PART I

Introduction to Deontic Logic

1

The Language of Logic and the Possibility of Deontic Logic

1.1. VALIDITY, TRUTH, AND LOGICAL FORM

An argument can be defined as a set of statements of a given language in which the truth of one of them (the conclusion) is assured by the truth of the others (the premises). Logic has to do with the study of arguments, but it is not a discipline oriented to describe our actual practices of argumentation and reasoning. It does not purport to examine the actual psychological processes or states of mind of people. In particular, it is not an empirical task whose outcome varies with different universes of analysis (i.e., different human groups and times). Instead, logic is concerned with a critical evaluation of argumentation; with patterns of *correct* reasoning.

However, this is still inaccurate as a proper characterization of logic, because any critical assessment of argumentation depends on the implicit goals ascribed to it, and there is not one but a plurality of goals we may seek to accomplish in our argumentative practices. For example, if we intend to persuade an audience, arguments will be evaluated as good or bad according to their merits to the extent that the audience is actually persuaded by our words. That would be an example of a *rhetorical* assessment of argumentation.

No doubt, the control of quality that logic exerts over argumentation is related to such rhetorical assessment, but this relation is not one of identity. Were an audience purely rational, it would only be persuaded by arguments whose premises warranted the conclusion, and this is what seems to be involved in the logical assessment of arguments. In other words, the aim that logic assigns to argumentation, and that is taken as a parameter to judge the quality and correction of arguments, is to preserve truth in the passage from the premises to the conclusion. From this point of view, it may be said that logic seeks to codify argumentative schemas that warrant the conclusion to be true if the premises are true. When an argument satisfies that condition, we say it is logically *valid*.

Logical validity, although connected with truth, does not truly depend on either the premises or the conclusion to be true. The connection between validity and truth is not so straightforward; to say that an argument is valid is to say that, *were* the premises true, the conclusion *could not* be false. Therefore, as the link between validity and truth is merely *conditional*, a valid argument may have true premises and a true conclusion, false premises and a false conclusion, or false premises and a true conclusion. The only combination validity excludes is the case in which an argument has true premises and a false conclusion. When an argument is not only valid but its premises (and, thus, its conclusion) are in fact true, we say that it is a *sound* argument.

Logic aims to isolate argumentative schemas in which the truth of the premises warrants the truth of its conclusion. That is why it focuses on the structure or *logical form* that links premises and conclusion. Consider the following argument:

(1) Some logicians are not boring; Lewis Carroll was a logician. Therefore, Lewis Carroll was not boring.

Although both premises and the conclusion are true, this is not a valid argument. This becomes obvious if we change the content of the statements, but preserving the same structure. Thus, replacing "logician" by "number," "not boring" by "prime" and "Lewis Carroll" by "8," we obtain:

(2) Some numbers are prime; 8 is a number. Therefore, 8 is prime.

Here, although the structure is exactly the same as in (1), the conclusion is obviously false. Thus, validity is not determined by the content of the premises but by the formal structure of statements. Let us see now examples of valid arguments:

(3) All Parisians are French; all French people are European. Therefore, all Parisians are European.
(4) Either Spain is the FIFA World Cup winner or Germany is; but Germany is not the FIFA World Cup winner. Therefore, Spain is the FIFA World Cup winner.

In these examples, the truth of the premises leaves no room for the conclusion to be false. The meaning of "to be a Parisian," "to be French" and "to be European" in (3), as well as the meaning of "Spain," "Germany" and "to be the FIFA World Cup winner" in (4), are irrelevant for their validity, in the sense that if we replaced those expressions by any other, but preserving the structure of the statements, we would equally obtain conclusive arguments.[1] In other words, any argument that exemplifies the schema:

(3') All *P*'s are *Q*; all *Q*'s are *R*. Therefore, all *P*'s *are R*.

or the schema:

(4') Either *p* or *q*; but not *q*. Therefore, *p*.[2]

will be valid. And they will be valid even if the premises are not true, because validity depends on the satisfaction of one condition: *The conclusion cannot be false if the premises are true*. Thus,

(5) All the fans of John Lennon marry Japanese women. All those who marry Japanese women are Japanese. Therefore, all the fans of John Lennon are Japanese.

and

(6) Either Al Pacino or Robert De Niro had the leading role in *The Shining*; but Robert De Niro did not have the leading role in *The Shining*. Therefore, Al Pacino had the leading role in *The Shining*.

are both valid arguments, even though their conclusions are false.

As the examples show, the meaning of certain expressions, such as "all," "some," "no," "and," "or," "not," "if . . . then," and so on, is of the utmost importance for linking the truth-values of the premises with the truth-value of the conclusion. By contrast, there are other aspects of the meaning of the premises and the conclusion of an argument that can be set aside to assess its validity. That is the reason why, since Aristotle, it has been said that the logical validity of an argument depends on its *form*. However, this claim requires a more thorough explanation, because it seems to suggest that each and every argument has a unique logical form, and this is not the case. For instance, the structure of (3) may be equally represented by (3') or, more simply, by:

(3") *p*, *q*. Therefore, *r*

where *p* represents "all Parisians are French," *q* represents "all French are European," and *r* represents "all Parisian are European." Now, although all

[1] For the different ways to define validity and logical truth, see Quine 1970: chapter 4.

[2] Later, we explain the reason for using capital letters in (3') and lowercase letters in (4') to stand for the variable contents.

arguments represented by (3') are valid arguments, that is not necessarily so with arguments that can be represented by (3").

Therefore, we might hold, on the one hand, that if an argument is an instance of more than one structure or form, and is valid according to one of them, then the argument is valid; and, on the other hand, if an argument is an instance of more than one structure or form, the richest of them should be preferred for its representation. However, although the first claim holds, the second is not justified; all arguments are capable of a diverse degree of complexity in their representation, but in many cases more complexity adds nothing relevant for the assessment of validity.[3]

In the formal languages developed by logicians, correctness in the patterns of inference may be evaluated either from a semantic or from a syntactic point of view. From a semantic point of view, a statement S (conclusion) in some language *L* is called a *semantic consequence* of a set α of statements of *L* (premises) ($\alpha \models$ S), if and only if S is true for every interpretation of *L* in which all the statements in α are also true. In such case, we say that the sequence $\alpha \models$ S is *semantically valid* or that it is a *logical truth*.

From a syntactic point of view, a statement S in language *L* is called a *syntactic consequence* of the set of statements α in *L* ($\alpha \vdash$ S), if and only if there is a finite sequence of statements $A_1 \ldots A_n$, where A_n = S, and each of the statements in the sequence is either an axiom of *L*, or an element of α, or it follows from previous statements in the sequence using a set of primitive rules of inference of *L*. The sequence $A_1 \ldots A_n$ is said to be *syntactically valid*, or that A_n is *demonstrable*.

The semantic approach, expressed in terms of interpretations and truth, has in a certain sense a *universal character*, because in order to prove that a statement *is not* a semantic consequence of a set of premises, it is sufficient to show the existence of an interpretation in which the premises turn out to be true and the conclusion false. By contrast, the syntactic approach, expressed in terms of axioms and primitive rules of inference, has in a certain sense an *existential character*, because to prove that a statement *is* a consequence of a set of premises, it is sufficient to show the existence of a finite sequence of statements allowing the derivation of the conclusion.

The ideal would be for these two notions of validity to correspond to one another – that is, that all logical truths (semantic consequences) were demonstrable (syntactic consequences), and all demonstrable formulas (syntactic consequences) were logical truths (semantic consequences). In the first case,

[3] See Haack 1978: 24.

the system would be *complete*; in the second case, the system would be *sound*. Unfortunately, both are contingent properties of formal systems, in the sense that they have to be proven for each formal system.

The intuitive and ordinary notion of validity that we presented at the beginning of this section corresponds to the semantic approach of formal validity, the one that has been dominant in contemporary logic. From a syntactic approach, logic seems at least at first sight to be the product of purely conventional and arbitrary choices. Being so, the creation of a logical system would be as free as the invention of a game. This seems to give conceptual priority to the semantic approach over the syntactic one. Under this view, what would guide the choice among different syntactical axiomatic calculi to identify a logical system would be that the axioms and theorems of the system are logical truths, and that the rules of inference guarantee that the truth of the premises is preserved in the conclusion.

Nevertheless, this remark should be refined in two respects. First, the intuitive notion of validity that is used to assess informal arguments of ordinary language does not have a perfect correspondence with the formal concepts of validity. Although logical systems have been developed for a rigorous representation of informal patterns of inference, precisely on account of this reason they cannot reproduce all their complexities, subtleties, inaccuracies, and vagueness. Therefore, the relation between the informal notion and the formal concepts of validity might be explained as follows: We take the intuitive and informal judgments of validity as a basis for the development of formal logical systems (i.e., rigorous theoretical systems that offer general principles for the assessment of validity). Now, in case of discrepancy between our intuitive judgments and the evaluation provided by those formal systems, we sometimes sacrifice our intuitive judgments, and sometimes sacrifice the general principles,[4] in a process analogous to the Rawlsian *reflective equilibrium*.[5] The selection of logical principles is not "fixed once and for all"; it requires adjustments between our normative standards and our intuitions.[6]

Second, the priority of the semantic over the syntactic approach of formal validity has not only been a controversial issue in the history of logic, but is also affected by a highly complex difficulty that we try to examine in the coming pages. Once again, it has to do with the connection between validity and truth. In spite of all the explanations and qualifications offered here, a strong

4 See Haack 1978: 25.
5 See Rawls 1971: 42–43.
6 See Engel 1989: 320.

connection still remains between validity and truth in the semantic approach; if validity of an argument is ultimately characterized as the preservation in the conclusion of the truth of the premises, the domain of logic would be restricted to the realm of truth. In other words, logical relations could not hold among entities that are incapable of truth-values, which seems to be a very deep and problematical limitation.

1.2. THE LANGUAGE OF LOGIC

Propositional calculus (PC) is the most simple and basic logical system.[7] The variables of PC are *propositional letters*. They represent sentences describing states of affairs that may or may not be the case. Each variable expresses a proposition (i.e., the meaning of sentences capable of independent truth-values). The constants of PC are *sentential connectives*, which may affect a whole formula (monadic connectives) or relate two or more formulas (dyadic connectives). Those formulas composed only by a single propositional variable will be called *atomic formulas*; the rest will be called *molecular formulas*.

Language of PC:

Propositional letters or variables: p, q, r, and so on. They range over propositions, and it is assumed that there is an unlimited number of them.

Propositional connectives:

- Monadic connective: $\sim$ (negation).
- Dyadic connectives: $\wedge$ (conjunction); $\vee$ (disjunction); $\rightarrow$ (conditional); $\leftrightarrow$ (biconditional).

Auxiliary signs: ( ) (brackets). They indicate a certain order within combinations of formulas. The rules for their use will not be stated, as they should be obvious from the context.

The rules to combine all these signs to produce admissible or *well-formed formulas* (*wffs* in the sequel) of PC are the following:

Formation rules of wffs for PC (recursive definition of wff of PC):[8]

- A propositional variable is a wff of PC;
- if α is a wff of PC, then $\sim\alpha$ is a wff of PC;

[7] For a more detailed presentation of the basic notions of propositional calculus and predicate logic, any introductory text dedicated to elementary logic may be consulted. Our suggestions: Gamut 1991 and Makinson 2008: 189 ff.

[8] Greek letters are used here to represent any arbitrary formula. They are not symbols of the language of PC itself, but (metalinguistic) signs to refer to the expressions of our language.

- if α and β are wffs of PC, then $(\alpha \wedge \beta)$, $(\alpha \vee \beta)$, $(\alpha \rightarrow \beta)$ y $(\alpha \leftrightarrow \beta)$ are wffs of PC;
- only those formulas constructed according to the previous clauses in a finite number of steps are wffs of PC.

A basic assumption of PC is that every proposition is either true or false, but not both. We will say that the truth-value of p is T if p is true, and F if it is false:

p
T
F

This representation is called a *truth table*; it shows the truth-values of a formula for all possible values of its constituent parts. The meaning of a molecular formula is built systematically from the meaning of its component parts. Thus, the truth-value of a molecular formula is determined by the truth-values of its atomic components. In this sense, the truth-values of compound formulas may be seen as a *function*, the domain being the set of all propositions of the language, and the range being the set {T,F}.

For any molecular formula α, with n being the number of different propositional letters in α, and there being two possible truth-values of each propositional letter, the number of cases to be analyzed in a truth table is 2^n. PC only deals with truth-functional connectives (i.e., those where the truth-values of the formulas over which they operate depend exclusively on the truth-values of their components). Therefore, logical connectives in PC only capture a certain aspect of their natural language counterparts,[9] and their meaning can be defined in terms of the truth-values of the formulas connected by them:

p	q	$\sim p$	$p \wedge q$	$p \vee q$	$p \rightarrow q$	$p \leftrightarrow q$
T	T	**F**	**T**	**T**	**T**	**T**
F	T	**T**	**F**	**T**	**T**	**F**
T	F	**F**	**F**	**T**	**F**	**F**
F	F	**T**	**F**	**F**	**T**	**T**

According to this, negation ($\sim$) switches the truth-value of the subsequent formula, so that the negation of a formula is true if the formula over which

9 See Suppes 1957: 5.

it operates is false, otherwise the negation is false, and thus it approximately corresponds to the meaning of the English word "no," and similar expressions. The conjunction ($\wedge$) of two formulas is true if both components are true, and false otherwise, and thus it approximately corresponds to the meaning of the English word "and" and similar expressions. The disjunction ($\vee$) of two formulas is true if at least one of them is true, and false otherwise, thus approximately corresponding to the meaning of the English word "or" and similar expressions. A conditional ($\rightarrow$) formula – often called "material conditional"[10] – is true if its antecedent is false or its consequent is true, otherwise it is false, and thus approximately corresponds to the meaning of the English clause "if . . . then," "only if," and similar expressions. Finally, a biconditional ($\leftrightarrow$) formula is true if its constituent formulas have the same truth-value, and false otherwise, approximately corresponding thus to the meaning of the English clause "if and only if," "just in case," and similar expressions.

Using the sign v to refer to the valuation of a formula and "iff" as the abbreviation of "*if and only if*," the semantics of PC might de expressed through the following clauses:

- $v(\sim\alpha) = \mathrm{T}$ iff $v(\alpha) = \mathrm{F}$
- $v(\alpha\wedge\beta) = \mathrm{T}$ iff $v(\alpha) = \mathrm{T}$ and $v(\beta) = \mathrm{T}$
- $v(\alpha\vee\beta) = \mathrm{T}$ iff $v(\alpha) = \mathrm{T}$ or $v(\beta) = \mathrm{T}$
- $v(\alpha\rightarrow\beta) = \mathrm{T}$ iff $v(\alpha) = \mathrm{F}$ or $v(\beta) = \mathrm{T}$
- $v(\alpha\leftrightarrow\beta) = \mathrm{T}$ iff $v(\alpha) = v(\beta)$

Now, take a formula such as $(p\vee\sim p)$. Its truth table is:

p	$\sim p$	$p\vee\sim p$
T	F	**T**
F	T	**T**

This formula is true regardless of the truth-values of its constituent atomic formulas. In other words, it is true for all possible truth-values of its variables. These formulas will be called *tautologies*. Tautologies are valid formulas of PC (i.e., they are true [value T] under any assignment of value to its variables). Take now a formula such as $(p\wedge\sim p)$. Its truth table is:

10 Material conditional must be distinguished from other kinds of conditional connectives, some of them stronger (like *strict* conditional) and other weaker (like *defeasible* conditional). We return to this subject when considering the formal representation of conditional norms and the alleged defeasible character of rules. For an introduction to material conditional, see Quine 1950: chapter 3.

p	$\sim p$	$p \wedge \sim p$
T	F	**F**
F	T	**F**

This is the opposite case: the formula is false regardless of the truth-values of its constituent atomic formulas. These formulas will be called *contradictions* (or *unsatisfiable*). Formulas that are neither contradictions nor tautologies will be called *contingencies*. Contingent formulas are those in which the corresponding truth tables contain at least one value T and one value F.

Truth tables can be used to analyze whether certain inference patterns are logically valid in PC. Take the following argument:

(1) If John rides his bicycle, he will arrive at his job on time; John rides his bicycle. Therefore, John will arrive at his job on time.

This argument may be represented as:

(1') $p \rightarrow q$, p. Therefore, q.

The first two formulas are the premises of the argument, and the third its conclusion. It may also be represented as a unique conditional formula in which the antecedent is formed by the conjunction of the two premises and the consequent by its conclusion:

(1") $((p \rightarrow q) \wedge p) \rightarrow q$

Under this second representation, validity can be tested through its truth-table:

p	q	$((p \rightarrow q) \wedge p) \rightarrow q$
T	T	T T **T**
F	T	T F **T**
T	F	F F **T**
F	F	T F **T**

As this formula is a tautology, there is no logical possibility for the premises to be true and the conclusion false, so the argument is valid.

Dyadic connectives in PC are interdefinable. If we take negation and any dyadic connective, it is possible to define all the rest.[11] Using the symbol = as

[11] A biconditional can obviously be reduced to a conjunction of two conditionals: $(p \leftrightarrow q) = ((p \rightarrow q) \wedge (q \rightarrow p))$.