GAMUT

and renumber all of the lines in the derivation of $\phi \to \psi$: all numbers (don't forget those after the formulas) are raised by m. And then we add an application of E→ at the bottom, deriving the conclusion ψ:

1.	.	
	.	
	.	
	.	
m.	ϕ	
m + 1.	.	
	.	
	.	
m + n.	$\phi \to \psi$	
m + n + 1.	ψ	E→, m + n, m

The end result is a derivation of ψ.

All of the rules apply equally with premises. As an example, we shall prove (vi*) in the following form:

(vi*) $\psi_1, \ldots, \psi_n \vdash \neg\neg\phi$ iff $\psi_1, \ldots, \psi_n \vdash \phi$

Proof ⇒: Suppose we have a derivation of $\neg\neg\phi$ from $\psi_1, \ldots, \psi_n$. This can very easily be turned into a derivation of ϕ from $\psi_1, \ldots, \psi_n$ by adding one new line in which the conclusion ϕ is drawn from $\neg\neg\phi$ by means of the $\neg\neg$-rule.

⇐: The required alterations can be read from the following schema:

1.	ψ_1	
	.	
	.	
	.	
n.	ψ_n	
	.	
	.	
	.	
m.	ϕ	
m + 1.	$\neg\phi$	assumption
m + 2.	$\bot$	E¬, m + 1, m
m + 3.	$\neg\neg\phi$	I¬

□

Exercise 16

Show that $\vdash \phi \wedge \psi$ iff $\vdash \phi$ and $\vdash \psi$.

4.4 Soundness and Completeness

In this section we shall (without giving rigorous proofs) go into the connections between the semantic and syntactic approaches to logical inference, that

is, between $\vDash$ and $\vdash$. As we have said, these amount to the same thing. Or to put it more precisely, for any sentences $\phi_1, \ldots, \phi_n, \psi$ in any language L of predicate logic, we have $\phi_1, \ldots, \phi_n \vDash \psi$ if and only if $\phi_1, \ldots, \phi_n \vdash \psi$. This can be divided into two implications, one in each direction. We shall treat them separately, formulating them as two theorems.

Theorem 14 (*Soundness Theorem for Predicate Logic*)

For all sentences $\phi_1, \ldots, \phi_n, \psi$ (in some language L of predicate logic), if $\phi_1, \ldots, \phi_n \vdash \psi$, then $\phi_1, \ldots, \phi_n \vDash \psi$ too.

Strong

Theorem 15 (*Completeness Theorem for Predicate Logic*)

For all sentences $\phi_1, \ldots, \phi_n, \psi$ (in some language L of predicate logic), if $\phi_1, \ldots, \phi_n \vDash \psi$, then $\phi_1, \ldots, \phi_n \vdash \psi$ too.

These theorems are primarily statements about the rules we have given for the system of natural deduction. The soundness theorem establishes that the rules are sound: applying them to some premises $\phi_1, \ldots, \phi_n$, all of which are true in some model **M**, can only give rise to conclusions which are themselves true in **M**. In order to prove this theorem (which we shall not do), it is sufficient to check the above for each of the rules in turn. The introduction rule for $\wedge$ is sound, for example, since if both $V_M(\phi) = 1$ and $V_M(\psi) = 1$, then we can be sure that $V_M(\phi \wedge \psi) = 1$ too. The proof for the other rules poses no real problems, although there are a few complications in the rules for the quantifiers, which we encountered in connection with the restrictions to the rules. The soundness theorem assures us that the restrictions are sufficient to block all undesirable conclusions which might otherwise be drawn. In the special case of $n = 0$, it can be seen that the soundness theorem reduces to: if ϕ can be derived without premises, then ϕ is universally valid.

The completeness theorem assures us that the rules are complete in the sense that if $\phi_1, \ldots, \phi_n/\psi$ is valid, i.e., if $\phi_1, \ldots, \phi_n \vDash \psi$, then there are enough rules to enable us to derive ψ from $\phi_1, \ldots, \phi_n$. In other words, the rules are in themselves sufficient to generate all valid argument schemata; nothing has been forgotten. It is clear that this result is much less obvious than the soundness theorem, even if we thought we could obtain all valid argument schemata while forming the rules, and in particular, that we could derive all tautologies and other universally valid formulas without premises (see the discussion on the EFSQ and $\neg\neg$-rules). And this result is not only less obvious, it is also less easily proved.

But the soundness and completeness theorems are not only statements about the derivation rules. They also say something about semantics, about the concept of semantic validity. What is characteristic of derivation rules is that they leave absolutely no room for doubt about what combinations of symbols are proper derivations and what combinations are not. This is true of natural deduction, but it is equally essential to other existing formal proof systems. What soundness and completeness theorems say is that valid argument

argued

decidability where
D ⊢ S a finite
proof of ϕ from Γ

schemata are precisely those which can be obtained as derivations in the formal system in question. This is by no means always the case: it holds for predicate logic, but it does not hold for second-order logic or for mathematics in general (for more on this, see below). AWK

It should be noted that the completeness theorem in no way contradicts Church's Theorem on the undecidability of predicate logic, which was briefly mentioned in §4.2. If a given argument schema happens to be valid, then we are assured that there is some finite derivation of its conclusion from its premises. So we have a method at our disposal which is guaranteed to show sooner or later that the schema is valid: we just start generating derivations and wait until the right one turns up. The problem is with the schemata which are not valid; we have no method which is guaranteed to discover this for us. Generating derivations will not help us here, since in that case we would have to wait to make sure that the argument schema does *not* turn up. And since there are infinitely many possible derivations, we could never be sure.

The completeness theorem can also be presented in another form which is of some interest. But first we will introduce the concept of consistency and prove a few simple things about it.

Means: ⊢ every formula

Definition 4

$\phi_1, \ldots, \phi_n$ is said to be Syntactically *inconsistent* if $\phi_1, \ldots, \phi_n \vdash \perp$; $\phi_1, \ldots, \phi_n$ is said to be *consistent* if it is not inconsistent, that is, if $\phi_1, \ldots, \phi_n \nvdash \perp$.

Theorem 16

(a) $\phi_1, \ldots, \phi_n, \psi$ is inconsistent iff $\phi_1, \ldots, \phi_n \vdash \neg\psi$.
(b) $\phi_1, \ldots, \phi_n, \psi$ is consistent iff $\phi_1, \ldots, \phi_n \nvdash \neg\psi$.
(c) $\phi_1, \ldots, \phi_n, \neg\psi$ is consistent iff $\phi_1, \ldots, \phi_n \nvdash \psi$.

Proof (a) $\Rightarrow$: Suppose $\phi_1, \ldots, \phi_n, \psi$ is inconsistent, that is, suppose $\phi_1, \ldots, \phi_n, \psi \vdash \perp$. Then there is a derivation of $\perp$ from $\phi_1, \ldots, \phi_n, \psi$. This derivation can be converted into a derivation of $\neg\psi$ from $\phi_1, \ldots, \phi_n$ by adding I$\neg$ as a last step:

1.	ϕ_1	assumption
	.	
	.	
	.	
n.	ϕ_n	assumption
n + 1.	ψ	assumption
	.	
	.	
	.	
m.	$\perp$	
m + 1.	$\neg\psi$	I$\neg$

$\Leftarrow$: Suppose $\phi_1, \ldots, \phi_n \vdash \neg\psi$. Then a derivation of $\neg\psi$ from $\phi_1, \ldots, \phi_n$ is given. Now form a derivation starting with the assumptions $\phi_1, \ldots, \phi_n, \psi$, followed by the remainder of the given derivation (some of the numbers will have to be adapted). This will result in a derivation of $\neg\psi$ from $\phi_1, \ldots, \phi_n, \psi$ (in which no real use is made of ψ). This derivation can now be turned into a derivation of $\perp$ from $\phi_1, \ldots, \phi_n, \psi$ by adding E$\neg$ as a last step:

1.	ϕ_1	assumption
	.	
	.	
	.	
n.	ϕ_n	assumption
n + 1.	ψ	assumption
	.	
	.	
	.	
m + 1.	$\neg\psi$	
m + 2.	$\perp$	E$\neg$, n + 1, m + 1

□

(b) is an immediate consequence of (a).
(c) $\phi_1, \ldots, \phi_n, \neg\psi$ is consistent iff (according to (b)) $\phi_1, \ldots, \phi_n \nvdash \neg\neg\psi$ iff (according to (vi*) given in §4.3.7) $\phi_1, \ldots, \phi_n \nvdash \psi$. □

Before we present the completeness theorem in its alternative form, consider first its contraposition: if $\phi_1, \ldots, \phi_n \nvdash \psi$, then $\phi_1, \ldots, \phi_n \not\models \psi$. Now if the antecedent of this is replaced by means of theorem 16c, then we obtain: if $\phi_1, \ldots, \phi_n, \neg\psi$ is consistent, then $\phi_1, \ldots, \phi_n \not\models \psi$. Reformulating the consequent of this, we obtain: if $\phi_1, \ldots, \phi_n, \neg\psi$ is consistent, then there is a model **M** suitable for $\phi_1, \ldots, \phi_n, \psi$ and such that $V_M(\phi_1) = \ldots = V_M(\phi_n) = 1$ and $V_M(\psi) = 0$. Or in other words, if $\phi_1, \ldots, \phi_n, \neg\psi$ is consistent, then there is some suitable model **M** such that $V_M(\phi_1) = \ldots = V_M(\phi_n) = V_M(\neg\psi) = 1$. If in order to keep things short we just say that **M** is a *model* for the string of formulas $\chi_1, \ldots, \chi_m$ just in case $V_M(\chi_1) = \ldots = V_M(\chi_m) = 1$, then we see that the completeness theorem is equivalent to the following result.

Theorem 17 (*Consistency Theorem*)

If the string of sentences $\chi_1, \ldots, \chi_m$ is consistent, then there is a model for $\chi_1, \ldots, \chi_m$.

And the soundness theorem can be shown to be the reverse of theorem 17 in exactly the same manner:

Theorem 18

If the string of sentences $\chi_1, \ldots, \chi_m$ has a model, then $\chi_1, \ldots, \chi_m$ is consistent.

Nowadays it is usual to prove theorem 17 instead of proving the completeness theorem directly. One assumes that a set of sentences is consistent and then tries to provide it with a model. This idea was initiated by Henkin (1949). The original proof of the completeness theorem, the one given by Gödel (1930), was more direct.

All these theorems demonstrate a striking peculiarity of modern logic: its ability to theorize about its own systems and prove significant results about them. This 'self-reflecting' activity is sometimes called *metalogic*. In modern metalogic there are many more concerns than those touched upon so far. For instance, one can inquire into soundness and completeness for systems other than standard predicate logic, such as intuitionistic logic or higher-order logic (see chap. 5). But there are also other important metalogical theorems about predicate logic itself; we will survey a few, taking our first cue from an earlier theme.

We said before that the validity of an inference may be described as the absence of any counterexamples. And we also noted what a staggering task is involved in determining the latter state of affairs, since all interpretations in all models might be eligible in principle. But perhaps our apprehension in the face of 'the immense totality of all interpretations' seems a little exaggerated. After all, in propositional logic one can manage by checking the finite list of interpretations which are relevant to the validity of any given schema. In this, as in so many other respects, however, propositional logic is hardly representative of logical theories. Thus, all structures with arbitrary domains D have to be taken into account when evaluating schemata in predicate logic. And there is indeed an 'immense' number of these. The domain D may be finite or infinite, and within the latter type there are different varieties: among these some are countably infinite (like the natural numbers) and some are uncountably infinite (like the real numbers, or even bigger). In 1916, L. Löwenheim proved that predicate logic is at least insensitive to the latter difference between infinite sizes:

> If an inference has a counterexample with an infinite domain, then it has a counterexample with a *countably infinite* domain.

The true force of this result can probably only be appreciated against the background of a working knowledge of Cantor's set theory. But the following stronger formulation, which it received in the hands of D. Hilbert and P. Bernays in 1939, must still be quite surprising:

> If an inference has a counterexample, then it has a counterexample in *arithmetic* gotten by exchanging the predicate letters for suitable predicates of arithmetic, the formulas thus being seen as propositions about natural numbers.

This theorem led W. V. O. Quine (1970) to an interesting insight. In good nominalistic style, he compares the notions of validity that we have considered, $\vdash$ and $\models$, to the 'substitutional account' of validity: every substitution of suitable linguistic expressions in $\phi_1, \ldots, \phi_n, \psi$ which renders all of the premises true also renders the conclusion true.

It can easily be checked that syntactic derivability implies this form of validity. But conversely, nonderivability also implies (according to the Completeness Theorem) the existence of a counterexample, which in turn (according to the Hilbert-Bernays result) provides a counterexample in arithmetic which can serve as a nominalistic counterexample. Now nominalists do not believe in abstract structures like those involved in the definition of $\models$. The effect of Quine's idea is that the nominalists can nevertheless be reconciled to the notion: at least with regard to predicate logic, there is nothing wrong with $\models$. So metalogical theorems can sometimes be used to make philosophical points.

We have now seen how we can use finite and countably infinite structures to determine validity in predicate logic. If there are counterexamples to be found, then they are to be found among these structures. Can this be improved upon? Can we perhaps use just the *finite* structures? The answer is that we cannot. Every finite model of $\forall x \neg Rxx$ (the irreflexivity of R) and $\forall x \forall y \forall z((Rxy \wedge Ryz) \rightarrow Rxz)$ (the transitivity of R), for example, has an R-maximal element ($\exists x \forall y \neg Rxy$). But the derivation of the last of these formulas from the first two is nevertheless invalid. As a counterexample we have, for example, the natural numbers with R interpreted as *less than* (compare this with what is promised by Hilbert and Bernays's Completeness Theorem). Even worse, as was proved by B. Trahtenbrot in 1950, there can be no completeness theorem for the class of predicate-logical inferences which are valid on finite structures. These insights are also of at least some importance for the semantics of natural language. Given that the structures which natural language sentences are intended to pertain to are generally finite, the above shows that the infinite structures are not just a theoretical nicety: they are indispensable if we are to have a syntactically characterizable notion of validity.

In 1969, P. Lindström proved that the metaproperties which we have discussed are essentially characteristic of predicate logic. (We are now concerned only with languages with the same nonlogical vocabulary as predicate logic.)

> Any logical system plus semantics which includes predicate logic and such that a completeness theorem and Löwenheim's theorem hold, must coincide with predicate logic.

This is not put very precisely: finding an exact *formulation* for this metalogical theorem was actually a nontrivial part of Lindström's achievement.

But the idea amounts to the following. Extending predicate logic means losing at least one of the metaproperties of completeness, or the Löwenheim result. In particular, the stronger system of *second-order logic* is *incomplete*, as will appear in more detail in §5.4. There is no analogue of the completeness theorem for it, because its class of universally valid statements is too complex to admit of effective axiomatization. (Similar Lindström effects appear in connection with the *generalized quantifiers* which will be considered in vol. 2, like 'most' and 'for infinitely many'.) This is the phenomenon which, on the one hand, makes predicate logic so felicitous, and on the other, makes all of its extensions so mysterious and such a challenge to investigate.

Another aspect of inferences which has been studied quite extensively is their *decidability*. Is there, for some logical system, an effective method for deciding whether a given inference is valid or not? For propositional logic, there is. As we have seen, for instance, using the truth table test:

> Being a valid argument schema in propositional logic is a decidable notion.

and a fortiori

> Being a tautology of propositional logic is a decidable notion.

Moreover, by somewhat more complicated methods, one can also establish decidability for *monadic* predicate logic: that part of predicate logic which uses unary predicate symbols. Predicate logic taken as a whole, however, is *undecidable*. In 1936, A. Church proved his previously mentioned negative result (*Church's theorem*):

> Being a universally valid formula of predicate logic is not a decidable notion.

So the same must apply to predicate logic's set of valid argument schemata. The following *is true*, however, in the light of our earlier discussion:

> The set of argument schemata valid in predicate logic has an effective syntactic axiomatization.

For this assertion is always true of a system with a syntactic proof calculus which is complete with respect to its notion of inference. And predicate logic is such a system. For incomplete logical systems, however, like the previously mentioned second-order logic (or the theory of types to be presented in vol. 2), there isn't even an analogue of the last-mentioned result. The set of argument schemata which are valid in these systems has no effective syntactic characterization. This does not mean that one cannot use calculi of natural deduction in such cases: in fact, there exist interesting sound syntactic proof calculi for second-order logic too. But in view of the inescapable incompleteness of the system, they can never produce all of its universally valid formulas.

All of these metaresults give insights into the powers and limitations of the logical apparatus of deduction. But concrete reasoning always involves two distinct factors: there is inference and there are the initial *knowledge structures* from which inference must follow. The second formal aspect has also been studied extensively by logicians from a mathematical perspective, in a long tradition of research into the foundations of mathematics (and occasionally also other sciences). This involves investigations into the logical structure of axiomatized mathematical theories, the various metalogical properties which the theories can have, and the logical relationships they can have to each other in the web of scientific knowledge. Many different facets of our logical apparatus become relevant to the study of such issues as efficient representation and communication of knowledge. They range from the choice of an optimal vocabulary in which to formulate it to the choice of a suitable system of inference by which to develop and transmit it. For example, illuminating results have been achieved about the role of *definitions* in scientific theories (Beth's Theorem). Although foundational research tends to take place within an environment which is more concerned with scientific language than natural language, it is a source of inspiration for general logical and semantic studies too. (See Barwise 1977 for a comprehensive survey). ?

Exercise 17 ◇

Some logic textbooks are based on *maintaining consistency* rather than drawing inferences as the basic logical skill. So it is interesting to study the basic properties of consistency. Prove or refute the following assertions for sets of formulas X, Y and formulas ϕ:

(i) If X and Y are consistent, then so is their union $X \cup Y$.

(ii) If X is consistent, then so is $X \cup \{\phi\}$ or $X \cup \{\neg\phi\}$.

(iii) If X is inconsistent and ϕ is not universally valid, then there is a maximal consistent $Y \subseteq X$ which does not imply ϕ. Is this Y unique?

Exercise 18 ◇

Although full predicate logic is undecidable, many of its *fragments* are better behaved. As was observed in the text, for example, *monadic* predicate logic with only unary predicates is decidable. Another useful instance is the fragment consisting of *universal formulas*, i.e., formulas with arbitrary predicates but only universal quantifiers restricted to occurrences in front of quantifier-free formulas.

(i) Which of the earlier requirements on binary relations (see §3.3.8) are universal?

(ii) Prove that valid consequence among universal formulas is decidable, by showing that only certain finite models need be considered for its assessment.