Notre Dame Journal of Formal Logic
Volume 39, Number 2, Spring 1998

The Laws of Distribution for Syllogisms

WILFRID HODGES

Abstract The laws of distribution follow at once from Lyndon's interpolation theorem and the fact that the fallacy of many terms is a fallacy.

1 Introduction Since at least the seventeenth century, logic textbooks which discuss syllogisms have usually quoted some laws that a valid syllogism must obey. Different authors give different lists, but the following two laws are usually on the menu:

(a) The middle term must be distributed in at least one premise. (Some authors add: exactly one premise.)

(b) If a term is distributed in the conclusion, then it must be distributed in the premise in which it occurs. (Some authors add: if it is undistributed in the conclusion, then it must be undistributed in the premise in which it occurs.)

(In a moment I shall discuss the meanings of the technical terms in these laws.) The laws (a) and (b) are variously known as the *laws of distribution* or the *laws of quantity*. Kneale and Kneale ([2], p. 273) find the first parts of the two laws in the writings of the Jesuits of Coimbra in 1607. The second parts are apparently later; Peirce ([6], p. 350) had them in 1886.

One can justify the laws by checking that they hold for all valid syllogisms—there are only a small finite number to check. But many authors tried to give some general argument which covered all cases. These arguments were always unconvincing; but this was hardly surprising, since the early authors never managed to find suitable definitions of "distributed" and "undistributed." With twentieth century tools there is no problem in writing down sound definitions of these notions—in fact there are several ways of doing it—and then Lyndon's interpolation theorem [4] gives the laws of distribution almost immediately.

This paper is a revised version of the results of a discussion I had with Colwyn Williamson in March 1993. There is nothing original in it, beyond the easy observation that Lyndon's theorem gives the distribution laws. I wrote it up for publication because I became aware that the facts have not reached print. In fact they

Received August 27, 1998; revised March 11, 1999

haven't even reached the grapevine. Recently I came across a team whose research project revolved around answering the question which Lyndon's theorem has already answered.

2 Syllogisms We work with a signature σ consisting of three 1-ary relation symbols P, Q, R. So a σ-structure M consists of a set $\mathrm{dom}(M)$ (the *domain* of M) and three subsets P^M, Q^M, R^M of $\mathrm{dom}(M)$. I write $\mathcal{L}$ for the first-order language of signature σ. The language $\mathcal{S}$ of syllogisms has four types of atomic sentence. I write them below, each with its common English reading and a sentence of $\mathcal{L}$ which I shall call its *first approximant*. The symbols X, Y are metavariables ranging over the set of symbols $\{P, Q, R\}$; when we talk about $\mathcal{S}$, these three relation symbols are called *terms*.

(1) $A(X, Y)$, 'Every X is a Y.'
$\forall x\ (Xx \to Yx)$.

(2) $E(X, Y)$, 'No X is a Y'.
$\forall x\ (Xx \to \neg Yx)$.

(3) $I(X, Y)$, 'Some X is a Y'.
$\exists x\ (Xx \wedge Yx)$.

(4) $O(X, Y)$, 'Some X is not a Y'.
$\exists x\ (Xx \wedge \neg Yx)$.

The sentences of $\mathcal{S}$ are the atomic sentences of $\mathcal{S}$ and any built up from these by truth-functions $\neg, \wedge, \vee, \to, \longleftrightarrow$. There has been some confusion about how to interpret an atomic sentence of $\mathcal{S}$ in a σ-structure where one or both of the terms names an empty set. I shall assume for the moment that each atomic sentence means the same as its first approximant; but we shall have to come back to the question later.

We say that a σ-structure M is a *model of* the sentence φ if φ is true in M. The expression

$$\varphi_1, \ldots, \varphi_n \ \vdash\ \psi,$$

where $\varphi_1, \ldots, \varphi_n, \psi$ are sentences of $\mathcal{L}$ or $\mathcal{S}$, is called a *sequent*; its *premises* are $\varphi_1, \ldots, \varphi_n$ and its *conclusion* is ψ. It is said to be *valid* if every σ-structure which is a model of the premises is also a model of the conclusion. Instead of saying that the sequent is valid, we may say that $\varphi_1, \ldots, \varphi_n$ *entail* ψ. We say that sentences φ and ψ are *equivalent* if they are true in exactly the same σ-structures; or equivalently, if φ and ψ entail each other.

A *syllogism* is a sequent

$$\varphi_1, \varphi_2 \ \vdash\ \psi$$

where $\varphi_1, \varphi_2, \psi$ are atomic sentences of $\mathcal{S}$ and each of the three symbols P, Q, R occurs in precisely two of $\varphi_1, \varphi_2, \psi$. The term which occurs only in the premises is called the *middle* term of the syllogism.

For technical reasons we shall sometimes expand the signature σ by adding a new 1-ary relation symbol $Q^\star$. The new signature with this symbol added is called $\sigma^\star$. The notions of validity, equivalence, and so forth work for $\sigma^\star$-structures in the same way as for σ-structures.

3 The fallacy of many terms Suppose

$$\varphi_1, \varphi_2 \vdash \psi$$

is a valid syllogism. Suppose also that

$$\varphi_1^\circ, \varphi_2^\circ \vdash \psi^\circ$$

is a sequent in signature $\sigma^\star$ which comes from the syllogism by replacing the term Q in *one* of its occurrences by $Q^\star$. (There are two such sequents, depending on which occurrence of Q we replace.) We say that this sequent in $\sigma^\star$ commits *the fallacy of many terms*.

It will be important for us to know that the fallacy of many terms really is a fallacy. In other words:

$$\text{The sequent} \quad \varphi_1^\circ, \varphi_2^\circ \vdash \psi^\circ \quad \text{is not valid.}$$

To see this, suppose for definiteness that Q is the middle term and that we replace it by $Q^\star$ in φ_2. Let M be any $\sigma^\star$-structure in which ψ is false, and the sets P^M and R^M are not empty and not the whole domain of M. (Clearly there is such a structure.) Form a new $\sigma^\star$-structure N which agrees with M in its domain and its interpretations of P and R, but has interpretations of Q and $Q^\star$ which make the two premises φ_1° and φ_2° true. (Again it is clear that we can find such an N, since each of the symbols Q, $Q^\star$ occurs only once. For future reference we note that both Q^N and $Q^{\star N}$ can be chosen so that they are not empty and not the whole of the domain.) Since M and N agree in their interpretations of P and R, the conclusion ψ° (which is the same as ψ) is still false in N. Hence N is a counterexample to the sequent.

4 Distribution Suppose φ is a first-order sentence in which the symbols $\rightarrow$ and $\longleftrightarrow$ never occur. A relation symbol X is said to *occur positively* in φ if it has an occurrence which lies within the scope of an even number of negation signs; it is said to *occur negatively* in φ if it has an occurrence which lies within the scope of an odd number of negation signs.

The definitions of positive and negative occurrence in arbitrary first-order sentences are more complicated. For us it will be easiest to stick with the definitions above and use the fact that every first-order sentence is equivalent to one in which $\rightarrow$ and $\longleftrightarrow$ never appear.

Lyndon's interpolation theorem [4] states:

> If φ and ψ are first-order sentences and φ entails ψ, then there exists a sentence θ such that φ entails θ, θ entails ψ, and every relation symbol which occurs positively (negatively) in θ occurs positively (negatively) in both φ and ψ.

We shall need an immediate corollary of Lyndon's theorem:

Corollary 4.1 *Suppose φ and ψ are first-order sentences such that φ entails ψ, and a certain relation symbol Q occurs positively in at most one of φ and ψ, and also occurs negatively in at most one of φ and ψ. Suppose also that we introduce the new relation symbol $Q^\star$ and write $\psi^\star$ for ψ with every occurrence of Q replaced by $Q^\star$. Then φ entails $\psi^\star$.*

The point is that if θ is the Lyndon interpolant, then φ entails θ and θ entails ψ; but since Q doesn't occur in θ, θ must also entail $\psi^\star$.

Now we can define "distributed."

Definition 4.2 A term X is said to be *distributed* in an atomic sentence φ of $\mathcal{S}$ if X occurs in φ and there is a first-order sentence φ' which is equivalent to φ, in which X doesn't occur positively. Likewise X is said to be *undistributed* in φ if the same holds but with "negatively" for "positively."

For example, if we use first approximants to interpret $\mathcal{S}$, the sentence $A(P, Q)$ is equivalent to

$$\forall x\ (\neg Px \vee Qx).$$

We see that P is distributed and Q is undistributed in $A(P, Q)$. Likewise P and Q are both distributed in $E(P, Q)$ and both undistributed in $I(P, Q)$. Finally P is undistributed and Q distributed in $O(P, Q)$.

5 Proof of the laws of distribution I prove the laws of distribution first under the assumption that the first approximants give the meanings of atomic sentences of $\mathcal{S}$.

Consider the first half of law (a). For contradiction, let

$$\varphi_1, \varphi_2 \ \vdash \ \psi$$

be a valid syllogism in which the middle term Q is undistributed in both premises. Then we can translate the syllogism into a sequent of $\mathcal{L}$,

$$\varphi_1', \varphi_2' \ \vdash \ \psi',$$

in which Q doesn't occur negatively in either φ_1' or φ_2' and doesn't occur at all in ψ'. This sequent is equivalent to

$$\varphi_1' \ \vdash \ (\neg\varphi_2' \vee \psi'),$$

where Q doesn't occur negatively on the left of $\vdash$ and doesn't occur positively on the right. By Corollary 4.1, it follows that the sequent

$$\varphi_1' \ \vdash \ (\neg(\varphi_2^\star)' \vee \psi')$$

is valid, where $\varphi_2^\star$ is φ_2 with Q replaced by the new term $Q^\star$, and $(\varphi_2^\star)'$ is the corresponding first-order translation of $\varphi_2^\star$. (Recall that Q doesn't occur in ψ'.) Then translating back from $\mathcal{L}$ to $\mathcal{S}$,

$$\varphi_1, \varphi_2^\star \ \vdash \ \psi$$

is valid. But this sequent commits the fallacy of many terms and hence is invalid. Thus we reach a contradiction.

This proves the first part of law (a) of distribution. The proof of the second part is exactly the same but with "positively" and "negatively" transposed. For law (b) we argue in the same way but with the sequent

$$\varphi_1' \wedge \varphi_2' \ \vdash \ \psi',$$

noting that if a term occurs in the conclusion then it will occur in just one of φ_1' and φ_2'.

6 Other interpretations of $\mathcal{S}$ In dealing with syllogisms, many people restrict the class of σ-structures to those M in which P^M, Q^M, R^M are all nonempty. Some move of this kind is needed if we want Darapti (which I quote below) to be a valid form, as Aristotle took it to be. Another way of achieving exactly the same effect is to make no restriction on σ-structures, but to interpret an atomic sentence of $\mathcal{S}$, say with terms X, Y, as equivalent to the conjunction of its first approximant and the two sentences

$$\exists x Xx, \quad \exists x Yx.$$

These two sentences are called the *existence assumptions*. On this interpretation, if a syllogism is valid, then so is the corresponding first-order sequent where we add the existence assumptions only to the premises, not to the conclusion. (If the premises entail $\psi_1 \wedge \psi_2$, then they entail ψ_1 on its own.)

Now there are two ways of proceeding. The first route is to conjoin the existence assumptions when we test for distribution. This has the consequence that X is no longer distributed in $A(X, Y)$, since X occurs positively in $\exists x Xx$. I think nobody who works with syllogisms would be willing to go down this route. The second route, which I shall follow, is to use only the first approximant when testing for distribution; this way we still get the traditional division into distributed and undistributed.

The existence assumptions have no negative occurrences of any terms; so adding them will not make any difference to which terms occur negatively in the premises. On this interpretation, the first parts of laws (a) and (b) of distribution are proved just as before, recalling that when we proved that the fallacy of many terms is a fallacy, we always used nonempty sets.

Unfortunately the proofs of the second parts of the two laws go astray in this interpretation: if a symbol occurs positively on the right-hand side of the sequent, and the corresponding existence assumption appears on the left-hand side, then the symbol can occur in the interpolant. But this is as it should be, because under the present interpretation there are counterexamples to these second parts. A counterexample to the second part of (a) is Darapti, which has Q distributed in both premises:

$$A(Q, P), A(Q, R) \vdash I(R, P).$$

A counterexample to the second part of (b) is Bramantip:

$$A(P, Q), A(Q, R) \vdash I(R, P).$$

Here P is distributed in the first premise but undistributed in the conclusion.

On another interpretation sometimes attributed to the medievals, one should read the atomic sentences of $\mathcal{S}$ as follows.

$$\begin{array}{lll} A(X, Y) & : & \forall x\,(Xx \to Yx) \wedge \exists x Xx. \\ E(X, Y) & : & \forall x\,(Xx \to \neg Yx). \\ I(X, Y) & : & \exists x\,(Xx \wedge Yx). \\ O(X, Y) & : & \exists x Xx \to \exists x\,(Xx \wedge \neg Yx). \end{array}$$

(I assume as before that we are sticking with the traditional division into distributed and undistributed.) This interpretation disrupts the proofs of the second parts of laws

(a) and (b) for the same reason as before, with the same counterexamples. But the first parts of both laws are still intact by essentially the same proof as before. I leave the details to the reader. (For example when $O(X,Y)$ occurs as a premise of a valid sequent, the sequent remains valid when we drop the "$\exists x Xx \rightarrow$.")

7 Some historical remarks The Port-Royal Logic (Arnauld and Nicole, [1], III.3, for example) paraphrases "X is distributed in φ" as

> $[X]$ *doit être pris universellement*.

Several centuries earlier, Peter of Spain in his *Tractatus De Distributionibus* ([7], p. 209) defined distribution as

> *multiplicatio termini communis per signum universale facta*.

At first glance one might take these formulations as two ways of saying the same thing. But what the authors of the Port-Royal Logic meant was very different from what Peter of Spain meant.

Let us take Peter first. He explains that a common noun such as "man" is distributed if it appears in the context "every man" or "no man" or "whatever man" or a similar phrase. (See [7], p. 209 for the details.) I think it is not far from Peter's intentions if we say that a term X in a sentence φ is "distributed" if X appears just once in φ and φ has the form

$$\forall x\ (Xx\ \rightarrow\ \cdots).$$

Let us express this condition by saying that X is *universally quantified* in φ. Likewise we can say that X is *existentially quantified* in φ if X appears just once in φ and φ has the form

$$\exists x\ (Xx\ \wedge\ \cdots).$$

Peter has no name for this second condition.

On Peter's definition of distribution, the second term in an atomic sentence of $\mathcal{S}$ is never distributed; Peter never claims otherwise. So his definition is not the one needed for the traditional classification in syllogisms.

We turn to the Port-Royal authors. It seems that their formulation is meant to express that φ says about *all subsets* of X whatever φ says about X. For example, when the authors discuss the fact that Y is distributed in both $E(X,Y)$ and $O(X,Y)$, they comment that this fact means the same as the dictum that "If the genus is denied, the species also is denied" ([1], II.19).

Of course from the standpoint of today's logic there is no relevant difference between species of the genus and arbitrary subsets of the genus. In modern terms one would rewrite the Port-Royal definition as follows. Let us say that a term X in a formula φ is *downward monotone* in φ if for every structure M in which φ is true, φ is also true in N whenever N is the same as M except that X^N is a proper subset of X^M. Then a modern version of the Port-Royal Logic definition says that a term X is distributed in a sentence φ if and only if X is downward monotone in φ. (See Makinson [5] for a similar analysis.)

By analogy we say that X is *upward monotone* in φ if for every structure M in which φ is true, φ is also true in N whenever N is the same as M except that X^M is a proper subset of X^N.

If we read “distributed” as “downward monotone,” do we get the traditional classification of terms in syllogisms? Yes we do, for the following reasons. The Tarski truth definition quickly implies two facts:

If X doesn’t occur positively in φ then it is downward monotone in φ.

If X doesn’t occur negatively in φ then it is upward monotone in φ.

The usual way that logicians check upward or downward monotonicity is by means of these two facts. In any case we see at once that if a term is distributed (according to Definition 4.2), then it is downward monotone; and if it is undistributed then it is upward monotone.

For full first-order logic there is a converse which is more complicated. But for atomic sentences of $\mathcal{S}$ without repeated terms, one can check directly that terms which are downward (upward) monotone in a sentence φ of $\mathcal{S}$ don’t in fact occur positively (negatively) in the first approximant of φ. One can also check directly that each term in these sentences is either upward monotone or downward monotone, and not both; so the Port-Royal logic was entitled to take “undistributed” to mean “not distributed.”

The outcome of all these facts is that for the syllogistic sentences that we are considering, “distributed” and “undistributed” in the Port-Royal sense coincide exactly with “distributed” and “undistributed” in our sense. Hence the Port-Royal definition does lead correctly to the traditional classification of terms, unlike Peter of Spain’s definition.

It is interesting to read the proof of the first part of law (a) in the Port-Royal Logic ([1], III.3):

> Now if the middle term is taken twice [undistributed], it can be taken for two different parts of the same whole; and hence one cannot draw any conclusion (or at least, any necessary conclusion). This is enough to make an argument invalid, since one calls a syllogism valid only if the conclusion cannot be false when the premises are true.

The authors’ proof is precisely our proof, except for the fact that they skip over the reason why the fallacy of many terms arises. But this is exactly the step that Lyndon’s interpolation theorem gives us.

Note that in order to apply Lyndon’s theorem, we had to give separate and dual definitions of “distributed” and “undistributed.” It would not have been enough to define “undistributed” as “not distributed,” as the Port-Royal authors did. To the best of my knowledge, none of the traditional writers on distribution saw the need for a separate definition of “undistributed.”

The connection between distribution and monotonicity is certainly not new in the Port-Royal Logic. Sánchez Valencia [8] reports a number of medieval writers, including Lambert of Auxerre, Ockham and Burley, who saw some relationships between the two notions. One remark of Lambert ([3], p. 141) is particularly interesting:

> Again, it is important to know that not only universal signs have the power of distributing, but also negation.

Lambert goes on to note that adding a negation sign swaps a term from distributed to undistributed or vice versa.

When Lambert speaks of "universal signs" he has in mind the same condition that Peter of Spain used as a definition of "distributed," namely that the term comes with a word such as "every." Lambert's own definition is obscure, but most of what he says is consistent with the assumption that he means downward monotone. We should take a moment to compare "universally quantified" with "downward monotone."

In first-order notation a sentence

$$\ldots \text{ every man } \ldots .$$

goes over into

$$\forall x\, (\neg(x \text{ is a man }) \vee \ \ldots x \ldots).$$

Assuming that "man" doesn't appear anywhere else in the sentence, we see at once that "man" or "is a man" occurs only negatively, so that it is undistributed in our sense and hence downward monotone.

So universally quantified implies downward monotone. The converse is certainly not true, as Peter himself notes ([7], p. 224f, *"Utrum negatio habeat vim distribuendi"*). His conclusion is that negation is like universal quantification in that it is downward monotone (*"destructo genere destruitur quelibet eius species"*), but it doesn't distribute because it doesn't introduce a word such as "every." Peter's own examples are not syllogistic sentences in our sense. But he could have taken the term Y in $O(X, Y)$ as an example of a term which is downward monotone because of negation but not universally quantified.

When a term is universally quantified, one naturally asks whether the distribution comes from the $\forall x$ or the $\neg$ or both together or some other feature of the formula. Lyndon's theorem has told us the answer: it comes from the $\neg$, and the $\forall x$ is completely irrelevant. No doubt Lambert would have spotted this if he had known the first-order translation.

During the twentieth century the traditional logicians have suffered a bad press for their treatment of distribution. A fairer assessment would be that the best traditional logicians got remarkably close to the truth, given the inadequacy of their tools.

8 Rules of quality In passing let me mention another group of traditional rules. One counts the sentences $A(X, Y)$ and $I(X, Y)$ as *positive* and $E(X, Y)$ and $O(X, Y)$ as *negative*. (In fact the notation for atomic sentences of $\mathcal{S}$ comes from this division, using the Latin words "*AffIrmo*" and "*nEgO*.")

(c) If both premises of a valid syllogism are positive then the conclusion is positive.
(d) If one premise of a valid syllogism is negative then the conclusion is negative.
(e) There is no valid syllogism with both premises negative.

Law (c) is not hard to prove. Let us say that a σ-structure M is *terminal* if it has exactly one element and this element is in all of P^M, Q^M, R^M. (Terminal structures are the terminal objects in the category of σ-structures and homomorphisms.) One checks that (on any of the interpretations discussed above) if φ is an atomic sentence of $\mathcal{S}$ and M is terminal, then φ is true in M if and only if φ is positive.

For those who know the terminology, another way of stating this proof of (c) is that the positive sentences are strict Horn and the negative sentences are nonstrict Horn. There is no valid implication from strict Horn premises to a nonstrict Horn conclusion.

Law (d) follows by a switch. Suppose

$$\varphi_1, \varphi_2 \vdash \psi$$

is a valid syllogism which violates (d). Then

$$\varphi_1, \neg\psi \vdash \neg\varphi_2$$

is a valid sequent which violates (c). But if φ is an atomic sentence of $\mathcal{S}$ then $\neg\varphi$ is equivalent to an atomic sentence of $\mathcal{S}$ (under the existence assumptions, if we are using these).

I know no interesting general principle which gives (e). This law is very sensitive to Aristotle's choice of atomic sentences. Had he introduced a sentence $Y(X, Y)$ with the meaning "There is something which is neither an X nor a Y," the sequent

$$E(Q, P), E(Q, R) \vdash Y(P, R)$$

would have been a counterexample to (e) in the interpretation where all terms are required to be nonempty. As the referee kindly notes, (e) also fails if we allow negative terms such as "not-man."

Acknowledgments My thanks to Colwyn Williamson; and also to Johan van Benthem, Jan van Eijck, David Makinson, Víctor Sánchez Valencia and the anonymous referee for their comments.

REFERENCES

[1] Arnauld, A., and P. Nicole, *La Logique, ou l'Art de Penser*, Paris, 1662. 7, 7, 7

[2] Kneale, W., and M. Kneale, *The Development of Logic*, Clarendon Press, Oxford, 1962. Zbl 0100.00807 MR 26:2353 1

[3] Kretzmann, N., and E. Stump, editors, *The Cambridge Translations of Medieval Philosophical Texts*, vol. 1, Cambridge University Press, Cambridge, 1988. 7

[4] Lyndon, R., "An interpolation theorem in the predicate calculus," *Pacific Journal of Mathematics*, vol. 9 (1959), pp. 129–42. Zbl 0093.01002 MR 21:5555 1, 4

[5] Makinson, D., "Remarks on the concept of distribution in traditional logic," *Noûs*, vol. 3 (1969), pp. 103–08. 7

[6] Peirce, C. S., "Qualitative logic," pp. 323–71 in *Writings of Charles S. Peirce*, vol. 5, edited by C. J. W. Kloesel, Indiana University Press, Bloomington, 1993 (original 1886). 1

[7] Peter of Spain, *Tractatus*, called afterwards *Summule Logicales*, first critical edition from the manuscripts, with an introduction by L. M. de Rijk, Van Gorcum and Co., Assen, 1972. 7, 7, 7

[8] Sánchez Valencia, V., “Monotonicity in medieval logic,” *Language and Cognition*, vol. 4, Yearbook 1994 of the Research Group for Theoretical and Experimental Linguistics of the University of Groningen, Faculteit der Letteren, Rijksuniversiteit Groningen, pp. 161–74. 7

Wilfrid Hodges
School of Mathematical Sciences
Queen Mary and Westfield College
Mile End Road
London E1 4NS
UNITED KINGDOM
email: w.hodges@qmw.ac.uk