

Aristotle, Boole, and Categories

Vaughan Pratt

October 12, 2015

Abstract

We propose new axiomatizations of the 24 assertoric syllogisms of Aristotle's syllogistic, and the 2^{2^n} n -ary operations of Boole's algebraic logic.

The former organizes the syllogisms as a 6×4 table partitioned into four connected components according to which term if any must be inhabited. We give two natural-deduction style axiomatizations, one with four axioms and four rules, the second with one axiom and six rules. The table provides immediately visualizable proofs of soundness and completeness. We give an elementary category-theoretic semantics for the axioms along with criteria for determining the term if any required to be nonempty in each syllogism.

We base the latter on Lawvere's notion of an algebraic theory as a category with finite products having as models product-preserving set-valued functors. The benefit of this axiomatization is that it avoids the dilemma of whether a Boolean algebra is a numerical ring as envisaged by Boole, a logical lattice as envisaged by Peirce, Jevons, and Schroeder, an intuitionistic Heyting algebra on a middle-excluding diet as envisaged by Heyting, or any of several other candidates for the "true nature" of Boolean logic. Unlike general rings, Boolean rings have only finitely many n -ary operations, permitting a uniform locally finite axiomatization of their theory in terms of a certain associative multiplication of finite 0-1 matrices.

1 Introduction

Whereas the ancient Romans excelled at civil engineering and economics, the ancient Greeks shone in geometry and logic. Book I of Euclid's *Elements* and Aristotle's assertoric syllogisms dominated the elementary pedagogy of respectively geometry and logic from the 3rd century BC to the 19th century AD.

In the middle of the 17th century Euler abstracted Euclid's geometry to affine geometry by omitting the notions of orthogonality and rotation-invariant length, and two centuries later Boole generalized validity of Aristotle's unconditionally valid syllogisms to zeroth-order propositional logic by inventing Boolean rings. Yet subsequent literature has continued to find the original subjects of great interest. The 2013 4th World Congress and School on Universal Logic for example featured a score of presentations involving syllogisms.

Of these, logic is closer to the interests of my multidecadal colleague, coauthor, and friend Rohit Parikh. I shall therefore focus here on logic generally, and more specifically on the contributions of Aristotle and Boole, with categories as a common if somewhat slender thread.

Starting in the 1970s, Rohit and I both worked on dynamic logic [11, 8], a formalism for reasoning about behaviour that expands the language of modal logic with that of regular expressions. Dynamic logic witnessed the introduction into program verification of the possible-world semantics of modal logic [11], and into logic of multimodal logic with unboundedly many modalities, so it would be only logical to write about Aristotle's modal syllogistic.

However I was a number theorist before I was a logician [10], and it has puzzled me as to why the number of Aristotle's two-premise assertoric syllogisms should factor neatly as 6×4 , the inclusion of a few odd-ball syllogisms whose middle term did not interpolate its conclusion notwithstanding. A concrete and even useful answer to this question will hopefully prove of greater interest, at least to the arithmetically inclined, than more analytic observations on Aristotle's modal syllogistic.

2 Aristotle's logic

As a point of departure for our axiomatization, the following three subsections rehearse some of the basic lore of syllogisms, which has accumulated in fits and starts over the 23 centuries since Aristotle got them off to an excellent albeit controversial start. As our interest in these three subsections is more technical than historical we emphasize the lore over its lorists.

2.1 Syntax

An assertoric syllogism is a form of argument typified by, *no cats are birds, all wrens are birds, therefore no wrens are cats*.

In the modern language of natural deduction a *syllogism* is a sequent $J, N \vdash C$ consisting of three *sentences*: a maJor premise J , a miNor premise N , and a Conclusion C .

Sentences are *categorical*, meaning that they express a relation between two terms each of which can be construed as a category, predicate, class, set, or property, all meaning the same thing for our purposes.

The language has four relations:

- (i) a *universal* relation XaY (X all-are Y) asserting all X are Y, or that (property) Y holds of every (member of) X;
- (ii) a *particular* relation XiY (X intersects Y) asserting some X are Y, or that Y holds of some X;
along with their respective *contradictories*,
- (iii) XoY (X obtrudes-from Y) asserting some X are not Y, or Y does not hold of some X, or $\text{not}(XaY)$; and

- (iv) XeY (X empty-intersection-with Y) asserting no X are Y, or Y holds of no X, or not(XiY).

Set-theoretically these are the binary relations of inclusion and nonempty intersection, which are considered *positive*, and their respective contradictories, considered *negative*. Contradiction as an operation on syllogisms interchanges universal and particular and changes sign (the relations organized as a string *aeio* reverse to become the string *oiea*) but is not itself part of the language. Nor is any other Boolean operation, the complete absence of which is a feature of syllogistic, not a bug.

It follows from all this that a syllogism contains six occurrences of terms, two in each of the three sentences. A further requirement is that there be three terms each having two occurrences in distinct sentences.

The following naming convention uniquely identifies the syllogistic form. The conclusion is of the form S-P where S and P are *terms* denoting subject and predicate respectively. S and P also appear in separate premises, which are arranged so that P appears in the first or *major* premise and hence S in the second or *minor* premise.

The third term is denoted M, which appears in both premises, either on the left or right in each premise, so four possibilities, which are numbered 1 to 4 corresponding to the respective positions LR, RR, LL, and RL for M in the respective premises. That is, Figure 1 is when M appears on the left in the major premise and the right in the minor, and so on. The header of Table 1 in Section 2.5 illustrates this in more detail (note the Gray-code order).

So in the example at the start of the section, from the conclusion we infer that S = *wrens* and P = *cats*, so M has to be *birds*. That M appears on the right in both premises (RR) reveals this syllogism to be in Figure 2. Since P (*cats*) appears in the first or major premise there is no need to switch these premises.

When the occasion arises to transform a syllogism into an equally valid one, the result may violate this naming convention. For example we may operate on the conclusion and the major premise by contradicting and exchanging them, amounting to the rule of *modus tollens* in the context of the other premise. After any such transformation we shall automatically identify the new subject, predicate, and middle according to the naming convention, which in general could be any of the six possible permutations of the original identifications. And if this results in the otherwise untouched minor premise now containing P it is promoted to major premise, i.e. the premises are switched.

Syllogisms are divided syntactically into $4^3 = 64$ *moods* according to which of the 4 relations are chosen for each of their 3 sentences. The wren example has the form PeM, SaM ⊢ SeP and so its relations are *e*, *a*, and *e* in that order. This mood is therefore notated EAE, to which we append the figure making it the *form* EAE-2.

2.2 Semantics

Syntactically, figures and moods are independent, whence there are $4 \times 64 = 256$ forms. Semantically however not all of them constitute valid arguments. If for example we turn the above EAE-2 wren example into an EAE-4 syllogism by replacing its minor premise by *all birds are two-legged* and its conclusion by *no two-leggeds are cats*, we obtain a syllogism with premises that, while true, do not suffice to rule out the possibility of a two-legged cat. Hence EAE-4 must be judged invalid.

Taking S to be *warm-blooded* instead of *two-legged* might have made it easier to see that EAE-4 was invalid, since the conclusion *no warm-bloodeds are cats* is clearly absurd: all cats are of course warm-blooded. The two-legged example draws attention to the sufficiency of a single individual in a counterexample to EAE-4.

The semantics of syllogisms reduces conveniently to that of first order logic via the following translations of each sentence $t = P\text{-}Q$ to a sentence \hat{t} of the (monadic) predicate calculus.

$$\text{PaQ: } \forall x[\neg P(x) \vee Q(x)]$$

$$\text{PeQ: } \forall x[\neg P(x) \vee \neg Q(x)]$$

$$\text{PiQ: } \exists x[P(x) \wedge Q(x)]$$

$$\text{PoQ: } \exists x[P(x) \wedge \neg Q(x)]$$

Call a sentence of first order logic *sylogistic* when it is either of the form $\forall x[L_1(x) \vee L_2(x)]$ where the L_i 's are literals with distinct predicate symbols (thereby precluding $\forall x[P(x) \vee \neg P(x)]$), or the negation of such a sentence, i.e. of the form $\exists x[L_1(x) \wedge L_2(x)]$. Call a set of sentences, of any cardinality, *sylogistic* when every member is sylogistic and every pair of members shares at most one predicate symbol.

Theorem 1. *Any sylogistic set of sentences having at most one universal sentence is consistent.*

Proof. Let S be such a set having u as its only universal sentence if any. Form a model of S by taking its universe E to consist of the existential sentences. For each member e of E , set to *true* at e the literals of e and, if u exists, the literal of u whose predicate symbol does not appear in e . Set the remaining values of literals to *true*. This model satisfies every sentence of S , which is therefore consistent. \square

With one exception, this construction does not extend to sylogistic sets with two or more universal sentences because those sentences may contain a complementary pair of literals. The exception is when S contains no existential sentence, in which case the construction produces the empty model and all the universal sentences are vacuously satisfied regardless of their contents.¹

The following interprets a syllogism as a 3-element sylogistic set whose third element is the negation of the translation into predicate calculus of its

¹We take the traditional proscription in logic of the empty universe to be a pointless superstition that creates more problems than it solves.

conclusion. The validity of the syllogism is equivalent to the unsatisfiability of that set.

Corollary 2. *A valid syllogism must contain exactly one particular among its premises and contradicted conclusion.*

For if it contains two particulars it contains only one universal, whence Theorem 1 produces a counterexample, while if it contains no particulars then the exception to the non-extendability of the theorem to multiple universals produces a counterexample.

Hence of the $4^3 = 64$ moods, only $2^3 \times 3 = 24$ can be the mood of a valid syllogism. In conjunction with any of the four figures, there are therefore $24 \times 4 = 96$ candidate forms, call these *presyllogisms*. It follows that a valid syllogism must have at least one universal premise. Furthermore if it has a particular premise then the conclusion must also be particular, and conversely if it doesn't then the conclusion must be universal.

Theorem 3. *A presyllogism $J, N \vdash C$ is valid if and only if its translation $\hat{J} \wedge \hat{N} \wedge \neg \hat{C}$ into propositional calculus is unsatisfiable.*

Proof. By the corollary any counterexample requires only one individual witnessing the one particular and any other individuals may be discarded. But in that case both $\forall x$ and $\exists x$ act as the identity operation. In the translation to predicate calculus the quantifiers may be dropped and every literal $P(x)$ simplified to a propositional literal P . \square

So to decide validity of a form, verify that it is a presyllogism, translate it as above, and test its satisfiability.

The $24 \times 4 = 96$ possible forms of a valid syllogism are then those whose translation into propositional calculus is a tautology. These turn out to be the 15 in the largest of the four regions of Table 1 in Section 2.5 (the top four rows less AAI-4). A C program using this method to enumerate them can be seen at <http://boole.stanford.edu/syllenum.c>, requiring 1.33 nanoseconds on my laptop to translate and test each presyllogism (today's i7's are fast).

2.3 The problem of existential import

It is generally agreed today that there are 24 assertoric syllogisms. What of the 9 that the procedure judged invalid?

One of them, AAI-1, is illustrated by the syllogism, *all unicorns are ungulates* (hooved), *all ungulates are mammals, therefore some unicorns are mammals*. But if unicorns don't exist this is impossible.

The truth of assertions about empty classes could go either way, or even both: Schrödinger's cats could be both dead and alive if he had none.² To avoid inferring the existence of unicorns, and also to reflect intuitions about

²This makes more sense when phrased more precisely as a property of each *member* of the empty class; with no members no opportunity for an inconsistency ever arises.

natural language (at least Greek), Aristotle found it convenient to deem positive assertions about empty classes to be false. This convention justifies Aristotle's principle of **subalternation**: SiP is subalternate to, or implied by, SaP, one edge of the traditional Square of Opposition [9].

But Aristotle's convention obliges the truth of not-(*all unicorns are unguulates*). Aristotle himself dealt with this by not identifying it with *some unicorns are not unguulates* but many medieval logicians found it natural to make that identification, some seemingly without noticing the inconsistency.

The long history is recounted by Parsons.[9] To cut to the chase, the modern view is that the nine extra syllogisms are **conditionally** valid, conditioned on the nonemptiness of one of its terms. For example the foregoing AAI-1 example is valid provided unicorns exist. This approach weakens subalternation in the Square of Opposition [9] to something more subtle than envisaged by Aristotle.

One such subtlety is exposed by the following principle.

Say that Y **interpolates** X and Z when (i) YaZ and either XaY or XiY , or (ii) the same with X and Z interchanged. Interpolation is of interest because by transitivity we may infer in case (i) either XaZ or XiZ according to the relation of X to Y , and in case (ii) the same with X and Z interchanged.

Theorem 4. (Interpolation principle) *In any unconditionally valid syllogism, M' interpolates S and P' where M' may be either M or its complement \bar{M} , and likewise for P' , independently.*

For example when the mood is EIO in any figure, the premises express $SiMa\bar{P}$ whence $Si\bar{P}$, that is, SoP. For OAO-3 we have $\bar{P}iMaS$ whence $\bar{P}iS$ or $Si\bar{P}$ or SoP, and so on.

This reveals interpolation as the engine of unconditional validity.

We defer the proof to the end of Section 2.5, which will be much easier to see after absorbing Table 1.

The subtlety is that it is not the interpolation principle that drives the following conditionally valid syllogism.

EAO-4: *No donkeys are unicorns, all unicorns are mammals, therefore some mammals are not donkeys.*

Certainly the middle term seems ideally located to be the interpolant. But the conclusion SoP, that is, $Si\bar{P}$, follows from $Ma\bar{P}$ and MaS , not by transitivity of anything but from S and \bar{P} having M in common under the assumption that M is nonempty. EAO-4 shares this subtlety with two more conditionally valid syllogisms, EAO-3, which works the same way, and AAI-3 which replaces \bar{P} by P .

So how would Aristotle have justified these three? Simple: by subalternation, which permits deducing AAI-3 from the unconditionally valid AII-3, namely by "strengthening" the minor premise MiS to MaS , and similarly for the other two.

If we are to drop subalternation yet keep this little flock of three black sheep, some alternative means of support must be found for them. In the next section we offer just two: taking AAI-3 as an axiom, or permitting *modus tollens*, our rule of last resort, to be applied even to conditional syllogisms.

2.4 Yet another axiomatization of syllogistic deduction

The first major effort to bring Aristotle's syllogistic up to the standards of rigor of 20th century logic was carried out by Lukasiewicz in his book *Aristotle's Syllogistic* [7]. Lukasiewicz cast syllogistic deduction in the framework of predicate calculus axiomatized by a Hilbert system, rendering entire syllogisms such as EIO-1 as (in notation we might use today) single formulas $e(M, P) \wedge i(S, M) \supset o(S, P)$.

But while technically correct, the variables S, M, P range over classes and the relations a, e, i, o as binary predicates hold between classes, making this a second-order logic. Moreover it introduces Boolean connectives into a subject that had previously only seen them in translations of individual sentences such as in the third paragraph of Section 2.2 but not in systems of syllogistic deduction such as Aristotle's. The latter are much closer to natural deduction, as the case of Gentzen's sequent calculus with one consequent, than to Hilbert systems.

Oddly enough Lukasiewicz himself had previously been an early and effective contributor to natural deduction, for which Aristotle's syllogistic would have been an ideal application, so it is strange that he chose a different framework.

Sequent calculi typically reverse the ratio of axioms to rules, having many rules but as few as one axiom, frequently $P \vdash P$. More significantly, they lend themselves to systems like Aristotle's by not requiring Boolean connectives. That such connectives are frequently encountered in sequent calculus systems, for example in rules such as \wedge -introduction and \forall -elimination, is not an intrinsic property of sequent calculi but only of the particular language being axiomatized by the particular deduction rules. The four connectives a, e, i, o of Aristotle can be axiomatized just as readily in the sequent calculus as can the three connectives \wedge, \vee, \neg of Boole, in each case without assistance from other connectives.

Aristotle's first system had only two axioms, AAA-1 and EAE-1, mnemonically named *Barbara* and *Celarent*, which he viewed as self-evident and therefore not in need of proof. He derived his remaining syllogisms via a number of rules based on the Square of Opposition [9], including the problematic notion of subalternation. This discrepancy with Lukasiewicz's second-order Hilbert-style account was pointed out by Corcoran in a series of papers in the early 1970's, culminating in a 1974 paper [3] proposing formal systems D and D2 based on Aristotle's systems, along with variants D3 and DE.

Our axiomatization, which we call D4p as a natural successor to Corcoran's D3, with "p" for preliminary as will be seen shortly, is in the same natural deduction format as those of Aristotle and Corcoran. The essential differences are in its choice and formulation of axioms and rules, its preservation only of validity and not meaning (Rules 2 and 3), and its emphasis on visualizing validity and completeness graphically to make their proofs easily seen simply by staring at Table 1.

As axioms we share *Barbara* with Aristotle, but in place of *Celarent* we take instead the three conditionally valid syllogisms of mood AAI, taking care to keep them logically independent. As a further departure from Aristotle, instead

of taking the axioms as self-evident we shall justify them semantically in terms of certain posets lending themselves to a very elementary category-theoretic treatment.

As rules we have as usual the *converse* of the intersection sentences, provided for by rules R1 and R4, namely XeY as YeX and XiY as YiX , which do not change their meaning. But Rules R2 and R3 also allow the *obverse* of any sentence, which changes only the sign of the relation (XaY to XeY , XoY to XiY , etc.) leaving unchanged the terms and whether universal or particular. This changes the meaning by complementing a term, either P or M respectively, and is therefore only applicable to two sentences with the same right hand side, which exist in every figure save Figure 4.

Recall that a presyllogism is a syllogistic form whose premises and contradicted conclusion contain exactly one particular. No conditionally valid syllogism is a presyllogism.

System D4p

Axioms

- A1. AAA-1 (Barbara)
- A2. AAI-1 (Barbari)
- A3. AAI-3 (Darapti)
- A4. AAI-4 (Bamalip)

Rules

- R1. (*cj* or *cn*) Convert any *e* or *i* premise. (E.g. $SeM \rightarrow MeS$.)
- R2. (*ojc*) In Figures 1 and 3, obvert the major premise and the conclusion.
- R3. (*ojn*) In Figure 2, obvert both premises.
- R4. (*cc*) In Figure 3, convert any *e* or *i* conclusion and interchange the premises.
- R5. (*mtj* or *mtn*) For any presyllogism, contradict and exchange the conclusion and a premise.

Although every rule modifies the syllogism it applies to, rules R1-R3 do not modify the identities of the subject, predicate, and middle term, though R2 and R3 change the sign of both instances of a term on the right. Rule R4 converts the conclusion as P-S, making P the new subject and vice versa; the premises remain untouched other than switching the identities of S and P, requiring exchanging which is major and minor. Rule R5 can entail any of the five non-identity permutations of S, P, and M.

2.5 6x4 tabulation of the assertoric syllogisms

Table 1 lays out the 24 valid and conditionally valid syllogisms in a form designed to facilitate visual verification of soundness and completeness, and also to make it easy to see what adjustments to the axioms are needed in response to certain adjustments of the rules.

1	2	4	3
M-P S-M	P-M S-M	P-M M-S	M-P M-S
-cn- EIO	-cj- EIO	-cn- EIO	-cj- EIO
<i>ojc</i>	<i>ojn</i>		<i>ojc</i>
-cn- AII	AOO	AAI	AII -cn-
			<i>cc</i>
EAE -cj-	EAE	IAI -cj-	IAI
<i>ojc</i>	<i>ojn</i>	<i>mtj</i>	<i>ojc</i>
AAA	AEE -cn-	AEE	AOA
EAO -cj-	EAO	EAO -cj-	EAO
<i>ojc</i>	<i>ojn</i>		<i>ojc</i>
AAI	AEO -cn-	AEO	AAI

Table 1 Interdeducibility from the four axioms (shown in boldface)

The four columns are indexed by their respective figures, numbered in Gray-code-plus-one as 1,2,4,3 and wrapping around to form a cylinder. The connectors between adjacent syllogisms denote applications of single rules. Where no single rule applies a line is drawn to indicate blocking and the need to go round via a chain of deductions, if even possible. Two instances of Rule R1 labeled *-cn-*, converting the minor premise in each case, wrap around to connect figures 1 and 3.

Rule R1 appears as *-cj-* six times and *-cn-* five times (the two at far left are the same occurrences as at far right) denoting the converse of respectively major and minor premises, changing figure but not mood. Rule R2 appears six times as *ojc* denoting complementation of the major premise and conclusion, changing mood but not figure. Rule R3 appears three times as *ojn* denoting complementation of both premises, also changing only mood.

It is straightforward and not even terribly tedious to verify that every possible application of Rules R1-R3 appears explicitly in the table. Only two other connections are made, one instance of Rule R4, namely *rc* in Figure 3, and one instance of Rule R5, namely *mtj* in Figure 4 (the *j* names the particular premise in the syllogism IAI-4 as the one to contradict), both changing mood but not figure.

These 22 horizontal and vertical connections create a graph with four connected components each containing one axiom. If for a moment we replace *cc*

and mtj by a blocking line, corresponding to allowing only Rules R1-R3, we now have six connected components. If we add AII-1 and IAI-4 as Axioms A5 and A6 respectively, each component now contains exactly one axiom.

In the following theorems we take the criterion for completeness to be deducibility of all 24 syllogisms from the listed axioms using the listed rules.

The foregoing observations give a quick proof of the following.

Theorem 5. *Axioms A1-A6 and Rules R1-R3 completely axiomatize Aristotle's syllogisms.*

The only other instance of Rule 4, not shown, is accounted for by noting that it carries AAI-3 to itself and therefore creates no new connections. This gives another easy theorem trading off axioms for rules.

Theorem 6. *Axioms A1-A5 and Rules R1-R4 completely axiomatize Aristotle's syllogisms.*

Rule R5 is much more broadly applicable. The instance mtj completes the connection of the 15 presyllogisms, those forms having one particular after contradicting the conclusion, namely all but AAI-4 in the top four rows. However it is straightforward to show (and intuitively obvious) that R5 carries presyllogisms to presyllogisms, whence all further instances of its use can have no further impact on the connectivity of Table 1.

In this way we have given a largely visual proof of the main theorem promised in the abstract:

Theorem 7. *Axioms A1-A4 and Rules R1-R5 completely axiomatize Aristotle's syllogisms.*

Rule R4's limitation to Figure 3 allowed for a visually simple proof of Theorem 7. It is again easy to see that dropping that restriction adds three more connections, between AII-1 and IAI-4, EAE-1 and AEE-4, and AAI-4 and AAI-1. Since the first two are within the unconditional component they change nothing. However connecting AAI-4 and AAI-1 brings the former out of solitary confinement to permit a further reduction in axioms merely by broadening the applicability of a rule, summarized thusly.

Theorem 8. *Axioms A1-A3 and Rules R1-R5 with R4 unrestricted completely axiomatize Aristotle's syllogisms.*

Although Aristotle eventually adopted Figure 4, his initial identification of it with Figure 1 can be understood as viewing all three of AAI-4, IAI-4, and AEE-4 as being at least morally part of Figure 1.

Similarly lifting the restriction on Rule 5 will enable the further elimination of Axiom A3. However before tackling this it is worthwhile investigating the meaning of the axioms in order to have a clearer picture of the implications of deducing one universal conditional syllogism from another.

But first this interruption: we now have enough material for the promised

Proof. (of Theorem 4) The principle is easily verified by hand for AAA-1 and AII-1 as its two canonical instances for each of XaY and XiY . All remaining unconditionally valid syllogisms are reachable from these two via Rules R1-R4. Rule R1 (cj and cn) involves no relabeling of S, M, or P while Rule R4 (rc) only interchanges S and P, whence the need for part (ii) of the definition of interpolation. Rule R2 (ojc) complements P while Rule R3 complements M. Since no rule complements S there is no need for the more general S'. \square

2.6 Axiom semantics

We propose to interpret the three terms of each of axioms A1-A4 as three vertices S, M, P of a graph; each sentence of the form XaY as a path from X to Y , and the conclusion XiY of the conditional axioms as a pair of paths to respectively X and Y , both starting from a fourth vertex 1 and sharing a common first edge to one of S, M, or P.

Table 2 exhibits the explicitly given major and minor premises as edges labeled respectively j and n , and the tacit assumption of nonemptiness of a term as an edge from 1 to that term.

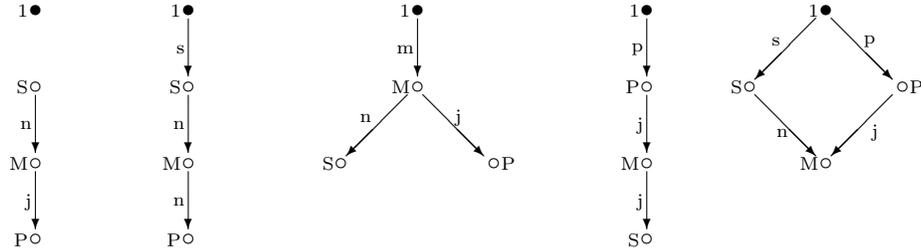


Table 2 (a) AAA-1 (b) AAI-1 (c) AAI-3 (d) AAI-4 (e) AAI-2

The conclusion of each axiom is then readily established by the existence of a path satisfying the conditions given above for propositions of the form XaY and XiY . In 2(a) the path $n; j$ from S to P verifies the conclusion SaP , while in 2(b) the paths s and $s; n; j$ from 1 to respectively S and P verify SiP , and similarly for 2(d). In the case of 2(c), AAI-3, even though the paths $m; n$ and $m; j$ to S and P diverge they both start with m which we interpret as a single individual, and hence suffice to witness SiP .

Table 2(e) is included to show the difficulty in meeting these conditions for AAI-2. In order to have two paths from from 1 to S and P that start out with the same edge to one of S, M, or P, there would need to be additional tacit assumptions besides those representable by edges from 1.

Now if by “categories” in the Introduction we had meant those of Aristotle, that would indeed be a slender thread connecting Aristotle to Boole. Rather, we had in mind that each of these five graphs freely generates a category C , namely the category whose objects are the vertices and whose morphisms are the paths. Composition of morphisms is the converse of path concatenation, that is, the concatenation $n; j$ corresponds to the composition jn and so on. Concatenation

is associative, whence so is its converse. Allowing the empty path at each vertex (not shown) ensures that every object has at least an identity morphism.

The role of the black vertex in each of the five categories C in Table 2 is as the object representing the forgetful functor $U : C \rightarrow \mathbf{Set}$, namely via C 's homfunctor. For all we know the objects of C may have much structure, including potentially billions of other members, not relevant to verification of the associated axiom's validity, which we forget by taking U to be $C(\bullet, -) : C \rightarrow \mathbf{Set}$. This represents each object v of C as merely the set of morphisms (edges) from \bullet to v sufficient to witness nonemptiness when needed, and each morphism $f : u \rightarrow v$ of C as the function $C(\bullet, f) : C(\bullet, u) \rightarrow C(\bullet, v)$ that maps each member $a : \bullet \rightarrow u$ of $U(u)$ to the member $fa : \bullet \rightarrow v$ of $U(v)$.

In all five examples \bullet has only one morphism to it from \bullet , namely its identity morphism 1_\bullet , making it a singleton. In 2(a) the remaining objects could for all we care be completely empty, which they can afford to be because the problem of existential import does not arise with presyllogisms. In 2(b)-(d) they all have at least one member each, witnessed by the forgetful functor.

In 2(e) S and P have respective members s and p . We specified that C is *freely* generated, which simply means that distinct paths remain distinct as morphisms, i.e. the square does not commute and M therefore has (at least) two members, namely ns and jp . There is therefore no visible means of support in this semantics of axioms for the conclusion SiP in AAI-2.

These examples suggest the following general criterion for any syllogism with universal premises and a particular conclusion. If the diagram formed by the premises has a least member then that member is the one whose nonemptiness must be assumed. In the absence of such a member, the assumption of nonemptiness even for all the terms of the syllogism is insufficient to warrant judging that syllogism as conditionally valid, with Table 2(e) serving as the generic example justifying this rule.

With these insights about conditionally valid syllogisms we can now understand the meaning of applying Rule R5 to them. Formally R5 takes AAI-3 to either EAO-1 or AEO-2 depending on whether it is applied to the major or minor premise respectively. It is a nice exercise to trace the progress of the term required to be nonempty under these deductions and evaluate its meaning in the light of Table 2. In any event we have the following.

Theorem 9. *Axioms A1 and A2 and Rules R1-R5 with R4 and R5 unrestricted completely axiomatize Aristotle's syllogisms.*

If the goal is to minimize the number of axioms, this is clearly the way to go. But if Table 2 has any appeal, keeping the three conditionally valid components separate by observing the restrictions on rules R4 and R5 is the better choice, because the unrestricted use of Rule R5 does violence to the logical structure of conditional syllogisms. As the abstract makes clear we prefer the latter, making Theorem 7 the *Hauptsatz* for this section, but this is more a matter of personal taste than anything deep.

Consideration of *mtj* as the passage in Table 1 between the two halves of the unconditionally valid syllogisms also reveals great violence: whereas in IAI-4 M

interpolates as PiMaS, mtj changes the interpolation radically to $Si\overline{M}a\overline{P}$.

If violence to logic is a concern at all, this would be a strong argument for stopping at Theorem 6 and putting up with five axioms and four rules. Or better yet, dropping the restriction on Rule R4, which was only to avoid having to insert a distracting crosslink from AAI-4 to AAI-1, thereby permitting replacing the ridiculously lonely Axiom A4 by A5, AII-1, which is much nicer as an axiom than AAI-4. We'll resume this train of thought in Section 2.8.

2.7 Normalization

Taking Table 1 as a complete enumeration of all valid syllogisms, conditional or not, validity of syllogistic forms is decidable simply by searching the table. It is nonetheless of interest to ask whether there is some less brute-force way of deciding validity, and moreover one which in the case of conditionally valid syllogisms automatically identifies the term whose nonemptiness must be assumed. We propose the following normalization procedure aimed at moving as briskly as possible from anywhere in Table 1 towards an axiom.

- Step 1. If there exists a particular premise, apply Rule R5 to render it and the conclusion universal.
- Step 2. If not already at an axiom, work through the list ojc , cj , ojn , cn of rule instances until finding one that is applicable, and apply it.
- Step 3. Work backwards through the list from that point applying the previously inapplicable instances if now applicable.

The result is the normal form.

Step 1 takes a long jump towards the given syllogism's axiom if any. Step 2 measures the remaining short distance from the axiom, and Step 3 then takes that many steps to reach it if possible.

While it might seem simpler to skip Step 2 and just apply the list in the other order, unless Step 1 lands at the maximal distance from an axiom among syllogisms with universal premises, this naive method would guarantee taking a step away from the axiom instead of towards it and then getting stuck.

Theorem 10. *A syllogistic form is valid if and only if its normal form is one of the four axioms.*

Proof. (Only if, completeness) By inspection of Table 1 the procedure carries all entries to one of the four axioms.

(If, soundness) All operations of the procedure are reversible and preserve validity in both directions. Hence the procedure cannot carry any syllogism outside the table into the table, whence any well-formed syllogism leading to an axiom must have been valid to begin with. \square

The axiom arrived at then tells which term in the original syllogism if any needs to be nonempty. Note that the subject-predicate-middle identities can

change in the course of the deduction; what matters is the actual term, such as *unicorn*, that arrives in that position. Thus if the unicorn started out as the predicate P and ended up as M in AAI-3, it is P in the initial syllogism that needs to be nonempty, not M.

While the procedure uses Rule R5 to leapfrog many steps, modifying it to work without it is a matter of coming up with an extension of Step 2 that can carry IAI-4 and OAO-3 etc. all the way round to AII-1. A case could be made for taking EIO as the axiom for that component, ignoring its figure, which can be reached from anywhere in the component in at most three steps.

This procedure may well be similar to at least some extant normalization procedures in the literature, but its use of Table 1 to make the reasoning easy to follow and its uniform method of handling the problem of existential import is to our knowledge new.

2.8 A role for subalternation

Taking Parson's history of the Square of Opposition [9] as definitive of the state of the art of our understanding of subalternation, all attempts to date at restoring it to the good graces of logical consistency have failed. With Table 1 giving a clear picture of what is at stake, we can analyze why, and see whether there is any way round the inconsistencies.

As motivation for even considering subalternation, Axioms A2-A4 of our system D4p serve to justify just the nine conditionally valid syllogisms, costing an average of a third of an axiom per syllogism. One might just as well dispense altogether with axioms and rules, along with issues of soundness and completeness, and simply list the nine along with the term each one needs to be nonempty, as typically done.

But we also struggled in D4p to connect IAI-4 and AEE-4, either via a fifth rule or giving up and settling instead for a fifth axiom. Given that the only purpose served by Rule R5 is to make one single connection, that makes it even more expensive than the conditionally valid axioms!

Subalternation to the rescue. As noted earlier it permitted deducing AAI-3 from AII-3, namely by licensing the strengthening of the minor premise from MiS to MaS. In similar fashion we may deduce AAI-4 from IAI-4, and AAI-1 from AAA-1 (weakening the conclusion instead of strengthening a premise). This dispenses with all three conditional axioms, leaving us with just Axiom A1 provided we keep Rule R5.

Moreover we can see at a glance which term needs to be nonempty for a conditional syllogism to hold, namely the subject of the sentence modified by subalternation.

And if in Figures 1 and 2, whose minor premise is S-M, we allow joint weakening of the minor premise and conclusion, we obtain the unconditionally sound Axiom A5, AII-1, from AAA-1, which Theorem 6 showed could replace Rule R5, with no requirement that any term be nonempty. We would then need only Axiom A1, Rules R1-R4, and subalternation in suitable form.

But as pointed out by Russell, one can expect anything to follow from an inconsistent rule, so perhaps we should not be amazed that so much can follow from this naive appeal to subalternation.

The constructive approach in such a case is to limit the offending construct to a domain of applicability small enough to avoid any inconsistency. Table 1 facilitates seeing where those limits should be set in the case of subalternation by explicitly listing all possible applications of Rules R1-R3, making it easy to see at a glance that none of them lead out of the region of the conditionally valid syllogisms. And it is equally easy to see that the region is closed under Rule R4 as well after noting that it is applicable only to AAI-1, AAI-4, and AAI-3 in that region, interchanging the first two and carrying the third to itself.

We are now in a position to improve our preliminary system D4p to D4. Rules R1-R4 are unchanged from D4p. In rules R5-R6, by “weaken” we mean replace a or e by i or o respectively, and the reverse for “strengthen”.

System D4

Axioms

A1. AAA-1 (Barbara)

Rules

R1. Convert any e or i premise.

R2. In Figures 1 and 3, obvert the major premise and the conclusion.

R3. In Figure 2, obvert both premises.

R4. In Figure 3, convert any e or i conclusion and interchange the premises.

R5. Strengthen a particular premise or weaken a universal conclusion.

R6. With premise SaM or SeM and conclusion SaP or SeP, weaken both.

The foregoing discussion, combined with the easily seen fact that Rules R5 and R6 are not applicable to any conditionally valid syllogism, makes the following immediate.

Theorem 11. *System D4 is sound and complete.*

Soundness entails consistency. The system has the additional benefit that Rule R5, as the only rule deducing a conditional from an unconditional syllogism, provides the term whose nonemptiness is required, namely the subject of the modified sentence, which Rules R1-R4 subsequently do not modify within the region of conditionally valid syllogisms.

The one axiom and six rules of System D4 is arguably a significant improvement over the four axioms and five rules of System D4p. And it is very much in the spirit of those sequent calculi that require only a single axiom, with AAA-1 as the natural counterpart in Aristotle’s syllogistic of $P \vdash P$.

3 Boole’s logic

3.1 The language problem

In 1847 George Boole wrote a short pamphlet [1] entitled *The Mathematical Analysis of Logic, being an essay towards a calculus of deductive reasoning*. In the preface he wrote “In the spring of the present year, my attention was directed to the question then moved between Sir W. Hamilton and Professor De Morgan; and I was induced by the interest which it inspired, to resume the almost-forgotten thread of former inquiries.”

Boole was referring to an acrimonious debate that had lately arisen between University College London’s Augustus De Morgan and Edinburgh University’s Sir William Hamilton, 9th Baronet. The dispute regarded a matter of priority in setting Aristotle’s logic on a firmer footing. It continued for nearly a decade until Hamilton’s death in 1856 at age 68.

In his preface Boole asked his readers not to dismiss his algebraization of truth as inconsistent but to consider the whole without concern for its departure from conventional algebra. Taking for his language that of algebra, namely addition, subtraction, and multiplication (but not division), he lists as his “first principles” three properties of multiplication: distributivity $x(y + z) = xy + xz$ over addition, commutativity $xy = yx$, and idempotence, $x^2 = x$.

Now whereas the first two of these properties are the laws making an additive (i.e. abelian) group a commutative ring, the third makes it what today we call a Boolean ring, the basic such being the ring of integers mod 2.

This ring witnesses the consistency of Boole’s logic.

Boole himself failed to recognize this, despite giving explicit definitions of both inclusive or and exclusive or. The stumbling block for both him and his critics appeared to have been their reluctance to accept $x + x = 0$ as a property of exclusive-or. Instead Jevons, Peirce, and Schroeder made the decisive move from rings to lattices by focusing on the logical operations of conjunction and disjunction as the basic operations of an algebra of logic. Peirce however believed incorrectly that every lattice was distributive, leaving to Schroeder the distinction of being the first to properly axiomatize a Boolean algebra. In modern language Schroeder defined it as a bounded distributive lattice (one with 0 and 1 as lower and upper bounds) that was *complemented*, meaning it had a unary operation \neg of negation satisfying $x \vee \neg x = 1$ and $x \wedge \neg x = 0$.

The lattice-based approach to Boolean logic then developed slowly for a few more decades. But then in 1927, three-quarters of a century after Boole’s ring-based proposal, Ivan Ivanovich Zhegalkin realized [14] that there was nothing wrong with that proposal after all. Zhegalkin proposed to define a Boolean operation as what we now call variously a Zhegalkin polynomial, algebraic normal form, or Reed-Muller expansion.

Nine years later, communication between Moscow and the west having been greatly slowed by the fallout from the 1917 revolution, topologist Marshall Stone independently realized essentially the same thing [13] in 1936. Stone had been writing up the duality he’d discovered between Boolean algebras and what we

now call Stone spaces. He was in the middle of explaining Boolean algebras by analogy with rings when it dawned on him that his explanation was more than a mere analogy: a Boolean algebra actually *was* a ring! As an approach to Boolean algebras rings are an alternative to lattices. Paul Halmos’s little book on Boolean algebras [4] adopts the ring-based definition without even mentioning the alternatives.

But there is yet another approach, via the propositional fragment of the intuitionistic logic proposed by Luitzen Egbertus (“Bertus”) Jan Brouwer [2] and formalized by his student Arend Heyting [5]. Its language is that of bounded lattices: conjunction, disjunction, 0 and 1, supplemented with implication $x \rightarrow y$. Implication is defined implicitly in terms of the lattice by either $x \wedge y \leq z$ iff $y \leq x \rightarrow z$ or its equational equivalent. Negation is defined explicitly by $\neg x = x \rightarrow 0$ and hence need not be included among the basic operations and constants, which are $\wedge, \vee, \rightarrow$, and 0 (1 is definable as $0 \rightarrow 0$). These are the properties required of a Heyting algebra. A Boolean algebra can then be defined as a Heyting algebra satisfying double negation, $(x \rightarrow 0) \rightarrow 0 = x$, or equivalently the law of excluded middle, $x \vee (x \rightarrow 0) = 1$.

These insights into Boolean algebra create something of an ontological dilemma: what exactly is a Boolean algebra, and how should it be defined? In particular, on what language should it be based? A language consists not of operations but of operation *symbols*. Should we take the symbols to be those of complemented lattices, $\wedge, \vee, \neg, 0, 1$, or of rings with unit, $+, *, 0, 1$, or of intuitionistic logic, $\wedge, \vee, \rightarrow, 0, 1$, or something else again?

The following offers one approach that is indisputedly neutral, yet not as impractical as it might at first seem. The approach exploits the fact that Boolean algebras form a locally finite variety, a rarity in algebras with binary operations that we discuss in more detail later. We first described the approach in a Wikipedia article [12]; its account here clarifies the connection with categories while fixing some technical problems as well as the problem that prior to now it was “original research”, officially *verboden* on Wikipedia even though we seem to have gotten away with it.

3.2 The language MaxBool

We define the language MaxBool as follows.

MaxBool is the language whose operation symbols are the finitary operations on $\{0, 1\}$.

The m input bits of an m -ary operation have 2^m possible combinations of values. The operation assigns one bit to each. Since the assignments are independent, there are 2^{2^m} possible combinations of such assignments. Hence MaxBool has that many operation symbols f, g, \dots of each finite arity $m \geq 0$.

Ordinarily the difference between syntax and semantics is that syntax is what you write while semantics is what the writing denotes in some universe. Here we are doing the opposite: interpreting the semantics of an m -ary operation in the universe $\{0, 1\}$ as an operation symbol. How could reversing the roles of syntax and semantics in this way make sense?

An m -ary operation on $\{0,1\}$ can be understood syntactically as a string of 2^m bits,³ namely its truth table. Every string of length 2^m constitutes an operation symbol, whose arity m is given by the logarithm to the base two of its length. Figure 5 gives five binary and seven ternary operation symbols as examples.

x	1010	x	10101010
y	1100	y	11001100
$x \wedge y$	1000	z	11110000
$x \vee y$	1110	$x \wedge y$	10001000
$x \rightarrow y$	1101	$y \wedge z$	11000000
		$z \wedge x$	10100000
		$(x \wedge y) \vee (y \wedge z) \vee (z \wedge x)$	11101000

Figure 5. Five binary and seven ternary operation symbols of MaxBool.

Above the line are the two binary projections or variables x and y and the three ternary such. (The sense in which variables are projections will gradually become clear, in Section 3.7 if not sooner.) Reading the columns from right to left it can be seen that the projections taken together are counting in binary notation (without leading zero suppression): from 0 to 3 for the binary projections and from 0 to 7 for the ternary ones. Each column of the projections uniquely identifies its horizontal position (distance from the right end) and can therefore be understood as a column index.

Below the line are some examples of truth tables: three binary and four ternary operations formed coordinatewise as Boolean combinations of operations above them. The last is the ternary majority operation, returning whichever of 0 or 1 is in the majority at its inputs.

The set of all bit strings of the same length is closed under the Boolean operations applied coordinatewise and hence forms a finite Boolean algebra. For those of length 2^m , the closure under the Boolean operations of the variables x_1, \dots, x_m as represented above is the whole Boolean algebra, and constitutes the free Boolean algebra \mathcal{F}_m on m generators.

A fringe benefit of taking the language to consist of all operations on $\{0,1\}$ is that separate provision for variables becomes unnecessary because they arise as projections. The variable names x, y, z above are notationally convenient synonyms for the first three variables in the infinite supply $x_1, x_2, x_3, x_4, \dots$. Each variable x_i for $i \geq 1$ is of arity i or more depending on context as treated in the next section.

3.3 MaxBool terms

As customary in algebra, a MaxBool term is either an atom or an application. The atoms include the constants 0 and 1 and the variables x_1, x_2, \dots , while the

³When succinctness is important symbols can be shortened by a factor of four by writing them in hexadecimal, exploited in the C program mentioned at the end of Section 2.2.

applications have the form $f(t_1, \dots, t_n)$ where f is an n -ary operation symbol and the t_i 's are MaxBool terms constituting the n *arguments* of f . Atoms are of height one while the height of an application is one more than that of its highest argument.

A novelty with MaxBool atoms is that every operation symbol is permitted to be an atom, subject to the constraint that all the atoms of a term have the same arity. The arity of a term is the common arity of its atoms. Since in MaxBool the variables up to x_m are representable as m -ary operation symbols, those variables are permitted in an m -ary term but no others.

An m -ary term has two personalities. One is as an ordinary term in Boolean algebra each of whose atoms is understood as an m -ary operation symbol applied to a fixed tuple (x_1, \dots, x_m) of variables, the same tuple for every atom. In that personality the variables are the real leaves of the term, having height zero (the atoms still have height one) and taking values in the Boolean algebra $2 = \{0, 1\}$. A Boolean identity is a pair of m -ary terms having the same value, 0 or 1, for all values 0 or 1 of the m variables. These form the theory T_m consisting of all Boolean identities involving at most m variables. The terms of height zero need not be mentioned in T_m because every atom is applied to the same m -tuple of them.

The other personality is as a constant term, one whose atoms play the role of constants valued in the free Boolean algebra \mathcal{F}_m . In this personality the height-zero variables of the other personality disappear leaving the atoms as the leaves, which serve as constants having fixed standard interpretations in \mathcal{F}_m . Variables are still permitted but only in the personality of constants, namely height-one m -ary operation symbols, e.g. 1010 and 1100 when $m = 2$.⁴ A Boolean identity is a pair of m -ary terms evaluating to the same element of \mathcal{F}_m . The intent is for these to form the same theory T_m as for the other personality.

3.4 Reduction of height-two terms in \mathcal{F}_m

In its second personality, an m -ary MaxBool term interpreted in \mathcal{F}_m denotes an element of \mathcal{F}_m , namely an m -ary operation or bit string of length 2^m .

Atoms denote themselves.

Terms $f(t_1, \dots, t_n)$ of height two are trickier. The m -ary operation denoted by such a term is the result of substituting each m -ary operation t_i for the i -th variable of the n -ary operation. In ordinary algebra this would entail substituting n m -variable polynomials for the variables of an n -variable polynomial, for example substituting $x + y$ for a and $x - y$ for b in ab , which for larger polynomials would require considerable simplification to put the result in a suitable normal form.

In MaxBool the entire process can be accomplished with a form of matrix multiplication.

Represent an n -tuple of m -ary operation symbols as an $n \times 2^m$ bit matrix

⁴The C program mentioned at the end of Section 2.2 represents S, P, and M as respectively 10101010, 11001100, and 11110000.

whose n rows are the n operation symbols. This will be matrix B in the following. The n -ary operation f will be the matrix A with $p = 1$; the general case of arbitrary p is just as easy to describe.

Given a $p \times 2^n$ matrix A and an $n \times 2^m$ matrix B , define their **product** $A \circ B$ to be the $p \times 2^m$ result of replacing each column of B by the column of A indexed by the replaced column. Formally,

$$(A \circ B)_{ij} = A_{it} \text{ where } t = \lambda k. B_{kj} = B_{*j} \text{ (column } j \text{ of } B).$$

Figure 6 illustrates this product using the examples from Figure 5. Take A to be all seven rows of the ternary example, and B to be the bottom three rows of the binary example, i.e. omitting the projections.

A		B	=	A \circ B
10101010		1000		1000
11001100		1110		1110
11110000		1101		1101
10001000	o	1110		1000
11000000		1101		1100
10100000				1000
11101000				1100

Figure 6. The product $A \circ B$.

Reading B from left to right, the successive columns of B pick out columns 7, 6, 2, and 4 of A (equivalently, the columns of A whose top 3 rows match the replaced columns of B), which are then assembled in that order to form $A \circ B$.

We retained the horizontal line in A to make it block structured, mainly to show the result of multiplying the upper block of A by B , namely B itself. It should be clear that the upper block of A is in fact the 3×2^3 identity matrix for this product, yet another role for the projections taken collectively.

3.5 An axiomatization of Boolean algebra

The following system, MB_m , uses the foregoing product to axiomatize T_m , the theory of Boolean algebra based on m variables. The atomic subterms of terms are m -ary operation symbols; all other operation symbols, namely those applied to n -tuples, are of arity $n \leq m$.

System MB_m .

A1. $f(t_1, \dots, t_n) = f \circ T$ where f is an n -ary operation symbol, the t_i 's are atoms, and T is the $n \times 2^m$ matrix whose i -th row is t_i .

R1. $\vdash t = t$.

R2. $s = t \vdash t = s$.

R3. $s = t, t = u \vdash s = u$.

R4. $s_1 = t_1, \dots, s_n = t_n \vdash f(s_1, \dots, s_n) = f(t_1, \dots, t_n)$ where f is an n -ary operation symbol.

R1-R3 realize the properties of an equivalence relation, which R4 extends to a congruence.

Theorem 12. MB_m is complete, meaning that it proves every equation $s = t$ of T_m .

Proof. We begin with the case when t is an atom, namely an m -ary operation. We proceed by induction on the height of s .

For height one s is an atom. But distinct m -ary atoms must denote distinct operations, whence the only s for which $s = t$ can hold must be t itself, which R1 supplies.

For height two s must equal a uniquely determined m -ary atom, which A1 supplies.

For height $h > 2$ with s of the form $f(s_1, \dots, s_n)$, the induction hypothesis supplies atoms t_1, \dots, t_n such that $s_i = t_i$ has been proved for each i . Furthermore A1 supplies $f(t_1, \dots, t_n) = a$ for some atom a , which must therefore be t . Hence by R4 and R3, $s = t$,

For general t , if $s = t$ is an identity both must reduce to the same atom a . We have shown that MB proves $s = a$ and $t = a$, whence by R2 and R3 we have $s = t$. \square

The theory of Boolean algebras is then $T = \bigcup_m T_m$.

Note that this union is disjoint. This is because every term contains at least one atom, whose length determines which T_m it came from. However every member of T_m has a counterpart in T_{m+1} obtained by duplicating every atom to make it twice as long. In that sense every equation appears infinitely often, once in each T_m beyond a certain point.

As with any of the other languages for Boolean algebra and their axiomatizations, a Boolean algebra is a model of T .

3.6 Discussion

Propositional logic is notorious for having relatively simple theorems with long proofs. In contrast proofs in MB simply evaluate every term directly, in a number of steps linear in the number of symbols in the term, with no need for auxiliary variables. In terms of running time however there is no free lunch: MB replaces exponentially long proofs with exponentially long operation symbols, slowing down each proof step exponentially. They are not infinitely long however, which is what would happen if instead of representing say x_1 as 10 in MB_1 , 1010 in MB_2 and so on we represented it once and for all as the infinite string 10101010...

This approach to axiomatization would also become infinite if used to axiomatize say commutative rings. This is because whereas there are only finitely many Boolean polynomials in the variables x_1, \dots, x_m , there are infinitely many integer-coefficient polynomials in m variables. With the requirement $n \leq m$ the schema A1 expands to $2^{2^n} \times (2^{2^m})^n$ axioms summed over $0 \leq n \leq m$. For $m = 0, 1, 2, 3, \dots$ this comes to respectively 2, 18, 4096, 4296016898, ... axioms when the schema A1 is expanded out.

3.7 The categorical basis for MB

As defined by Lawvere [6], an *algebraic theory* is a category having for its objects the natural numbers, such that for all n , n is the n -th power of 1. This means that for each n there exists an n -tuple (x_1, \dots, x_n) of morphisms $x_i : n \rightarrow 1$, the *projections* of n onto 1, such that for each n -tuple (t_1, \dots, t_n) of morphisms $t_i : m \rightarrow 1$ there exists a unique morphism $t : m \rightarrow n$ such that $x_i t = t_i$. The morphism t encodes the n -tuple of morphisms to 1 as a single morphism to n .

MaxBool adapts Lawvere's notion to the case of locally finite varieties, which permit a straightforward syntactic definition of composition. The category \mathcal{C}_{MB} has as objects the natural numbers, as morphisms from m to n the $n \times 2^m$ bit matrices, and as composition the matrix multiplication defined above. The identity at m is the $m \times 2^m$ matrix (x_1, \dots, x_m) , whose rows constitute the projections making m the m -th power of 1.

Figure 6 illustrates this in the case $n = 3$. The requisite 3-tuple (x_1, x_2, x_3) of projections $x_i : 3 \rightarrow 1$ are the first three rows of A . In the case when (t_1, t_2, t_3) consists of the 1×2^2 truth tables for respectively $x \wedge y$, $x \vee y$, and $x \rightarrow y$ at lower left of Figure 5, there exists a morphism $t : 2 \rightarrow 3$ such that $x_i t = t_i$, namely the 3×2^2 matrix B , such that $x_i \circ B$ is t_i for each x_i . Furthermore any other 3×2^2 matrix must contain a row not equal to t_i for some i , whence B is the *unique* such morphism.

The category \mathcal{C}_{MB} is easily seen to be isomorphic to the full subcategory of **Set** consisting of the finite powers of $2 = \{0, 1\}$. That m is the m -th power of 1 makes more sense when m and 1 are interpreted concretely as the power sets 2^m and 2^1 . From the perspective of language theory the power 2^m can be identified with the set of bit strings of length m , with 2^0 consisting of just the empty string. Each function from 2^m to 2^n can be understood as n functions from 2^m to 2, i.e. n m -ary operations, one for each of the n bits of the result. The corresponding matrix takes these to be the n rows, each the truth table of an m -ary operation.

Two paths from m to n that are mapped by composition to the same morphism from m to n form a *commutative diagram*. The equations of an algebraic theory are realized as its commutative diagrams to 1.

A set-valued model of an algebraic theory T is a functor from T to **Set** that preserves products. Such a functor interprets 1 as a set X , n as X^n , and n -ary operation symbols $f : n \rightarrow 1$ as functions $f : X^n \rightarrow X$. Functors preserve commutative diagrams, whence any equation holding in the theory holds of every model of the theory.

A Boolean algebra is any model of \mathcal{C}_{MB} as an algebraic theory.

Algebraic theories do not assume a locally finite variety. Composition is always definable by substitution. MB is special because composition has a syntactically simple form that justifies organizing it as an axiom.

4 Conclusion

My main goal in this paper has been to contribute a few hopefully novel insights into two logics neither of which could be considered low-hanging fruit the way computer science was in the early 1970s. I laid out a system of deductions for Aristotle's syllogisms designed to permit evaluating their soundness and completeness visually. And I gave a locally finite axiomatization of Boolean algebras that did not need to start from any particular point of view such as that they are lattices, or rings, or Heyting algebras, etc.

I also drew attention to how such things can look when viewed through the lens of category theory. In the case of syllogisms, category theory enters at such a trivial level that the purpose served is not so much insight into syllogisms as into category theory, which is often pitched inaccessibly higher. In the case of Boolean algebra it was the other way round: it was only by contemplating Lawvere's presentation of an algebraic theory as a category closed under products that I realized how the subject could be made independent of choice of basis for the operations.

Although there have been logicians who have seen deeper into logic, particularly within the past century, to date Aristotle and Boole stand out as the inventors of two logics each of which stood the test of time longer than any other. Though Aristotle's logic dominated the field for over two millennia, it is fair to say that Boole's logic has pushed it into something of a backwater. I frequently hear calls from type theory, programming languages, category theory, and proof theory advocating the same fate for Boole's logic, but from my perspective as an exponent of both the hardware and software of digital logic it will be a while yet.

References

- [1] G. Boole. *The mathematical analysis of logic, being an essay towards a calculus of deductive reasoning*. Macmillan, London, 1847.
- [2] L.E.J. Brouwer. Intuitionistische mengenlehre. *Jahresber. Dtsch. Math. Ver.*, 28:203–208, 1920.
- [3] J. Corcoran. Aristotle's natural deduction system. In J. Corcoran, editor, *Ancient Logic and Its Modern Interpretations, Synthese Historical Library*, pages 85–131. Springer, 1974.
- [4] P.R. Halmos. *Lectures on Boolean Algebras*. Van Nostrand, 1963.
- [5] A. Heyting. Die formalen Regeln der intuitionistischen Logik. In *Sitzungsberichte Die Preussischen Akademie der Wissenschaften*, pages 42–56. Physikalische-Mathematische Klasse, 1930.
- [6] W. Lawvere. Functorial semantics of algebraic theories. *Proceedings of the National Academy of Sciences*, 50(5):869873, 1963.

- [7] J. Lukasiewicz. *Aristotle's Syllogistic (2nd ed.)*. Oxford: Clarendon Press, 1957.
- [8] R. Parikh. A completeness result for a propositional dynamic logic. In *Lecture Notes in Computer Science*, volume 64, pages 403–415. Springer-Verlag, 1978.
- [9] T. Parsons. The traditional square of opposition. In *Stanford Encyclopedia of Philosophy*. Stanford University, 1997.
- [10] V.R. Pratt. Every prime has a succinct certificate. *SIAM J. Computing*, 4(3):214–220, 1975.
- [11] V.R. Pratt. Semantical considerations on Floyd-Hoare logic. In *Proc. 17th Ann. IEEE Symp. on Foundations of Comp. Sci.*, pages 109–121, October 1976.
- [12] V.R. Pratt. Boolean algebras canonically defined. *Wikipedia*, August 2006.
- [13] M. Stone. The theory of representations for Boolean algebras. *Trans. Amer. Math. Soc.*, 40:37–111, 1936.
- [14] I.I. Zhegalkin. On the technique of calculating propositions in symbolic logic. *Matematicheskii Sbornik*, 43:9–28, 1927.