let T be the empty theory in Σ, and let T' be the theory in Σ' that says that r is symmetric, i.e., $r(x, y) \to r(y, x)$. Let $F : \Sigma \to \Sigma'$ be the reconstrual that takes p to $\exists z\, r(x, z)$. Is F essentially surjective? (This exercise will be a lot easier to answer after Chapter 6.)

EXERCISE 4.5.25 Let Σ be the signature with a single unary predicate symbol p, and let Σ' be the empty signature. Let T' be the theory in Σ' that says "there are exactly two things," and let T be the extension of T' in Σ that also says "there is a unique p." Is there an essentially surjective translation $F : T \to T'$? (This exercise will be a lot easier after Chapter 6.)

In the case of propositional theories, we saw that a translation is an equivalence iff it is conservative and essentially surjective. We now show the same for first-order theories.

PROPOSITION 4.5.26 *Suppose that $F : T \to T'$ is one-half of a homotopy equivalence. Then F is conservative and essentially surjective.*

Proof The proof here is structurally identical to the one for propositional theories. Suppose that $G : T' \to T$ is the other half of a homotopy equivalence so that $T \vdash \phi \leftrightarrow GF\phi$ and $T' \vdash \psi \leftrightarrow FG\psi$. To see that F is conservative, suppose that $T' \vdash F\phi$. Then $T \vdash GF\phi$, and since $T \vdash \phi \leftrightarrow GF\phi$, it follows that $T \vdash \phi$. Therefore, F is conservative. To see that F is essentially surjective, let ψ be a Σ'-formula. Then $G\psi$ is a Σ-formula, and we have $T' \vdash \psi \leftrightarrow FG\psi$. Therefore, F is essentially surjective. □

PROPOSITION 4.5.27 *Suppose that $F : T \to T'$ is conservative and essentially surjective. Then F is one-half of a homotopy equivalence.*

Proof Again, the proof here is structurally identical to the proof in the propositional case.

Fix a relation symbol p of Σ'. Since F is essentially surjective, there is a formula ϕ_p of Σ such that $T' \vdash p \leftrightarrow F\phi_p$. Define $Gp = \phi_p$. Thus, for each relation symbol p, we have $T' \vdash p \leftrightarrow FGp$ by definition. And since FG is (by definition) compositional, $T' \vdash \psi \leftrightarrow FG\psi$ for all formulas ψ of Σ'.

Now we claim that G is a translation from T' into T. Indeed, if $T' \vdash \psi$, then $T' \vdash FG\psi$, and since F is conservative, $T \vdash G\psi$. Therefore, $G : T' \to T$ is a translation.

Finally, given an arbitrary formula ϕ of Σ, we have $T' \vdash F\phi \leftrightarrow FGF\phi$, and hence $T' \vdash F(\phi \leftrightarrow GF\phi)$. Since F is conservative, it follows that $T \vdash \phi \leftrightarrow GF\phi$. Therefore, F and G form a homotopy equivalence. □

DISCUSSION 4.5.28 Using translations between theories as arrows, we could now define a category **Th** of first-order theories, and we could explore the features of this category. However, we will resist this impulse – because it turns out that this category isn't very interesting.

DISCUSSION 4.5.29 Does the notion of translation capture every philosophically interesting relation between theories? There are few questions here: First, can every interesting relation between theories be explicated syntactically? And if the answer to the first question is yes, then can every such relation be described as a translation? And if the answer to that question is yes, then have we given an adequate account of translation?

Recall that Carnap (1934) seeks a theory of science that uses only syntactic concepts. If we were to follow Carnap's lead, then we would have to answer yes to the first question. But of course, many philosophers of science have convinced themselves that the first question must receive a negative answer. Indeed, some philosophers of science claimed that the interesting relations between theories (e.g., equivalence, reducibility) cannot be explicated syntactically. We will return later to this claim.

4.6 Definitional Extension and Equivalence

Mathematicians frequently define new concepts out of old ones, and logicians have only begun to understanding the varieties of ways that mathematicians do so. We do have a sense that not all definitions are created equal. On the one hand, definitions can seem quite trivial, e.g., when we come up with a new name for an old concept. On the other hand, some definitions are overtly inconsistent. For example, if we said "let *n* be the largest prime number," then we could prove both that there is a largest prime number and that there is not. The goal of a logical theory of definition is to steer a course between these two extremes – i.e., to account for those definitions that are both fruitful and safe.

In this section, we'll look at some of the simplest kinds of definitions. Our general setup will consist of a pair of signatures Σ and Σ^+ with $\Sigma \subseteq \Sigma^+$. Here we think of Σ as "old concepts" and we think of $\Sigma^+ \backslash \Sigma$ as "new concepts."

DEFINITION 4.6.1 If p is a relation symbol in Σ^+, then an **explicit definition of p in terms of** Σ is a Σ^+-sentence of the form

$$\forall\vec{x}(p(\vec{x}) \leftrightarrow \phi(\vec{x})),$$

where $\phi(\vec{x})$ is a Σ-formula.

Here p can be thought of as "convenient shorthand" for the formula ϕ, which might itself be quite complex. For example, from the predicates "is a parent" and "is a male," we could explicitly define a predicate "is a father." Of course, the definition itself is a sentence in the larger signature Σ^+ and not in the smaller signature Σ.

DISCUSSION 4.6.2 What are we doing when we define new concepts out of old ones? Some philosophers might worry that defining a new concept amounts to making a theoretical commitment – to the existence of some worldly structure corresponding to that concept. For example, suppose that we initially have a theory with the concepts $\mathsf{male}(x)$ and $\mathsf{parent}(x)$. If we define $\mathsf{father}(x)$ in terms of these two original concepts, then are we committing to the existence of some further worldly structure, viz. the property of fatherhood?

The default answer in first-order logic is no. Using a predicate symbol r does not amount to any kind of postulating of worldly structure corresponding to r. Accordingly, adding definitions to a theory does not change the content of that theory – it only changes the resources we have for expressing that content. (We don't mean to say that these views are uncontroversial or mandatory.)

Not only can we define new relations; we can also define new functions and constants. Certainly a function can be defined in terms of other functions. For example, if we begin with functions g and f, then we can define a composite function $g \circ f$. In fact, this composite $g \circ f$ can be defined explicitly by the formula

$$((g \circ f)(x) = y) \ \leftrightarrow\ \exists z((f(x) = z) \wedge (g(z) = y)).$$

Similarly, a constant symbol c can be defined in terms of a function symbol f and other constants $d_1, \ldots, d_n$, namely

$$c \ := \ f(d_1, \ldots, d_n).$$

However, these ways of defining functions in terms of functions, and defining constants in terms of functions and constants, can be subsumed into a more general way of defining functions and constants in terms of relations.

DEFINITION 4.6.3 An explicit definition of an n-ary function symbol $f \in \Sigma^+$ in terms of Σ is a Σ^+-sentence of the form

$$\forall\vec{x}\forall y(f(\vec{x}) = y \leftrightarrow \phi(\vec{x}, y)), \tag{4.2}$$

where $\phi(\vec{x}, y)$ is a Σ formula.

DEFINITION 4.6.4 An explicit definition of a constant symbol $c \in \Sigma^+$ is a Σ^+-sentence of the form

$$\forall y(y = c \leftrightarrow \psi(y)), \tag{4.3}$$

where $\psi(y)$ is a Σ-formula.

Although they are Σ^+-sentences, (4.2) and (4.3) have consequences in the signature Σ. In particular, (4.2) implies $\forall \vec{x} \exists! y \phi(\vec{x}, y)$ and (4.3) implies $\exists! y \psi(y)$. These two sentences are called the **admissibility conditions** for the explicit definitions (4.2) and (4.3).

DEFINITION 4.6.5 A **definitional extension** of a Σ-theory T to the signature Σ^+ is a Σ^+-theory

$$T^+ = T \cup \{\delta_s : s \in \Sigma^+ \backslash \Sigma\},$$

that satisfies the following two conditions. First, for each symbol $s \in \Sigma^+ \backslash \Sigma$, the sentence δ_s is an explicit definition of s in terms of Σ. And second, if s is a constant symbol or a function symbol and α_s is the admissibility condition for δ_s, then $T \vdash \alpha_s$.

Example 4.6.6 Let $\Sigma = \{p\}$, where p is a unary predicate symbol, and let T be *any* theory in Σ. We can then define a relation r by means of the formula

$$r(x, y) \leftrightarrow (p(x) \leftrightarrow p(y)).$$

It's easy to see that r is an equivalence relation. Thus, every unary predicate symbol defines a corresponding equivalence relation. In fact, this equivalence relation r has precisely two equivalence classes.

The converse is not exactly true. In fact, suppose that a theory T entails that r is an equivalence relation with exactly two equivalence classes. One might try to define a predicate p so that the first equivalence class consists of elements satisfying p, and the second equivalence class consists of elements not satisfying p. However, this won't work, because the relation r itself does not provide the resources to name the individual classes. Intuitively speaking, r can't tell the difference between the two equivalence classes, but p can, and therefore p cannot be defined from r. We'll be able to see this fact more clearly after Chapter 6. ⌟

Example 4.6.7 The following example is from Quine and Goodman (1940). Let $\Sigma = \{r\}$, where r is a binary relation symbol. Now define a new relation symbol s by setting

$$s(x, y) \leftrightarrow \forall w(r(x, w) \to r(y, w)).$$

Then it follows that s is a transitive relation, i.e.,

$$\vdash (s(x, y) \wedge s(y, z)) \to s(x, z).$$

⌟

Example 4.6.8 Let T be the theory of Boolean algebras. Then one can define a relation symbol $\leq$ by setting

$$x \leq y \leftrightarrow x \wedge y = y.$$

It follows that $\leq$ is a partial order. ⌟

Let T^+ be a definitional extension of T. We now define two translations $I : T \to T^+$ and $R : T^+ \to T$. The translation $I : T \to T^+$ is simply the inclusion: it acts as the identity on elements of the signature Σ. The latter we define as follows: for each symbol r in $\Sigma^+ \backslash \Sigma$, let $Rr = \theta_r$, where θ_r is the Σ-formula in the explicit definition

$$T^+ \vdash \forall \vec{x}(r(\vec{x}) \leftrightarrow \theta_r(\vec{x}).$$

For $r \in \Sigma$, let $Rr = r$.

DISCUSSION 4.6.9 The translation $R : T^+ \to T$ is an example of a **reduction** of the theory T^+ to the theory T. Here we simply replace the definiendum r with its definiens θ_r. However, this particular R has another feature that not all reductions have – namely, it's an equivalence between the theories T^+ and T (as we show in the subsequent lemmas).

This kind of strict reduction is similar to what Carnap hoped to achieve in the *Aufbau* – and for which he was so severely critized by Quine. However, it should be noted that Carnap's permissible constructions were stronger than explicit definitions in the sense we've explained here. At the very least, Carnap permitted maneuvers such as "extension by abstraction," which are akin to the "extensions by sorts" that we consider in the following chapter.

Similarly, advocates of the old-fashioned mind–brain identity theory presumably believed that folk psychology (to the extent that it is accurate) could be reduced, in this strict sense, to neuroscience, and perhaps ultimately to fundamental physics.

Scientists do often talk about one theory T^+ being reducible to another T, e.g., thermodynamics being reducible to statistical mechanics. However, it's beyond doubtful that all such cases of successful reduction could be faithfully modelled as a simple expansion of definitions. We don't think, however, that the moral is that philosophers of science should resort to vague and imprecise accounts of reduction. Instead, they should find more sophisticated tools for their explications.

We now show that the pair I, R form a homotopy equivalence between T and T^+. For this we need a few auxiliary lemmas.

LEMMA 4.6.10 *For any term $t(\vec{x})$ of Σ^+, we have $T^+ \vdash (t(\vec{x}) = y) \leftrightarrow Rt(\vec{x}, y)$.*

Proof We prove this by induction on the construction of t. In the case that t is a variable x, the claim is

$$T^+ \vdash (x = y) \leftrightarrow (x = y),$$

which obviously holds. Now suppose that t is the term $f(t_1, \dots, t_n)$ and that the result holds for the terms $t_1, \dots, t_n$. That is,

$$T^+ \vdash (t_i(\vec{x}) = z_i) \leftrightarrow Rt_i(\vec{x}, z_i). \tag{4.4}$$

Since T^+ defines f in terms of Rf, we also have

$$T^+ \vdash (f(z_1, \dots, z_n) = y) \leftrightarrow Rf(z_1, \dots, z_n, y). \tag{4.5}$$

By the definition of R on terms, $R(f(t_1,\ldots,t_n))(\vec{x},y)$ is the formula

$$\exists z_1\cdots\exists z_n(Rt_1(\vec{x},z_1)\wedge Rt_n(\vec{x},z_n)\wedge Rf(z_1,\ldots,z_n,y)).$$

Thus, (4.4) and (4.5) imply that

$$T^+\ \vdash\ (f(t_1(\vec{x}),\ldots,t_n(\vec{x}))=y)\leftrightarrow R(f(t_1,\ldots,t_n))(\vec{x},y).$$

□

LEMMA 4.6.11 *For any Σ^+-formula ϕ, we have $T^+\vdash\phi\leftrightarrow R\phi$.*

Proof We prove this by induction on the construction of ϕ. Since R is defined compositionally on formulas, it will suffice to establish the two base cases.

1. Suppose first that ϕ is the formula $s(\vec{x})=t(\vec{x})$, in which case $R\phi$ is the formula $\exists y(Rs(\vec{x},y)\wedge Rt(\vec{x},y))$. By the previous result,

$$T^+\ \vdash\ (t(\vec{x})=y)\leftrightarrow Rt(\vec{x},y),$$

and

$$T^+\ \vdash\ (s(\vec{x})=y)\leftrightarrow Rs(\vec{x},y).$$

By assumption, we have

$$T^+\ \vdash\ \exists y(s(\vec{x})=y\wedge t(\vec{x})=y),$$

and the result immediately follows.

2. Suppose now that ϕ is the formula $p(t_1,\ldots,t_n)$, in which case $R\phi$ is the formula

$$\exists y_1\cdots\exists y_n(Rt_1(\vec{x},y_1)\wedge\cdots\wedge Rt_n(\vec{x},y_n)\wedge Rp(y_1,\ldots,y_n)).$$

By the previous result again,

$$T^+\ \vdash\ (t_i(\vec{x})=y_i)\leftrightarrow Rt_i(\vec{x},y_i).$$

Moreover, since T^+ explicitly defines p in terms of Rp, we have

$$T^+\ \vdash\ p(\vec{y})\leftrightarrow Rp(\vec{y}).$$

The result follows immediately. □

LEMMA 4.6.12 *For any Σ^+-formula ϕ, if $T^+\vdash R\phi$ then $T\vdash R\phi$.*

Proof To say that $T^+\vdash R\phi$ means that there is a finite family $\theta_1,\ldots,\theta_n$ of axioms of T^+ such that $\theta_1,\ldots,\theta_n\vdash R\phi$. By the substitution theorem (4.5.14),

$$\Delta,R\theta_1,\ldots,R\theta_n\ \vdash\ RR\phi,$$

where Δ consists of the admissibility conditions for function symbols in Σ^+. Since T^+ is a definitional extension, T implies the admissibility conditions in Δ. Moreover, since $RR\phi=\phi$, we have

$$T,R\theta_1,\ldots,R\theta_n\ \vdash\ R\phi.$$

Thus, it will suffice to show that $T\vdash R\theta_i$, for each i.

Fix i and let $\theta \equiv \theta_i$. Now, θ is either an axiom of T or an explicit definition of a symbol $s \in \Sigma^+ \backslash \Sigma$. If θ is an axiom of T, then it's also a formula in signature Σ, in which case $R\theta = \theta$ and $T \vdash R\theta$. If θ is an explicit definition of a symbol s, then $R\theta$ is the tautology $Rs \leftrightarrow Rs$, and $T \vdash R\theta$. □

PROPOSITION 4.6.13 *If T^+ is a definitional extension of T, then $I : T \to T^+$ and $R : T^+ \to T$ form a homotopy equivalence.*

Proof Since the axioms of T are a subset of the axioms of T^+, it follows that I is a translation from T to T^+. Next we show that R is a translation from T^+ to T – i.e., that if $T^+ \vdash \phi$, then $T \vdash R\phi$. By the previous two lemmas, we have $T^+ \vdash \phi \leftrightarrow R\phi$, and if $T^+ \vdash \phi$, then $T \vdash \phi$. Thus, if $T^+ \vdash \phi$, then $T^+ \vdash R\phi$, and $T \vdash R\phi$.

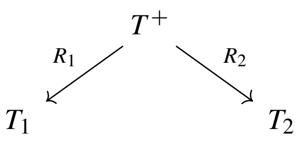
Next we show that $T \vdash \phi \leftrightarrow RI\phi$ and $T^+ \vdash \psi \leftrightarrow IR\psi$. For this, recall that both $I : \Sigma \to \Sigma^+$ and $R : \Sigma^+ \to \Sigma$ act as the identity on Σ-formulas. Thus, we immediately get $T \vdash \phi \leftrightarrow IR\phi$ for any Σ-formula ϕ. Furthermore, Lemma 4.6.11 entails that $T^+ \vdash \psi \leftrightarrow R\psi$. Since $R\psi$ is a Σ-formula, it follows that $T^+ \vdash \psi \leftrightarrow IR\psi$. □

COROLLARY 4.6.14 *If T^+ is a definitional extension of T, then T^+ is a conservative extension of T.*

The previous results show, first, that a definitional extension is conservative: it adds no new results in the old vocabulary. In fact, Proposition 4.6.13 shows that a definitional extension is, in one important sense, equivalent to the original theory. You may want to keep that fact in mind as we turn to a proposal that some logicians made in the 1950s and 1960s, and that was applied to philosophy of science by Glymour (1971). According to Glymour, two scientific theories should be considered equivalent only if they have a common definitional extension.

DEFINITION 4.6.15 Let T_1 be a Σ_1-theory and T_2 be a Σ_2-theory. Then T_1 and T_2 are said to be **definitionally equivalent** if there is a definitional extension T_1^+ of T_1 to the signature $\Sigma_1 \cup \Sigma_2$ and a definitional extension T_2^+ of T_2 to the signature $\Sigma_1 \cup \Sigma_2$ such that T_1^+ and T_2^+ are logically equivalent.

If T_1 and T_2 are definitionally equivalent, then they in fact have a **common definitional extension**, namely the theory $T^+ := \mathrm{Cn}(T_1^+) = \mathrm{Cn}(T_2^+)$. These three theories then form a span:



Here $R_i : T^+ \to T_i$ is the translation that results from replacing definienda in the signature $\Sigma_1 \cup \Sigma_2$ with their definiens in signature Σ_i. Note that if T_1 and T_2 are both Σ-theories (i.e., if they are formulated in the same signature), then T_1 and T_2 are definitionally equivalent if and only if they are logically equivalent.

Definitional equivalence captures a sense in which theories formulated in different signatures might nonetheless be theoretically equivalent. For example, although they are not logically equivalent, the theory of groups_1 and the theory of groups_2 are definitionally equivalent.

Example 4.6.16 Recall the two formulations of group theory from Example 4.5.3. Consider the following two $\Sigma_1 \cup \Sigma_2$-sentences.

$$\delta_{-1} := \forall x \forall y \big(x^{-1} = y \leftrightarrow (x \cdot y = e \wedge y \cdot x = e)\big)$$
$$\delta_e := \forall x \big(x = e \leftrightarrow \forall z (z \cdot x = z \wedge x \cdot z = z)\big).$$

The theory T_1 defines the unary function symbol $^{-1}$ with the sentence δ_{-1}, and the theory T_2 defines the constant symbol e with the sentence δ_e. One can verify that T_1 satisfies the admissibility condition for δ_{-1} and that T_2 satisfies the admissibility condition for δ_e. The theory of $\text{groups}_1 \cup \{\delta_{-1}\}$ and the theory of $\text{groups}_2 \cup \{\delta_e\}$ are logically equivalent. This implies that these two formulations of group theory are definitionally equivalent. ⌟

We're now ready for the first big result relating different notions of equivalence.

THEOREM 4.6.17 (Barrett) *Let T_1 and T_2 be theories with a common definitional extension. Then there are translations $F : T_1 \to T_2$ and $G : T_2 \to T_1$ that form a homotopy equivalence.*

Proof Let T^+ be a common definitional extension of T_1 and T_2. By Prop. 4.6.13, there are homotopy equivalences $I_1 : T_1 \to T^+$ and $R_2 : T^+ \to T_2$. Thus, $R_2 I_1 : T_1 \to T_2$ is a homotopy equivalence. □

We prove the converse of this theorem in 6.6.21.

DISCUSSION 4.6.18 In this section, we've discussed methods for defining relation, function, and constant symbols. It's commonly assumed, however, that other sorts of definitions are also possible. For example, we might define an exclusive "or" connective $\oplus$ by means of the recipe

$$\phi \oplus \psi \ \leftrightarrow \ (\phi \vee \psi) \wedge \neg(\phi \wedge \psi).$$

For more on the notion of defining new connectives, see Dewar (2018a).

The same might be said for quantifiers. Given existential quantifiers $\exists x$ and $\exists y$, we might introduce a new quantifier $\exists x \exists y$ over pairs. But does this new syntactic entity, $\exists x \exists y$, deserve to be called a "quantifier"?

Supposing that $\exists x \exists y$ does deserve to be called a quantifier, then we need to rethink the notion of the "ontological commitments" of a theory – and along with that, a whole slew of attitudes toward ontology that come along with it. It's common for philosophers of science to raise the question: "What are the ontological commitments of this theory?" The idea here is that if the scientific community accepts a theory, then we should accept that theory's ontological commitments. For example, some philosophers argue that we should believe in the existence of mathematical objects since our best scientific theories (such as general relativity and quantum mechanics) quantify over them. Others, such as

Field (1980) attempt to "nominalize" these theories – i.e., to reformulate them in such a way that they don't quantify over mathematical objects.

Both parties to this dispute about mathematical objects share a common presupposition: Once a theory is regimented in first-order logic, then its ontological commitments can be read off from the formalism. But this presupposition is brought into question by the fact that first-order theories can implicitly define new quantifiers. Thus, a theory might have *more* ontological commitments than are shown in its original quantifiers. Conversely, a theory isn't necessarily committed to the ontology encoded in its initial quantifiers. Those quantifiers might capture some derivative ontology, and the actual ontology might be captured by quantifiers that are defined in terms of those original quantifiers. In short, regimenting a theory in first-order logic does *not* settle all ontological disputes.

4.7 Notes

- Carnap (1935) gives a readable, nontechnical overview of his *Wissenschaftslogik* program. The amount of high-quality historical research on Carnap is on the steady rise. See, e.g., Friedman (1982); Awodey and Klein (2004); Andreas (2007); Creath and Friedman (2007); Hudson (2010); Friedman (2011). For the relevance of Carnap's views to contemporary issues, see, e.g., Price (2009); Blatti and Lapointe (2016).
- The substitution theorem is rarely proven in detail. One notable exception is Kleene (1952). We prove another, more general, version of the theorem in the following chapter.
- The word "reconstrual" comes from Quine (1975), where he uses it to propose a notion of theoretical equivalence. We find his notion to be far too liberal, as discussed in Barrett and Halvorson (2016a).
- **Definitional equivalence** and **common definition extension** have been part of the logical folklore since the 1960s, and many results about them have been proven – see, e.g., Hodges (1993, §2.6), de Bouvére (1965), Kanger (1968), Pinter (1978), Pelletier and Urquhart (2003), Andréka et al. (2005), and Friedman and Visser (2014) for some results. To our knowledge, Glymour was the first philosopher of science to recognize the significance of these notions for discussions of theoretical equivalence. For an application of definitional equivalence in recent metaphysical debate, see McSweeney (2016a).
- For overviews of recent work on scientific reduction, see Scheibe (2013); Van Riel and van Gulick (2014); Love and Hüttemann (2016); Hudetz (2018b). Nagel's pioneering work on the topic can be found in Nagel (1935, 1961). For recent discussions of Nagel's view, see Dizadji-Bahmani et al. (2010); Sarkar (2015). We discuss semantic accounts of reduction in Chapter 6.