Boole's Late Manuscript 'On the Nature of Thought' A Rewriting of the *Laws of Thought* without Uninterpretables

David Waszek
ITEM, École Normale Supérieure, Paris

Author accepted manuscript, 7 July 2025; published in *History and Philosophy of Logic*, 2025, DOI: https://doi.org/10.1080/01445340.2025.2550132

Abstract

This paper offers a commented edition of a late manuscript by Boole, which he likely put together in late 1863 as a response to William Stanley Jevons's criticisms of his system, in the hope of publishing his own views before Jevons's *Pure Logic* came out in early 1864. The manuscript, entitled 'On the Nature of Thought', is different in character from those that have been published to date. Boole does not attempt to rephrase his logic without algebraic symbolism. Instead, he amends the general problem-solving method presented in his *Investigation of the Laws of Thought* so that it obeys 'the express condition that no forms are to be employed which are not interpretable', without, however, making any change to his underlying logical calculus (in particular, Boole does not adopt Jevons's inclusive reading of +). Though quite terse in places, the manuscript is largely successful with respect to its stated goals. Moreover, it sheds light on Boole's thinking about interpretability, highlighting a tension in Boole's work between an indirect notion of interpretation based on the method of development, and a compositional notion of interpretation that, as the manuscript shows, Boole ended up emphasizing at the end of his life.

Keywords. George Boole; W. Stanley Jevons; uninterpretable expressions; unpublished manuscript

1 Introduction

The manuscript edited below, entitled 'On the Nature of Thought', is preserved among Boole's mathematical papers at the Royal Society in London. Written in the last years of his life, likely – as I shall argue – in the context of his correspondence with William Stanley Jevons (*Grattan-Guinness 1991*), it may well be Boole's last manuscript on logic. Its goal is to offer a rewriting of the *Investigation of Laws of Thought* in which every intermediate step is 'interpretable'; it is largely successful on its own terms, and in some ways brings Boole closer to *Schröder 1877*. Though noticed by previous researchers,[1] it has been left out from editions of Boole's manuscripts[2] and its significance has not been previously recognized – probably because it looks superficially too similar to the *Investigation of the Laws of Thought*, whereas readers of Boole's *Nachlass* were focused on his more radical, but less compelling attempts to rephrase his logical system entirely in natural language.[3]

The commentary offered below is meant as a reader's guide. Section 2 sets the stage: it reviews the main issues driving Boole's unpublished attempts at producing a sequel to the *Investigation of the Laws of Thought* (henceforth *LT*), and situates 'On the Nature of Thought' among these various projects. Section 3 then introduces the manuscript of 'On the Nature of Thought' (henceforth often abbreviated as NT) and argues that it was written in 1862–1864 – most likely in late 1863 – and is connected with Boole's correspondence with Jevons before the publication of the latter's *Pure Logic* (*Jevons 1864*).

The rest of the commentary delves into the technical contents of NT. As a preliminary, Section 4 reviews the multiple ways Boole used 'interpretation' and 'interpretable', highlighting how he altered and clarified his terminology in 'On the Nature of Thought'. With this in hand, Section 5 compares NT point by point with *LT* and shows how the manuscript carries out its goals. We shall see that NT builds on previous work by Boole, in particular on the little-read Chapter X of *LT,* while adding crucial ingredients – among other things, careful proofs of the validity of individual steps of Boole's method, a direct proof of his theorem of 'development' (one that does not rely on Taylor's theorem or otherwise on numerical algebra), and a new treatment of his much-criticized use of division in logic.

The edition itself is presented in an appendix. Throughout this paper, passages from 'On the Nature of Thought' are referred to by their manuscript page number, surrounded by square brackets (for instance, Boole's new proof of his theorem of development is on pages [29]–[33]).

2 Boole's logical projects after the *Investigation of the Laws of Thought*

Between the 1854 publication of his *Investigation of the Laws of Thought* (*LT*) and his death in 1864, Boole worked on and off on follow-up works on logic – expository works that, in contrast to *LT,* would be written 'for the general public' and not just 'for mathematicians', as

[1]*Hesse 1952,* which remains the single best source on Boole's late logical writings, quotes from it, and Grattan-Guinness and Bornet mention it without comment (*SMLP,* 205).

[2]Rush Rhees published a few as *Studies in Logic and Probability* (*1952*, henceforth *SLP*), and Grattan-Guinness and Bornet published a broader selection as *Selected Manuscripts on Logic and its Philosophy* (*1997*, henceforth *SMLP*). The majority of Boole's papers remain unpublished. For more on the state of Boole's *Nachlass*, see *SMLP,* xviii-xxv.

[3]The 20th-century neglect of 'On the Nature of Thought' may have an older origin as well. When Boole's daughter Alicia, together with her husband H. J. Falk, examined and reorganized her father's papers in 1889–1896, she produced typewritten transcriptions of a number of logical manuscripts (*SMLP,* xx), but not of 'On the Nature of Thought'. As a result of these transcriptions, the manuscripts Boole and Falk selected became much easier to study, which likely contributed to their being prioritized by later researchers.

he put it in one of his drafts (*SMLP,* 120). His correspondence, his manuscripts, and a note by his wife Mary Everest Boole amply testify to this project, though he never published any of it. But what presenting his system to a non-mathematical audience *actually amounted to* varied considerably across his attempts. In the manuscripts that have received the most scholarly attention so far, he tried to rephrase his logic entirely in 'ordinary language' (*SMLP,* 64), in a manner 'free from mathematical symbols' (*SMLP,* 124). By contrast, 'On the Nature of Thought' is just as symbol-heavy as *LT*. Roughly speaking, it is written for a reader like Jevons – one who had no objection to the use of symbols *per se* but had little familiarity with what Boole saw as more advanced mathematical methodology, which licensed the use of formal laws even when they led to (at least apparently) uninterpretable expressions.

In order to situate 'On the Nature of Thought' among Boole's various efforts at writing a follow-up to *LT,* the main goal of this section is to clarify exactly what the challenges were that his late logical projects were facing. These can conveniently be grouped under three headings: the use of symbols (Section 2.1); the use of formal reasoning (Section 2.2); and the issue of whether Boole's system depended on a specific branch of mathematics, namely arithmetic (Section 2.3). As we shall see, *foundational* concerns – how should logic be defined in contrast to mathematics, and does Boole's logic actually *depend* on mathematics thus construed? – were intertwined with merely *pedagogical* ones – how can Boole's system best be made intelligible to readers with no mathematical training?

2.1 Logic, mathematics, and the use of symbols

Our first task is to understand what Boole came to mean by the terms 'logic' and 'mathematics' – an issue on which his late manuscripts depart from his earlier published work – and how the distinction between mathematics and logic relates to the use of algebraic symbolism (for a fuller discussion, see also *Hesse 1952,* section 3).

Boole's published books on logic are explicitly presented as 'mathematical': the first one is entitled 'The Mathematical Analysis of Logic' (*MAL,* 1847), and the subtitle of the second refers to the 'mathematical theories of logic and probabilities' (*LT,* 1854). What 'mathematical' meant here deserves clarification, however. Boole's mathematics was rooted in 'symbolical algebra'[4] as defined by the mathematician Duncan Gregory *1840*: it was based on the possibility of representing and investigating operations of any kind – not just operations on numbers and magnitudes – using algebraic symbols. This led Boole to a broad definition of mathematics, as is made clear, for instance, by a lecture he delivered in 1851 (*SLP,* 195):

> I speak here, not of the mathematics of number and quantity alone, but of mathematics in its larger, and I believe, truer sense, as universal reasoning expressed in symbolical forms, and conducted by laws, which have their ultimate abode in the human mind.[5]

[4]For an introduction to 'symbolical algebra' as it developed in Britain at the time, see *Durand-Richard 1996* or *Parshall 2011,* which provides a good introductory bibliography at the beginning of its endnotes.

[5]The reference to 'the human mind' reflects Boole's view that the laws of symbols, in arithmetic just as much

Later in the same lecture, Boole added (*SLP,* 209):

> [. . .] it is simply a fact that the ultimate laws of Logic – those alone upon which it is possible to construct a science of Logic – are mathematical in their form and expression, although not belonging to the mathematics of quantity.

Thus, initially, what made his logic 'mathematical' for Boole was the use of a language of symbols subject to laws of combination – the use, we might say, of a kind of universal algebra (though the term would only be introduced later in the century). In this spirit, a non-mathematical version of his logic would indeed have to be one without symbols, but given that he regularly described the 'ultimate forms and processes' of logic as 'mathematical' (*LT,* 12), it is doubtful that he would, at this stage, have welcomed such a project.

In one sense, this conception of the relation between mathematics and logic was an unsurprising view for Boole to take. As mentioned, the starting point of his logical work was a development internal to mathematics, namely the rise of algebraic symbolization for operations that were not numerical or geometrical in nature. We would expect his logic to initially appear to him as a further step in the expansion of symbolical algebra, seen as a branch of mathematics, to more and more areas.

In another sense, however, this is a highly problematic position. For one thing, it raises foundational issues: should logic not be *prior* to mathematics? For another, it sits uneasily with Boole's extensive reflections on the relation between language and thought. Already in *LT,* it is clear that he saw no difference *in nature* between the symbols of algebra and the words of natural language – all of them were *signs*, he argued, and despite their confusing ambiguities, English words such as 'and' and 'or' were subject to laws of combination just like $+$ and $\times$ are in arithmetic. But if there is a continuity between natural and symbolic languages, it becomes unclear where exactly the boundary of mathematics should lie.

Boole's late drafts are responsive to these concerns. His strategy was to distinguish two senses of the term 'logic' – a move that emerges at the latest around 1856, in what may be his earliest substantial treatment of the subject after the *Investigation of the Laws of Thought*,[6] and remains in evidence all the way to 'On the Nature of Thought' (see [1]–[2]). What Boole called 'logic' *simpliciter* in his earlier work gets redefined as logic in a first, 'narrower' sense, which is about 'relations of Class e.g. those of genus and species, whole and part, identity and difference, & so on' (NT, [2]). Logic in this sense, Boole thought, is 'the Science of the Laws of Thought *as expressed in the terms of ordinary Language*' (NT, [2]): we can study it, as Boole did in *LT,* by examining the rules that, in natural language, govern such words as 'and', 'or', predicates, and adjectives.

Now, in order to express and systematically study the laws of this narrow 'Logic of Class', the earlier Boole would have said that we need *mathematics*, under the guise of the symbolical language of algebra. Instead, his late drafts argue that the 'Logic of Class' is 'subordinate' (*SMLP,* 129) to 'Logic in its primary and most general sense', which, he explained (*SMLP,* 126),

as in logic, come from the laws governing the mental operations that the symbols express – the normativity of the laws of symbols comes from the mind.

[6] This long manuscript was published as Chapter V of *SMLP;* for the relevant passage, see p. 66.

> might be said to be the philosophy of *all* thought which is expressible by signs, whatever the object of that thought, whatever the nature of these signs may be. [. . .] There is a philosophy of signs which governs and explains all their particular uses and applications, – which is equally manifested in the forms of ordinary speech and in the symbolical language of mathematics.

Logic in this second, broader sense is *not* a branch of mathematics. It is *simultaneously* prior to the 'Logic of Class' and to the symbolical language of algebra as it is used in (say) arithmetic.[7]

In Boole's new framework, it is no longer clear that an algebraic, symbolic language is *necessary* for his logic. He no longer described the laws of the 'Logic of Class' as essentially mathematical, that is, as requiring a symbolic language to be brought out. Instead, they are logical in his *broad* sense of the term – they are laws of signs, be these verbal or symbolic, that reflect laws of thought. Whether one uses symbols or just words to express such laws and clarify their nature then seems to become a mere choice of notation with no *foundational* significance. One can use symbols, as Boole did in 'On the Nature of Thought', without making logic dependent upon mathematics in any deep sense; conversely, one can choose to forgo symbols for pedagogical reasons, as Boole attempted – up to a point – in many other late drafts.[8]

Nevertheless, matters of notation were highly consequential for Boole, and even in his more radical attempts to write a non-symbolic logic, he never thought that using ordinary language instead of symbols would make no difference. Crucial to *LT* was a systematic problem-solving method, which Boole saw as *essentially tied* to his development of a symbolic calculus (*Waszek and Schlimm 2021*). Even when presenting the principles of his logic in verbal form in his later drafts, he maintained (*SMLP,* 86) that the method required to systematically solve problems

> is however so dependent upon language[,] and its exposition is so much facilitated by the employment of a proper system of notation[,] that it becomes if not necessary at least highly important to introduce such a system and avail ourselves of its aid in expression before proceeding further into the analysis of Logical Reasoning.

[7] Together with his distinction between two senses of logic, Boole developed a somewhat subtle account of their relative priority. On the one hand, he apparently thought that there were multiple independent instances of law-bound uses of signs – in other words, multiple independent developments of logic in its broader sense. For instance, numerical algebra and the narrow 'Logic of Class' (as studied in *LT*) appear to constitute two such independent domains for him. On the other hand, he granted that 'the Conception of Class is antecedent in the order of thought to all other scientific conceptions' and (at least for the sake of the argument) that 'we can throw every demonstration into a syllogistic form' (*SMLP,* 128). Accordingly, there is a sense in which all reasoning, including in mathematics, is reducible to the narrow 'Logic of Class'. Traces of this discussion appear in 'On the Nature of Thought', p. [4]. On this matter, see *Hesse 1952,* 65–67.

[8] The pedagogical argument for doing so, reiterated by Boole repeatedly, is that symbols would baffle readers with no mathematical training; for instance, he wrote: 'There are [. . .] persons, and private correspondence has acquainted me with some, who are interested to know all that can be known of the intellectual constitution and who yet may be unwilling or unable to pursue trains of reasoning conducted by symbols in which the laws of that constitution are[,] if we may use such an expression[,] embodied' (*SMLP,* 124).

Moreover, without such a system of notation, connections between the various laws of logic lack surveyability (*SMLP,* 86):

> It is important also to present the laws of the intellectual operations not as isolated truths but as constituent parts of a system governed by pervading relations; and to this object the employment of an adequate system of notation is equally requisite.

So Boole's ambition in non-symbolical presentations of his logic was never more than to offer 'an account freed *as far as possible* from the language of symbols' (*SMLP,* 64, my emphasis).

2.2 Formal reasoning

A central feature of Boole's account of algebra – be it in mathematics or in his logic – is that it licenses following formal laws even when they lead to expressions that are uninterpretable in the domain of investigation. One might expect him to see this as a characteristic feature of reasoning *with symbols*, one that has no place in verbal reasoning. Surprisingly, however, Boole sometimes retained it even in verbal expositions of his logic. Conversely, 'On the Nature of Thought' uses symbols liberally, but its explicit goal is precisely to *avoid* any dependence on uninterpretable formal reasoning.

Boole's conception of algebra, rooted in the work of such mathematicians as Duncan Gregory, is roughly as follows. First, one uses signs, such as + and −, to stand for operations on a certain domain. These operations are subject to conditions of possibility that depend on the domain in question: for instance, in arithmetic, a greater number cannot be subtracted from a lesser one; in logic, classes that are not disjoint cannot be 'aggregated' (the operation that Boole denoted by +). Second, one reads off the basic laws governing such operations in each domain from cases in which all operations involved are possible (an approach Boole describes explicitly below: NT, [22]). For instance, one observes that the equality[9]

$$x + y - z = x - z + y$$

holds when x and y are disjoint and z is included in x. Finally – this is the crucial step – one is allowed to follow such laws *formally* even when restrictions on the performability of operations are no longer satisfied: for instance, one may use the equality just given even when z is included in $x + y$ but not in x, in which case the right-hand side is, on its face, uninterpretable, as the first subtraction cannot be performed. Elsewhere, I dubbed this license Boole's *Principle of Formal Reasoning*, or PFR for short (see *Waszek 2025,* where this principle is examined thoroughly).

Boole came to consider that this feature of formal reasoning had nothing to do with symbolic languages in particular. In his *Treatise on Differential Equations* (*Boole 1859*), when discussing the principle that 'the mere processes of symbolical reasoning are independent of the conditions of their interpretation', he wrote that 'it would be an error to regard

[9] Boole assumes what we would call left-associativity, so this equality means $(x + y) - z = (x - z) + y$.

it as in any peculiar sense a mathematical principle. It claims a place among the *general* relations of Thought and Language' (*Boole 1859,* 389, his emphasis) – in other words, to use Boole's late terminology, it belongs to logic in its broad sense. A manuscript from around 1856, already discussed above (it is Boole's earliest sustained attempt to rewrite his logic for non-mathematicians), contains a revealing discussion of formal reasoning in the context of a verbally formulated logic. Boole's discussion starts from the principle that there are formal laws that are independent 'of the special meaning or content of the concepts involved', such as those expressed by Aristotle's argument schemas: 'Aristotle[,] in expressing the terms of propositions by letters[,] set an example of [the adoption of this principle] which nearly all subsequent writers have followed' (*SMLP,* 71). Such a principle, Boole added, holds far beyond syllogistic, but with a subtlety that is not usually noticed (*SMLP,* 72):

> the meaning of words is not always wholly independent of the form of the expression in which they occur. Thus the formula 'Xs and Ys' does not express an intelligible concept unless the symbols connected by the conjunction *and* be interpreted to signify classes of things wholly distinct.

From there, Boole introduced what I call his Principle of Formal Reasoning (*SMLP,* 72):

> This leads us to the threshold of perhaps the deepest question in the Philosophy of Logic viz. Are we bound when conducting the processes of reasoning by means of language to keep constantly in mind the conditions of interpretability and therefore to employ forms which impose such conditions then only when those conditions are actually satisfied? [. . .] Or is the intellectual procedure in Logic governed solely by a reference to abstract forms and laws? [. . .] I hold, as will be evident [. . .], the latter view.

Notice the conspicuous absence of symbolical algebra in the entire discussion: Boole thought that the formal laws of verbal signs (such as 'and' in English) license uninterpretable reasoning steps just as much as the formal laws of algebraic symbols do.

There is little indication that Boole ever doubted his PFR. He realized, however, that it raised objections, especially from 'those who knew nothing of mathematics' (*Grattan-Guinness 1991,* 26), as he put it in a letter to Jevons that I shall discuss below. In *LT,* he had described the PFR as a 'law of the mind', in line with Whewell's treatment of Peacock's Principle of Permanence (a close relative to Boole's PFR) in the *Philosophy of the Inductive Sciences* (*Richards 1980,* 350–353*; Waszek 2025,* section 6). In some drafts, he attempted to offer justifications of the PFR that would be more compelling for readers. For instance, he suggested that it could be justified 'upon the ground of a large induction upon the actual processes of mathematics' (*SMLP,* 146) – in other words, because we see that it leads to correct results in many cases. He also tried to argue that the *possibility* of *reinterpreting* symbols in such a way that uninterpretable algebraic manipulations in one domain become fully interpretable in another, could legitimate the manipulations in the first domain (*SMLP,* 147–8*; SLP,* 227–8), thus defending special cases of the PFR.

In 'On the Nature of Thought', Boole did not try to justify the PFR; his design was to avoid relying on it entirely. By contrast to *LT* – he wrote – in which 'although the final *results* of the method admit of being interpreted [. . .] into general theorems of the Logic of Class[,] yet the intermediate *processes* by which these results are obtained are not always so interpretable', his aim in NT was 'to show that it is possible to pass from the same system of primary laws to the same final results without transgressing on the way the limits of that kind of Thought with which the Logic of Class is concerned' (NT, [8]–[9]). He maintained, though, that going through uninterpretable steps as done in *LT* was legitimate – in one passage, for instance, he described one of the amendments offered in NT as 'one which in the purely formal procedure of thought is wholly unnecessary and which is made here only in order to enable us to present under a certain condition of interpretability forms of Thought which as to their essence are independent of such conditions' (NT, [43]). Boole did, at one point, also suggest that the procedures presented in NT would *vindicate* detours through uninterpretable steps. When presenting the laws of his symbols, he wrote (NT, [22]):

> Our knowledge of them is derived from cases in which the elementary operations are *possible*. [. . .] That the relations possess a truth beyond this will be shown hereafter.

This is not a theme Boole returns to in NT. He likely did not mean to offer a full justification of the PFR as it applies to his algebra of logic; plausibly, he just meant to vindicate those uninterpretable steps that he used in *LT* by justifying their fruits, i.e., by showing that one could obtain the same conclusions through fully interpretable reasoning.

2.3 Logic and arithmetic

On one point, all of Boole's late logical manuscripts agree: his logic did *not* depend on arithmetic, and insofar as *LT* made it seem that it did, it needs to be corrected. This theme plays an important role in the introduction to 'On the Nature of Thought' (pp. [6]–[8]) and is closely connected to, but distinct from that of formal reasoning.

To understand why this issue arises, we need to go back to *LT*. There, Boole noted that the 'laws' of his logic agree with those of a specific kind of numerical algebra, which he sometimes called 'dual algebra' (*SMLP,* 91–95), in which letters can only stand for the numbers 0 or 1 (so that letters satisfy Boole's law $x^2 = x$, on which more below). Importantly, despite its name, Boole's dual algebra is *not* arithmetic modulo 2: only *individual letters* are limited to the values 0 and 1, not complex expressions, which can stand for any number. From Boole's point of view, even though the operations denoted by + and − are subject to the same 'laws' in both domains, they are subject to different restrictions: for example, any two numbers can be added – so the operation + is always possible in dual algebra, with no restriction – whereas two classes can only be 'aggregated' in some cases, namely if they are disjoint. Thus, $x + x$ is always defined in the domain of numbers but never in that of classes (unless x is the empty class). Nevertheless, as discussed in the previous section, Boole believed that one would always get correct results if one reasoned formally according to the laws while *ignoring* restrictions on the possibility of operations. From the perspective of such formal reasoning, the algebras for logic and dual algebra are exactly the same: accordingly, theorems – Boole believed – could be transferred from one to the other (*LT,* 69–70). This kind of transfer is the main way in which the symbolical nature of algebra is employed in the work of Gregory, an important source of Boole (for more on this, see *Durand-Richard 2022,* 103–106 and *Waszek 2025,* section 6). Crucially, Boole used this license to ground the cornerstone of his system, his theorem of development (*LT,* 72–73), to which we shall return below.[10]

From a foundational point of view, Boole did not think that this way of proceeding actually made his logic *depend* on arithmetic. It appears that he reasoned as follows. First, he seemingly believed that, in principle, any methods or theorems that held in dual algebra could be *deduced* from the formal laws of combination of dual algebra – thus making what we might describe as an implicit completeness assumption. Second, since the laws of dual algebra are the same as the laws of his logic, any theorem holding in dual algebra, and hence derivable from the laws of dual algebra, would be derivable from the laws of his logic as well. Of course, deriving these theorems from the laws might involve going through steps *uninterpretable in logic* – if not in dual algebra – but this was licensed by the principle of formal reasoning *independently of the very existence of an arithmetical interpretation*.[11]

From this perspective, Boole saw the role of dual algebra in his earlier work as fundamentally *heuristic*, a point he emphasized repeatedly in his drafts. For instance, he wrote (*SMLP,* 120, Boole's emphasis):

> Is the analogy which has been referred to above as connecting the intellectual operations in the two distinct spheres of Logic and of the special Algebra under consideration essential to the full development of the former science? I have certainly never regarded it as such, freely as I have employed it for the discovery of *methods*.

Or again (*SLP,* 211):

> Some years ago I published a work in which the Science of Logic was developed in mathematical forms. This mode of expression was not founded upon any supposed relationship between the conceptions or ideas about which logicians and mathematicians are respectively conversant, but upon the fact established by actual examination that the formal laws of Thought in Logic are the same as those of Algebra or the science of Number would be if it were conversant

[10] Boole in fact offered two different justifications of his theorem of development, both relying on arithmetic. One of them consists in importing Taylor's theorem (*LT,* 72, footnote*;* see also *MAL,* 60*,* which talks of 'Maclaurin's' rather than Taylor's theorem). The other simply consists in noting that his formula of development holds in the arithmetical interpretation in which literal symbols can only take the numerical values 0 or 1, and in concluding from this that it holds in the logical interpretation as well.

[11] Notice that the point just made – that Boole saw the detour through dual arithmetic as licensed by formal reasoning, but in principle dispensable because one ought to be able to rely on formal reasoning directly – is a historical one: it is about what Boole believed and wrote repeatedly in his published writings and drafts. It should not be confused with the question of what we might find mathematically cogent from a retrospective point of view. From the latter perspective, the principle of formal reasoning would have to be rejected; on the other hand, the transfer of results from dual algebra to the algebra of classes turns out to be load-bearing, as it is robust in the kinds of cases where Boole is invoking it, though it would have to be justified in a completely different way (*Burris and Sankappanavar 2013*).

> not about all numbers but only about those which we designate as Unity and Nothing. [. . .] I think it possible that but for the light of this analogy I should have failed to raise upon the basis of formal law any such structure of methods and results as is exhibited in my work [. . .].

To avoid any misunderstanding, however, his drafts after *LT* generally attempt to present his logical system independently of any consideration of its arithmetical interpretation. This is true of 'On the Nature of Thought' as well; see, for instance, how careful he is when introducing his use of 0 and 1 in logic (NT, [16]–[17]):

> This mode of expressing the conceptions of Nothing & Universe is adopted from the 'Laws of Thought' where it is employed upon the ground of the identity which is there proved to exist between the formal laws of the conception Nothing in Logic and the number 0 in the science of Number and between the formal laws of the conception Universe in Logic and the number 1 in Arithmetic. Here[,] though we retain the notation[,] we dismiss for the present the analogy. No part of the following exposition would be affected if we represented the conceptions of Nothing and Universe by definite literal symbols just as we here express ordinary class conceptions, provided that those symbols were used in subjection to formal laws founded upon their peculiar interpretation – laws which would prove identical with those we shall establish for the symbols 0 and 1.

What marks out 'On the Nature of Thought' is that Boole also avoided uninterpretable formal reasoning entirely, and so had to carefully provide proofs each step of which was interpretable.

2.4 'On the Nature of Thought' among Boole's late drafts

We discussed above several features of Boole's published works on logic that made them forbidding – in Boole's eyes – for readers with limited mathematical training: the use of a symbolic language borrowed from mathematics; the use of formal reasoning even when it leads to uninterpretable expressions; and the wholesale transfer to logic, licensed by formal reasoning, of results from numerical algebra (most significantly to obtain the theorem of development). Eliminating these features from his logic involves trade-offs, and Boole made different choices in different drafts.

Many of Boole's logical drafts from after *LT* seem intended for 'logicians from Oxford' and attempt for the sake of such readers to 'avoid symbolism' (*SLP*, 212), as he put it in one manuscript. I already mentioned that in such manuscripts, despite the absence of a symbolic language, Boole often defended formal reasoning, though he no longer used it to import results from arithmetic and instead developed logic on its own terms. As we saw, however, the elimination of symbolism is not without its problems; in particular, from Boole's point of view, his symbolic calculus is crucial to enable a general method for solving logical problems. Thus, while some drafts do omit symbols completely,[12] others adopt the strategy

[12] See 'Logic and Reasoning', published as Chapter VI of *SLP*.

of introducing them eventually, but as gently and as late possible.[13]

In 'On the Nature of Thought', Boole made a different compromise. This manuscript is responsive to many of the same concerns as his other drafts. In fact, much of the material in the more philosophical introduction of NT (roughly [1]–[14]) is very close to other manuscripts by Boole, sometimes verbatim;[14] I will argue below that NT was likely put together in haste, on the basis of older material, in response to Jevons. However, the goals and technical content of NT are distinctive. Its intended reader seems to be someone like Jevons, who is not opposed to symbolization as a matter of principle, but sees the specific symbolic methods that Boole uses as 'unintelligible and mysterious' and as relying on 'mistaken but not entirely false analogies' (with arithmetic), as Jevons put it in his letters to Boole (*Grattan-Guinness 1991,* 25). Accordingly, 'On the Nature of Thought' uses symbols freely and exposes Boole's methods in their full generality, but 'without transgressing the conditions of interpretability' (NT, [49]), and without relying on the analogy with arithmetic.

3 The manuscript and its dating

The manuscript edited below is composed of 50 handwritten pages (mostly in ink in Boole's hand, with a few pencil annotation, some possibly in another hand) which, as of this writing, are kept at the Royal Society Archives in an individual folder labelled 'On the Nature of Thought, 1–50, W4 + C43'. The first page bears the title 'On the Nature of Thought'; there is no date. This section offers a material description of the document and, as far as is possible, reconstructs its history. As we shall see, the manuscript can be dated with high confidence to 1862–1864, and likely to late 1863, in the context of Boole's correspondence with Jevons.

3.1 Material description of the manuscript

First of all, the complicated history of Boole's *Nachlass* requires caution. Over multiple decades, numerous people examined, reorganized, and attempted to catalogue parts of Boole's manuscripts; what is likely the most extensive effort is due to Boole's daughter Alicia, together with her husband H. J. Falk, between 1889 and 1896 (see *SMLP,* xviii–xxv, 203–206). Thus, even the possibility that sheets currently labelled as a single text are in fact an archival artefact – assembled from unrelated material by later readers, rather than composed by Boole himself – cannot be dismissed out of hand.

[13]See 'On the Foundations of the Mathematical Theory of Logic and on the Philosophical Interpretation of Its Methods and Processes', partially published as Chapter V of *SLP* and in full as Chapter V of *SMLP*.

[14]In particular, it reproduces extensive passages from manuscript C57 (Chapter IX of *SMLP,* on which see pp. 213–214), usually thought to have been intended as the first chapter of a new book-length treatment of logic by Boole. Parallel passages can also be found in manuscript E.2, published as Chapter VI of *SLP* (e.g., compare pp. 212 and 215 of *SLP* with, respectively, pp. [1] and [3] and pp. [11]–[13]), which may have been used to prepare manuscript C57 or conversely. A systematic comparison, perhaps using the methods sometimes used for literary drafts (*Haffner 2024*), would be needed to establish the exact relationships between these various manuscripts.

Luckily, we can be confident that the 50 pages edited here originally belonged together. The sheets bear three different kinds of page numbering and labels. First, pages are numbered in pencil from 1 to 48, two consecutive pages bearing the number 13 and the last page being unnumbered; this numbering is plausibly due to Boole himself.[15] Second, the 50 pages are marked B.1 to B.50 in ink, seemingly not in Boole's hand; this labelling likely dates from one of the earliest examinations of Boole's papers, as a partial catalogue of them that was drawn up in the first years after Boole's death already lists 'On the Nature of Thought' under the signature 'B'.[16] Third, confusingly, one finds two additional classification marks (in pencil): 'W4' on the first page, 'C43 (?a)' on page B.49 – the penultimate page – and '? C43(b)' on the final page B.50 (hence the labelling of the text as 'W4 + C43' in the Royal Society Archives). These pencilled marks are consistent with the labelling conventions used by Alicia Boole and H. J. Falk, hence are presumably due to them; if so, the use of a different label for the last two pages indicates that they found these pages separated from the rest (*SMLP,* 205–206). Nevertheless, given that the 'B' numbering very likely predates the Boole/Falk work, and that there is clear textual continuity between pages B.48 and B.49, there is little reason to believe that the 50 pages were artificially assembled after Boole's death. The entire manuscript is also cohesive from the point of view of content.

The sheets are sometimes made of several pieces of paper pasted together, and hence are of uneven length. Boole's practice when finalizing a text for publication seems to have been as follows:[17] when a passage needed substantial corrections, he rewrote the page either in full or in part (by cutting out a piece of it and pasting a new piece of paper in its place), in both cases leaving the surrounding pages untouched. This process causes distinctive errors, usually flagged in footnotes to the edition below: sometimes words (or even a full line) are repeated on both sides of a page boundary, or a couple of words go missing.

The manuscript bears a few pencil annotations – some mere vertical lines in the margin flagging, for instance, the location of a missing equation number – the most significant of which are described in footnotes to the edition. Some of these annotations appear to be by Boole himself, especially the one on the back of p. [1], which reads:

> Explain on p 7′ the nature of *representative* thought as distinct from formal thought.

Most of them, however, are likely by other readers – either someone Boole would have asked to read the text (for instance his wife, Mary Everest Boole), or readers trying to make sense of the manuscript after Boole's death.

[15] The handwriting is compatible with Boole's, and the style of numbering – in the top right corner, with a rough quarter of a circle separating the number from the text – is similar to that found in other drafts by Boole, for instance in the last manuscript for *LT* (Royal Society Archives, collection MS/782, signature U).

[16] Royal Society Archives, collection MS/782, notebook R2 read from the back, on which see *SMLP,* 203–204.

[17] For instance, this is how the final manuscript for *LT* seems to have been produced (Royal Society Archives, collection MS/782, signature U); see also similar phenomena in the final draft of *Boole 1862* (Royal Society Archives, PT/64/10).

3.2 Context and dating

The dating of the manuscript relies on two lines of evidence: references, in the manuscript itself, to earlier publications by Boole; and mentions, in Boole's correspondence with Jevons, of a project similar to 'On the Nature of Thought'.

First, references to other works by Boole allow dating NT to between 1862 and Boole's death in late 1864. Most obviously, there are multiple explicit references to *LT* (1854) in the manuscript. In the first paragraph, Boole also mentions his recent work on the theory of probability, 'the analytical characters of which together with their consequences have lately been discussed by me in the Transactions of this Society.' At first sight, this sentence could refer to either of two papers on probability that Boole published in the *Transactions* of a Royal Society:[18] 'On the Application of the Theory of Probabilities to the Question of the Combination of Testimonies or Judgements', published in the *Transactions of the Royal Society of Edinburgh* (*Boole 1857*); and 'On the Theory of Probabilities', published in the *Philosophical Transactions of the Royal Society* (*Boole 1862*). The former applies the theory of probability from *LT* to specific problems, while the latter is concerned with the analytical foundations of the theory. Boole's mention, in our manuscript, of the 'analytical characters' of the theory of probability thus points to the 1862 paper (read at the Royal Society of London on 19 June 1862). Moreover, it is much likelier, on independent grounds, that Boole intended NT for the *Philosophical Transactions*, and thus that 'this Society' refers to the Royal Society of London: the *Philosophical Transactions* became Boole's preferred venue in the last years of his life, after his election as a Fellow in 1857, whereas he never published again in the *Transactions of the Royal Society of Edinburgh* – a journal he seems to have chosen in 1857 largely in order to be considered for that society's Keith Prize, which he won (*MacHale 2014,* 253). One further clue, though not conclusive by itself, is that Boole brought up his 1862 paper, this time unambiguously, in his first reply to Jevons from August 1863 (*Grattan-Guinness 1991,* 27):

> Your mention of my method for the solution of questions of Probability which occupies the latest portion of the work on the Laws of Thought encourages me to send you a paper of mine from the Philosophical Transactions which has a very important bearing upon the theory of the method.

This suggests that Boole saw his 1862 paper as an important piece of context at the time, and hence makes it more plausible that he would refer to it at the beginning of a new work. All of this points to NT having been composed after the summer of 1862.

The second line of evidence comes from the fact that Boole discusses manuscripts similar to 'On the Nature of Thought' in his correspondence with Jevons. In his first letter to Jevons, dated 17 August 1863, he wrote (*Grattan-Guinness 1991,* 26, Boole's emphasis):

> It is certainly possible to work out logical problems according to the general laws developed in my work, *with such added restrictions* as shall make all intermediate results interpretable. I have somewhere laid by, a paper on this

[18] See the full bibliography compiled by *MacHale 2014,* 315–320.

> subject which I wrote about two years ago. I cannot at this moment put my hand upon it, but I remember that it involved the application of such instructions as should make the elementary operation always interpretable in ordinary language. Thus $x - y = 0$ would become $x - xy = 0$, and so on. But I did this, not because I had any doubt of the validity of the processes of my work which are unrestricted by any such conditions, but in order to determine, for my then satisfaction, and prospectively with a view to publication, how far my system could be made intelligible to those who knew nothing of mathematics.

This describes 'On the Nature of Thought' (and especially the main technical part of it) quite well – better than it does any other manuscript I am aware of in the Boole *Nachlass*. Compare, for instance, the first sentence quoted above to the ending of NT ([49]–[50]):

> These propositions enable us to accomplish every object which lies within the scope of the formal Logic of class. And they enable us to do this without transgressing the conditions of interpretability.
>
> But it is seen that the freedom of our procedure is restrained by these conditions. [. . .] And the entire procedure of thought as manifested in the foregoing propositions is one in which while the result of each step is determined by formal laws[,] the *order* of the steps is restricted by the condition that each result as it arises shall be interpretable into actual representative thought.

The preceding quotes make it clear that NT belongs to the broad project Boole is gesturing at in his letter to Jevons. The chronology, however, does not quite fit. Here Boole mentions a manuscript from 'about two years ago', thus from around the summer of 1861, whereas, as I have argued, his reference to a 1862 paper at the beginning to NT point to it having been written no earlier than the summer of 1862. Perhaps Boole is misremembering and does refer to NT. His later letters, though, suggest a different hypothesis: 'On the Nature of Thought' is most likely a *reworked version* of the earlier material mentioned here, together with one or more *other* drafts.

Already in his letter to Jevons from August 1863, Boole had mentioned his intention 'of again publishing on the subject of Logic' (*Grattan-Guinness 1991,* 27). When Jevons, in a letter dated 5 November 1863, sent him the (as yet uncorrected) proofs of the critical part of his forthcoming book, *Pure Logic* (*Jevons 1864*), Boole refused to read them before publishing his own views (*Grattan-Guinness 1991,* 32):

> I received your letter with the accompanying proof sheets of your work on Logic [. . .]. But I have after careful consideration decided [. . .] not at present to read the sheets. And I have this day delivered them unopened into the hands of Dr. Ryall the Vice-President of the Queen's Coll. Cork [. . .].
>
> My reasons for doing this are 1st that I am at present working at another subject, in fact preparing for the press a second edition of my work on Diff^{l} Equations. [. . .] Secondly I have a large quantity of MS. unpublished on the subject of Logic in particular two papers one written a good many years ago and both bearing

on the question of the philosophy of the method in my published work. I am inclined to think that I could in a short time compile from these what would suffice at least to put my general views on the subject before the public and at present I feel disposed to do this – but I could not do this if I first engaged in the discussion of the points of difference between us.

Boole's efforts to secure a witness for the fact that he did not read Jevons's proofs suggest that he meant to stake a claim to priority by publishing his views as soon as possible – at any rate, before Jevons's book came out. If his final letter to Jevons, from about three months later, is to be believed, he indeed got down to work – though, again because of his book on differential equations, he eventually gave up: 'I threw aside my own unpublished papers on Logic *of which I commenced a month or two back to write an account*' (letter dated 30 January 1864, *Grattan-Guinness 1991,* 32, my emphasis).

It seems plausible that 'On the Nature of Thought' is the result of Boole's hasty attempt, in the last months of 1863, to pre-empt Jevons's criticisms. First, Boole's prestigious choice of venue is compatible with a priority claim. As argued above, the first paragraph of NT makes it clear that it was intended for the Royal Society of London. As a Fellow, Boole had easy access to publication in one of the Society's journals (the *Philosophical Transactions* or the *Proceedings*), and even though their reviewing procedures could make them slower than other venues, papers published there carefully indicated the date they were received;[19] moreover, Boole's past experience suggested that his paper would very swiftly be read at the Society and reviewed.[20] Second and most importantly, notice that in November 1863, Boole wrote not of *one*, as in his earlier letter, but of *two* main manuscripts on logic that he was thinking of compiling. This fits quite well with the composition of 'On the Nature of Thought'. As mentioned already (see especially footnote 14, p. 11), while its specific goals and technical details are distinctive within Boole's *Nachlass*, the philosophical introduction of NT has extensive parallels with other manuscripts: in particular, a number of passages are identical word for word with a fairly polished draft, possibly dating from around 1860, that has long been assumed to be the first chapter of a projected new book on logic. This points to Boole's composing 'On the Nature of Thought' by combining a manuscript specifically devoted to matters of interpretability with introductory philosophical considerations adapted from an earlier draft.

[19] On the Royal Society's reviewing and publication procedures at the time, see *McDougall-Waters and Fyfe 2022;* papers not accepted in the *Transactions*, or under consideration for an extended period of time, would typically have an Abstract (with date of reception) printed in the *Proceedings*, all but guaranteeing some publication with a certified date.

[20] Boole's earlier paper on probability, mentioned above (*Boole 1862*), was received and read on 19 June 1862 and reviewed within a week (the reports, by William Donkin and Arthur Cayley, dated 23 and 25 June, are preserved in the Royal Society Archives, RR/4/17 and RR/4/16); Boole knew before mid-July that it had been accepted for publication in the *Transactions* (*MacHale and Cohen 2018,* 149–150).

4 Boole's logical calculus and the meanings of 'interpretability'

Before we can turn to the technical details of 'On the Nature of Thought', we need to get clear on the slippery issue of what exactly 'interpretability' meant for Boole. It turns out that, especially in the *Investigation of the Laws of Thought*, he used the term in multiple different senses. In particular, there is a tension between a *compositional* notion of interpretation – one that appears natural to modern readers, and that Boole would foreground in 'On the Nature of Thought' – and an indirect, normal-form-based notion which, as we shall see, was central in *LT*. I have analyzed this confusing landscape, and traced its roots in the legacy of algebraic analysis, the approach to the calculus that Boole practised, at length elsewhere (*Waszek 2025*); the goal of this section is to provide just enough for understanding 'On the Nature of Thought' and its relation to *LT*.

First of all, one should realize that Boole did not use 'interpretation' and its cognates in a systematic, well-delineated *technical* sense. Broadly speaking, he tended to use the term whenever giving some kind of (non-symbolic) meaning to a string of symbols. For instance, in his logical problem-solving method, he offered a general way of 'interpreting' *any equation* – which amounted to reducing it to other equations and eventually *translating it back into sentences of ordinary language*, in the terms of the original problem.

The ambiguity that matters for understanding 'On the Nature of Thought', however, is more specific: it is about what Boole meant by calling a symbolic *expression* (i.e., what in first-order logic would be called a *term*) 'interpretable'. In line with modern readers' expectations, for him 'interpreting' a symbolic expression usually amounted to assigning a semantic value to it (in logic, this value would be a 'class', that is, a set). But in contrast to the notion of interpretation that is standard for terms in first-order logic today, this semantic assignment did not need to be compositional. Moreover, Boole sometimes called an expression 'interpretable' in yet another sense, to mean that it could be assigned a semantic value for *any assignment* of values to the individual variables occurring in it. Accordingly, the purpose of this section is to distinguish three senses of the word 'interpretable' as Boole applied it to symbolic expressions, which I shall call the *compositional*, the *global*, and the *formal* senses.

As a preliminary, here are some pieces of Boole's terminology. Boole's symbolic calculus from *LT* is made up of 'literal symbols', that is, letters such as x, y, z, etc., and of 'signs of operation', including $+$, $-$, and $\times$ (the latter often omitted, as usually done in algebra). These signs and symbols are subject to 'laws of combination', which are general equations including, for instance, $xy = yx$, $x^2 = x$, and $z(x + y) = zx + zy$.[21] The resulting calculus can be 'interpreted' in different domains. In one interpretation, letters stand for either the number 0 or the number 1, and the signs of operation for the usual operations on numbers. In another – the properly *logical* interpretation – letters stand for classes, and signs of operation for operations on classes: xy is the class of elements common to classes x and y – what we would call their intersection; $x + y$ is the class of elements that are either in x or in y, but

[21] Note that the letters occurring in such 'laws' are best described, in today's language, as metavariables for any 'literal symbol', but not for any expression. Importantly, as we shall see below, the 'law' $x^2 = x$ holds for any literal symbol instead of x, but not for any complex expression.

is only defined if x and y are disjoint; $x - y$ is the class of elements that are in x and not in y, but *is only defined if y is included in x*. For simplicity, I shall only discuss this logical interpretation from now on.

4.1 Compositional versus global interpretability

The first sense in which an expression can be 'interpretable' is straightforward. It is relative to an assignment of classes to the individual letters that occur in the expression (or to hypotheses that constrain possible assignments). An expression is interpretable in this first sense – is *compositionally* interpretable, as I shall say to disambiguate – if it can be assigned a class from the bottom up[22] according to Boole's definitions, given above, of the class operations he denotes by $\times$, $+$, and $-$. An expression will fail to be interpretable in the compositional sense if, when attempting to assign a class to it step by step, one hits upon an operation that cannot be performed, e.g., an addition of non-disjoint classes.

It is the compositional sense of interpretation that Boole has in mind when he writes that 'the expression $x + y$ seems [...] uninterpretable, unless it be assumed that the things represented by x and the things represented by y are entirely separate' (*LT,* 67); or the sense in which, discussing the equation

$$x + y - z = x - z + y,$$

he explains that for 'the forms themselves [to] become interpretable', one needs the two conditions that 'x and y have no members in common' and that 'all the members of z are contained in x' (NT, [22]).[23] Boole does not define compositional interpretability in so many words in *LT,* but is more explicit in 'On the Nature of Thought', at one point talking of expressions 'formed of elementary conceptions $x, y, z, \ldots$ by processes which are possible' (NT, [34]).

Because of Boole's conception of algebra, however, compositional interpretability is not the end of the story (at least before NT). As discussed in Section 2.2, he thought that one could continue to reason formally according to such laws as $x + y - z = x - z + y$ even in cases in which the expressions involved were no longer (compositionally) interpretable. But this opens the door to *indirectly* assigning classes to expressions. Suppose that x and y are disjoint, and that z is included in $x + y$, but not in x, so that $x + y - z$ (i.e., $(x + y) - z$) is compositionally interpretable, but $x - z + y$ (i.e., $(x - z) + y$) is not. The latter expression can then be assigned a class indirectly by first transforming it into the former. In fact, Boole has a systematic method for performing such indirect assignments, via a reduction of any expression to a *normal form* from which one can either assign a class to it or conclude that no such assignment is possible. In contrast to compositional interpretability, the notion of interpretability that results – which I shall call *global* interpretability – is a property of

[22]For compositional interpretability to be well defined in general, even – as is usual in Boole – when expressions are not fully parenthesized, one needs to specify the precedence of operations. It is clear from Boole's writings that, following standard practice in algebra, he takes multiplication to have precedence over addition and subtraction, and that he treats the latter two as left-associative.

[23]As Boole reads the equality as $(x + y) - z = (x - z) + y$, he does not require z to be included in y.

expressions that is invariant under all algebraic transformations licensed by Boole's 'laws'. An expression that is *globally* interpretable may also be compositionally interpretable, but need not be.

Boole's reduction of expressions to a normal form relies on his *theorem of development*. To introduce it, let us focus on the particular case of expressions V made up of two literal symbols x and y, which, following Boole, we can write $V = f(x, y)$ – a notation indicating that V is a 'function' of x and y, that is, an algebraic expression[24] composed of x and y (and perhaps of other literal symbols, though we shall assume that such is not the case here). Relative to two classes x and y, the universe of discourse can always be partitioned into the four classes xy, $x\bar{y}$, $\bar{x}y$, and $\bar{x}\bar{y}$ (where $\bar{a} = 1 - a$ is the complement of a), which Boole calls 'constituents': the classes made up of, respectively, the elements that are both in x and in y; those that are in x but not in y; those that are in y but not in x; and those that are in neither. Relative to n literal symbols, the universe of discourse can be partitioned into 2^n 'constituents' in the same way (incidentally, it was in order to clarify such partitions that John Venn, while preparing a textbook on Boole's logic, introduced the diagrams that now bear his name). Now, Boole claimed that any function of x and y could be expressed – in his words, 'expanded' or 'developed' – in terms of these four 'constituents':

$$f(x, y) = f(1, 1)\, xy + f(1, 0)\, x\bar{y} + f(0, 1)\, \bar{x}y + f(0, 0)\, \bar{x}\bar{y}.$$

The 'coefficients', such as $f(0, 1)$, that appear in this development are obtained by substituting 0 or 1 to each of x and y in the function f. Notice that the value of these coefficients will be invariant under algebraic transformations, and so will the development as a whole.

As a first example, consider the expression $f(x, y) = (x + y) - xy - xy$, or $(x + y) - 2xy$, which is not compositionally interpretable in general (though it will be if x and y stand for disjoint classes). Its development is $x\bar{y} + \bar{x}y$,[25] which is *always* compositionally interpretable and stands for the symmetric difference of x and y. By contrast, consider what happens to the expression $f(x, y) = x + y$. It gets expanded as

$$x + y = 2xy + 1x\bar{y} + 1\bar{x}y + 0\bar{x}\bar{y} = x\bar{y} + 2xy + \bar{x}y$$

where the intersection of x and y is, so to speak, counted twice. Since $xy + xy = 2xy$ cannot be assigned a class, the expression $x + y$ is not interpretable, *unless* xy vanishes – that is, unless the classes x and y are disjoint.

In general, the development of an expression shows whether an expression is globally interpretable, and if so, how. This goes as follows. If all 'constituents' with a coefficient other than 0 or 1 vanish, as is the case in the development of $x + y$ whenever we know that $xy = 0$, then the development exhibits the expression as a (compositionally interpretable) sum of disjoint constituents, which straightforwardly allows assigning a class to it. Otherwise, the expression cannot be assigned a class in general, although the development brings out the *conditions* that would be required for it to be globally interpretable.

[24]Boole used 'function' in the traditional sense of algebraic expression, not in the modern extensional sense of an arbitrary input-output relation.

[25]Indeed, we have $f(1, 1) = 1 + 1 - 2 = 0$; $f(1, 0) = 1 + 0 - 0 = 1$; $f(0, 1) = 0 + 1 - 0 = 1$; and $f(0, 0) = 0 + 0 - 0 = 0$, so that $f(x, y) = 0xy + 1x\bar{y} + 1\bar{x}y + 0\bar{x}\bar{y} = x\bar{y} + \bar{x}y$.

4.2 Formal interpretability

The most important sense in which Boole called expressions 'interpretable' in *LT* is yet a different one. When referring to it, he sometimes, but by no means consistently, clarified his meaning by adding adverbs, writing of 'independently' or 'generally' interpretable expressions; for my part, following *Brown 2009,* I shall talk of *formal* interpretability. As Boole put it in *LT,* 92*:*

> By an independently interpretable logical function, I mean one which is interpretable, without presupposing any relation among the things represented by the symbols which it involves.

In practice, Boole treated an expression V as formally ('independently') interpretable if it was *globally* interpretable *for any assignment* of classes to the variables occurring in it. It is clear that an expression V is formally interpretable in this sense if and only if *all* coefficients in its development are 0 or 1, so that no constituent needs to vanish for the development to allow assigning a class to the expression. Boole identified an expression V being formally (i.e., in his terms, 'independently' or 'generally') interpretable with its satisfying what he called the 'law of duality', that is, the equation $V^2 = V$ or $V(1 - V) = 0$ – or more precisely, with its *formally* satisfying it, that is, satisfying it without bringing in any assumed relation between the variables. He justified the equivalence by reference to the developments of both sides: replacing V by its development in $V^2 = V$, the law of duality will hold formally if and only if all coefficients a in the development of V satisfy $a^2 = a$, i.e., assuming no fractions are involved, if all of them are either 0 or 1; a condition which, as just explained, is equivalent to formal interpretability. As Boole put it (*LT,* 93):

> The condition $V(1 - V) = 0$ may be termed "the condition of interpretability of logical functions."

4.3 Discussion: interpretability in the *Laws of Thought*

Notice that the distinction between compositional and merely global interpretability is orthogonal to the distinction between formal interpretability and interpretability only under conditions. A formally interpretable expression is one that can be assigned a class for any assignment of classes to the variables it contains, but does not have to be compositionally interpretable – the assignment of a class to it may only be possible globally. Conversely, an expression that is not formally interpretable can nevertheless, under specific conditions, be compositionally (or globally but not compositionally) interpretable. Examples of all of these situations are provided in Table 1.

In sum, from the point of view of *LT,* we end up with the following landscape. Relative to a specific assignment of classes to the individual letters (i.e., class variables) that occur in it, an expression can be *compositionally* interpretable or *globally* but not compositionally interpretable (or, of course, not interpretable at all). Moreover, an expression can also be called interpretable in a *formal* sense, namely if it is *globally* interpretable without conditions, that is, for any assignment of classes to the letters that occur in it. One complication is that

	Formally interpretable	Interpretable only under conditions
Compositionally interpretable	$x + (1 - x)y$	$x + y$ *(requires $xy = 0$)*
Only globally interpretable	$x + y - 2xy$ $(= x\bar{y} + \bar{x}y)$	$2x - (x + y)$ $(= x - y$, *requires* $\bar{x}y = 0)$

Table 1: Examples of four ways for an expression to be interpretable, from *Waszek 2025*

Boole's method is algebraic, or *analytic* in the traditional sense of the term: one reasons about class variables as about unknowns that satisfy some equations, without having extensional knowledge of the corresponding classes. In such contexts, where no explicit assignment of classes to variables is provided, it might seem that the notions of compositional and global interpretability are not applicable (though formal interpretability always is). In practice, while formal interpretability does tend to take center stage in *LT,* an expression can nevertheless be called interpretable in the compositional or global senses as long as the assumptions one has made guarantee that the conditions required for compositional or global interpretability are satisfied.

4.4 From the *Laws of Thought* to 'On the Nature of Thought'

In 'On the Nature of Thought', Boole aims to avoid any reliance on formal reasoning beyond conditions of interpretation. As a consequence, he aims to avoid any expression that is globally but not compositionally interpretable.

This shift in point of view may be observed in Boole's treatment of equality. In *LT,* he initially introduced 'the symbol =' as expressing '*is* or *are*' (*LT,* 35), but in his practice, the sign = could not be systematically defined by the fact that the expressions on either side of it had the same value: this would only make sense in interpretable cases. Instead, as was usual in algebraic analysis (*Ferraro and Panza 2003,* section 3), an equality just meant that the expression on the left-hand side of = could be transformed into the one on the right-hand side. In NT, on the other hand, Boole wrote (NT, [17]):

> The sign = will be used to denote identity ie to denote that any two expressions between which it is placed represent classes of things which are identical as respects the individuals of which they consist[,] although these individuals are in the two expressions presented under different aspects of thought.

In other words, he moved to an explicit, fully extensional definition of equality.

The shift to an extensional perspective leaves only two notions of interpretability at play. All expressions ought to be *compositionally* interpretable (relative to a specific assignment of classes to variables, or to assumptions that constrain such assignments), corresponding to the top line of Table 1. They can additionally be *formally* (in NT, Boole says 'generally') interpretable, corresponding to the left-hand column of the table. Since Boole restricts himself to compositionally interpretable expressions, the relevant notion of formal interpretability is restricted to the top left corner of the table: it can be defined as compositional (and not, as before, global) interpretability for any assignment of classes to variables. As Boole puts it in the context of equations (NT, [34]):

> By a generally interpretable equation is meant one in which the members connected by the sign of equality = express general conceptions being formed of elementary conceptions x, y, z, ... by processes which are possible independently of the particular interpretation of x, y, z.

In practice, Boole in NT rewrites his problem-solving methods so that all expressions that occur in them 'express general conceptions' in the sense of this quote: they are both formally and compositionally interpretable, corresponding to the top left corner of Table 1.

5 Eliminating uninterpretable steps: the technical details

At the heart of Boole's *Investigation of the Laws of Thought* are a symbolic calculus for logic and a *general method*, using this calculus, for solving logical problems. At first sight, the technical parts of 'On the Nature of Thought' seem to offer no more than a new exposition of these two elements, and Boole does not always spell out the differences between what he is doing and his earlier work. However, while the calculus itself remains unchanged, the problem-solving method is altered systematically with a specific goal in mind, namely that of avoiding uninterpretable intermediate steps – steps which his conception of formal reasoning had previously licensed.

To bring out what 'On the Nature of Thought' achieves, I offer a reconstruction of Boole's main problem-solving method in *LT,* breaking it down into six steps and highlighting where issues with interpretability arise (Section 5.1). I then compare this reconstruction, first, to the amendments Boole suggested in the little-discussed Chapter X of *LT,* which – though it is not framed in that way – takes some steps toward solving the interpretability issues in question (Section 5.2); and, second, to 'On the Nature of Thought' itself (Section 5.3). This should elucidate where the problems lay that Boole intended to solve, and make it clear that he systematically, and largely successfully, addressed them.

5.1 Boole's method in *LT:* A blueprint

Boole's logic aimed at offering a general method to solve *logical problems*. What logical problems are is best introduced by an example (for a more thorough discussion, see for instance *Waszek and Schlimm 2021,* section 1). In *LT,* 134–137, Boole reconstructed a piece of reasoning from Aristotle's *Nicomachean Ethics*. In Boole's telling, Aristotle's starting point are various premises, such as 'Virtue is something according to which we are praised or blamed, and which is accompanied by deliberate preference', which can be phrased as equations connecting the classes of virtues (v), passions (p), faculties (f), habits (h), things accompanied by deliberate preference (d), etc. Aristotle's reasoning as Boole conceives it is a goal-directed process that aims at determining whether virtues are passions, faculties, or habits, a goal Boole expresses in equational terms: one seeks the strongest possible conclusion of the form $v = \ldots$, where the right-hand side is some expression containing p, f, and h, *but not* any of the other classes (such as 'things accompanied by deliberate preference'), which we treat as auxiliary terms to be eliminated.

In abstract terms, the general problem that Boole solves is as follows. Given equations connecting a set of class terms, find the strongest possible conclusion that involves a prescribed *subset* of these class terms ('eliminating' the others) and that is expressed in a prescribed form – usually $x = \ldots$, where x is a particular class term and the right-hand side does not contain x. Moreover, Boole's method typically yields an additional equation – he calls it an 'independent relation' – which expresses everything that follows from the premises concerning the non-eliminated class terms but cannot be fit into the prescribed form.

That being said, here is a sketch of Boole's general solution method in *LT,* roughly as it is presented up to and excluding Chapter X of that book (which will be discussed in the next section). For future reference, I have broken down the method into six successive steps and have introduced labels for them (which are not Boole's own, though they are inspired by terminology that he actually uses). My reconstruction is largely in line with that of *Brown 2009* and I have followed his terminology whenever possible, though my purposes require distinguishing more steps than he does. Note that the blueprint I offer is a simplification in several ways. First, I have left aside any consideration of Boole's 'indefinite' class symbol v, since NT does not offer much clarification on this particular topic and does not discuss particular propositions at all.[26] Second, I have omitted numerous variations in Boole's methods (in many cases, he presents multiple ways of accomplishing each step, and his steps can often be switched around to simplify computations – see especially Chapter IX of *LT*). My point is merely to give an overview, and most importantly to highlight where interpretability issues can arise: after each step, I discuss whether and how it can lead to uninterpretable expressions, be it in the compositional or in the formal sense. As a preliminary to the method proper, I have inserted Boole's process of development as Step 0, not because it comes first in actual problem solving but because it is an auxiliary tool Boole invokes repeatedly.

0. (*Development*) Any expression V can be 'developed' in terms of any literal symbol x as $V = Ax + B\bar{x}$, where A and B are expressions that do not contain x.

As mentioned above, to compute the development of an expression V with respect to x, Boole would write $V = f(x)$ (indicating that V is a 'function' of x, that is, an algebraic expression composed of x, possibly among other symbols) and write

$$V = f(1)x + f(0)\bar{x}, \qquad (*)$$

where $f(0)$ (resp. $f(1)$) is the expression obtained by replacing all occurrences of x in $V = f(x)$ by 0 (resp. 1). We have already discussed the fact that Boole's *justification* of

[26] NT is more explicit than *LT* on the fact that the meaning of his v will have to be context-dependent, sometimes referring to an 'absolutely indefinite' class that could be empty or equal to the universe, sometimes referring to a class that 'must be supposed to include at least one individual' (NT, [19]). For discussion of Boole's v, see *Hailperin 1986,* 152–155 and *Makinson 2022.*

the identity (∗) in *LT* presupposes the free use of formal reasoning (in that this licenses the transfer of results from the arithmetical case to the logical case; see footnote 10 p. 9 above). But what about the identity itself, disregarding its justification – is the right-hand side of (∗) always interpretable if V is? If V is formally interpretable, then $f(0)$, $f(1)$, and the whole right-hand side of (∗) will trivially be as well. On the other hand, Boole's development does not preserve mere compositional interpretability: if V is compositionally but *not* formally interpretable, the result of development might be *neither*. For instance, suppose $V = x + y$, assuming $xy = 0$. Then $f(1) = 1 + y$, which is not compositionally interpretable unless $y = 0$. (If the development is performed with respect to *all* literal symbols appearing in V, then this problem does not arise and mere compositional interpretability is preserved.)

1. (*Expression*) Express the premises by symbolical equations.

By construction, both sides of the equations expressing the premises will be (compositionally) interpretable, but note a subtlety: the expressions involved might not be *formally* interpretable, unless particular care is taken to ensure this. For instance, one premise could assert that x and y form disjoint classes (which could be expressed, say, as $xy = 0$). Another premise might assert that everything in the universe is either in class x or in class y. Since we know that x and y are disjoint, it is in principle allowable to express this second premise as $x + y = 1$. However, the expression $x + y$ is not formally interpretable; it can only be assigned a class under the assumption that the other premise is true. Boole always encourages the use of formally interpretable expressions when translating premises, but does not mandate it.

2. (*Transposition*) For each premise, bring the equation corresponding to it into the form $V = 0$.

The straightforward way of doing this in *LT* is by subtraction: transform an equation $S = T$ into $S - T = 0$. The resulting left-hand side $S - T$ will not be formally interpretable in general, even though it will be compositionally interpretable under the assumption that $S = T$. For instance, $x = y$ leads to $x - y = 0$, where $x - y$ is not formally interpretable.

3. (*Reduction*) Bring all equations $V_1 = 0, V_2 = 0, \dots, V_n = 0$ together into a single one of the form $V = 0$.

Boole's standard methods for doing this lead to V being neither compositionally nor formally interpretable in general, even if all the V_i are. His simplest method involves squaring the equations, then adding them together, leading to the single equation $V_1^2 + V_2^2 + \dots + V_n^2 = 0$. (If all V_i's are formally interpretable, one can skip the squaring step, as $V_i^2 = V_i$.) The reason the left-hand side will not be interpretable in general is that there could be terms in common in the developments of the V_i's. For instance, starting from the innocuous-looking premises $xz = 0$ ('x and z are disjoint') and $x\bar{y} = 0$ ('x is included in y'), one obtains $xz + x\bar{y} = 0$, which is neither compositionally nor formally interpretable unless one makes the additional hypothesis that $x\bar{y}z = 0$ (i.e., that there are no elements that are at the same time in x, in z, and not in y).

4. (*Elimination*) Eliminate any unwanted 'literal symbols' (class variables) from $V = 0$, obtaining a new equation $V' = 0$ in which the unwanted symbols no longer appear.

Elimination as performed in *LT* will preserve formal interpretability, but in general not mere compositional interpretability. Roughly, the method goes as follows. To eliminate s from V, Boole would write $V = f(s)$ and obtain V' as $f(0)f(1)$. If V is formally interpretable, then $f(0)$, $f(1)$, and $f(0)f(1)$ will be too; if in addition V is compositionally interpretable, all of these will be as well. On the other hand, elimination (much like development) does not preserve mere compositional interpretability: if V is compositionally but *not* formally interpretable, the result of elimination might be neither. Take $V = x + y$, assuming $xy = 0$. Then eliminating x by Boole's method yields $V' = (1 + y)y = 2y$, which is neither compositionally nor formally interpretable unless $y = 0$.

Moreover, note that the way Boole *justified* the fact that the result of eliminating s from $f(s)$ is $f(0)f(1)$ heavily relied on formal reasoning, including the use of division (*LT*, 101–103).

5. (*Solution*) Solve for x: develop $V' = 0$ as $Ex + E'\bar{x} = 0$, then write $x = \frac{E'}{E'-E}$ and develop the right-hand side with respect to all remaining class symbols. In the simplest case (if V', and hence E and E', are formally interpretable), this will yield an expression of the form

$$x = A + B\frac{0}{0} + C\frac{1}{0}.$$

Boole interprets $\frac{0}{0}$ as representing an indefinite class (i.e., any class whatsoever, including the empty class or the universe), for which he also sometimes used the symbol v, and $\frac{1}{0}$ as representing an impossible class, so that the full equation is only possible if $C = 0$. His full solution, then, is that

- $x = A + vB$, i.e., x is equal to A together with an indefinite portion of B;
- there is an 'independent relation' $C = 0$.

If V' is not formally interpretable, non-zero coefficients other than 1, $\frac{0}{0}$ and $\frac{1}{0}$ can pop up; these will then be treated just like $\frac{1}{0}$, so that the corresponding constituents can just be added to C.

This is the step in Boole's method that has seemed most mysterious to readers. One can distinguish three different problems. The first is that the expression $E' - E$ is neither compositionally nor formally interpretable in general. The second, which readers often found most glaring, is the use of division itself. As Boole did not give an explicit meaning to division in *LT*, it appears that, from the perspective of that book, the expression $\frac{E'}{E'-E}$ is not compositionally interpretable. Moreover, there is no way to see it as globally or formally interpretable either: in general, coefficients that do not represent classes, such as $\frac{1}{0}$, will appear in its development. The third problem is Boole's way of 'interpreting' the coefficients $\frac{0}{0}$ and $\frac{1}{0}$, which in *LT* appears decidedly ad hoc.

5.2 Chapter X of *LT*

A first step in the direction taken in NT appears in Chapter X of *LT*, which is rarely discussed by commentators (though see *Brown 2009*). In this chapter, entitled 'On the conditions of a perfect method', Boole writes (*LT*, 151):

> To make the one fundamental condition expressed by the equation
>
> $$x(1-x)=0,$$
>
> the universal type of form, would give a unity of character to both processes and results, which would not else be attainable.

In practice, he attempts to alter some steps in his methods so that both sides of all equations satisfy the 'law of duality' $V(1-V)=0$.

As we have seen, Boole called the equation $V(1-V)=0$ 'the condition of interpretability of logical functions' (*LT*, 93) – it characterizes what I called formal interpretability, that is, global interpretability for any assignment of classes to variables. Nevertheless, Boole did not describe his goal in Chapter X as that of ensuring the interpretability of intermediate expressions. This is less surprising once one realizes that in the context of *LT*, guaranteeing $V(1-V)=0$ did not and could not have *foundational* significance. Boole did not doubt the legitimacy of formal reasoning beyond interpretable cases, and, in contrast to 'On the Nature of Thought', was not trying to show that his methods did not depend on it. Moreover, the equation $V(1-V)=0$ can only track *global* interpretability: even if all expressions satisfy it, hence can be assigned classes, they may in principle still be uninterpretable *compositionally* – in which case formal reasoning would still be needed to make sense of them. Boole's set-up in Chapter X does not allow addressing compositional interpretability as a separate notion.

In terms of the blueprint above, Boole's amendments in Chapter X mainly concern Steps 2 (transposition) and 3 (reduction to a single equation); they also touch upon Step 1 (symbolic expression of the premises). While Boole also discusses Steps 4 (elimination) and 5 (solution), I shall leave these aside because he does not modify them to ensure formal interpretability, but only examines the simplifications that result if everything in their *starting point* is formally interpretable.

2. (*Transposition*) For each premise, bring the equation corresponding to it into the form $V=0$.

Referring to earlier chapters, Boole claimed that the equation $X=Y$, where X and Y satisfy the law of duality, could be replaced by $X(1-Y)+Y(1-X)=0$, whose left-hand side also satisfies it (*LT*, 152). Though he does not state the matter in such terms, the left-hand side of the latter equation is not only formally, but also compositionally interpretable if X and Y are. However, Boole only *asserted* the equivalence of $X=Y$ and $X(1-Y)+Y(1-X)=0$, without offering a careful proof that would not require detours through equations such as

$X - Y = 0$ (whose left-hand side is not formally interpretable in general).[27]

3. (*Reduction*) Bring all equations $V_1 = 0, V_2 = 0, \ldots, V_n = 0$ together into a single one of the form $V = 0$.

This is the step of most concern to Boole in Chapter X. Instead of just adding the equations (or their squares), Boole derives a single equation of the form

$$V_1 + (1 - V_1)V_2 + (1 - V_1)(1 - V_2)V_3 + \ldots = 0.$$

Notice that this equation, in contrast to $V_1^2 + V_2^2 + \ldots = 0$, is always formally *and* compositionally interpretable provided all the V_i's are. Boole's *justification* of it, though, requires applying his process of development to the various equations involved (*LT,* 153): thus, it presupposes formal reasoning.

After discussing Step 3, Boole goes back to Step 1:

1. (*Expression*) Express the premises by symbolical equations.

Boole's discussion of this step contains two remarks that bear upon formal interpretability. First, he noted that the equations obtained from the premises can contain expressions that are not formally interpretable 'if they involve material (as distinguished from formal) relations, which are not expressed' (*LT,* 153) – e.g., if we presuppose the premise $xy = 0$ when translating another premise as $x + y = 1$ (see the initial discussion of Step 1 above). Boole certainly encouraged translating premises in such a way that no such problem would arise; in fact, he had claimed earlier in the book (*LT,* 65) that a proposition such as $X = Y$,

> if founded upon a sufficiently careful analysis of the meaning of the "terms" of the proposition, will satisfy the fundamental law of duality which requires that we have
>
> $$X^2 = X \quad \text{or} \quad X(1 - X) = 0,$$
> $$Y^2 = Y \quad \text{or} \quad Y(1 - Y) = 0.$$

Second, Boole remarked that there are cases in which his amendment to Step 3 is not required in order to guarantee a formally interpretable result: namely when the premises are 'independent', that is, 'when it is not possible to deduce from any portion of the system any conclusion deducible from any other portion of it' (*LT,* 154). Concretely, independence means that the developments of all the equations (once put into the form $V = 0$ by Step 2) will never have constituents in common, and so can be added without redundancy – their simple sum will be formally interpretable. Boole noted that one could ensure independence by developing equations right from the start and tossing out any redundant constituents.

[27] When asserting that $X = Y$ is equivalent $X(1-Y)+Y(1-X) = 0$ in Chapter X, Boole pointed to Chapter VIII, in which the fact that the first implies the second is proved by two methods. In the first, $X = Y$ is transformed into $X - Y = 0$ and then into $X(1 - Y) + Y(1 - X) = 0$ by squaring (*LT,* 123). In the second, $X = Y$ is decomposed into $X = vY$ and $Y = vX$; v is 'eliminated' from each, yielding $X(1 - Y) = 0$ and $Y(1 - X) = 0$; and then the latter equations are added (*LT,* 124). But the elimination of v, as presented by Boole earlier in the book, itself requires going through such steps as $X - vY = 0$ (*LT,* 104). In both cases, the implication is justified via a formally uninterpretable subtraction.

5.3 Boole's amended methods in 'On the Nature of Thought'

For the purposes of 'On the Nature of Thought', the methods of Chapter X of *LT* are incomplete and unsatisfactory. They are incomplete because they do not address the whole range of interpretability issues that arise in the methods of *LT*. They are unsatisfactory because, on the face of it, they are only about ensuring *formal*, and not *compositional* interpretability; and also, importantly, because Boole did not justify them in a way that did not *itself* require formal reasoning. While it shows some continuity with Chapter X, 'On the Nature of Thought' systematically tackles all of these issues.

0. (*Development*) Any expression U can be 'developed' in terms of any literal symbol x as $U = Bx + C\bar{x}$, where B and C are expressions that do not contain x.

Boole offers an elementary proof of his theorem of development that does not depend on formal reasoning (NT, [29]–[33]). In contrast to the proofs in *LT,* it is explicitly restricted to 'expressions for conceptions' that are 'formed by processes of addition subtraction and composition from given elementary conceptions x, y, z &c' (NT, [31]). In other words, though Boole would only be fully clear on this a couple of pages later (NT, [34]), it is restricted to expressions that are compositionally interpretable for any values of the variables (a hypothesis Boole will maintain throughout his methods in NT, as we shall see), and in particular that do not contain fractions.

The strategy of the proof is essentially the same as later used in *Schröder 1877,* 15. If we examine U 'with respect to x', Boole wrote (NT, [31]),

> it is plain that but three kinds of terms can appear in its expression *viz* terms containing x, terms containing $\bar{x}$ and terms containing neither x nor $\bar{x}$.
>
> But since $x + \bar{x} = 1$[,] any term A not containing x or $\bar{x}$ can be replaced by $A(x + \bar{x})$[,] ie by the aggregate $Ax + A\bar{x}$. Thus U can be brought to a form in which *all* the terms contain either x or $\bar{x}$.
>
> Hence [U] can be reduced to the form $U = Bx + C\bar{x}$.

Notice that the last step, which requires factorizing by x and $\bar{x}$, is not always legitimate if one wants to preserve compositional interpretability: in a formula such as $xy + xz = x(y + z)$, the left-hand side can be compositionally interpretable while $y + z$, and hence the right-hand side, is not. Incidentally, Schröder would no longer have this problem since, like Jevons, he used an inclusive reading of +; this is a clear illustration of why the task of avoiding compositionally uninterpretable expressions is trickier for Boole than in later forms of the algebra of logic. Nevertheless, since Boole in effect presupposes that his expressions are compositionally interpretable *for any assignment of classes to the variables*, the factorization is indeed legitimate here: when x is assigned the universe (resp. nothing), U reduces to the terms which are to be gathered into B (resp. C), so that Boole's hypothesis guarantees that the sums forming B and C are indeed compositionally interpretable.

Boole generalizes this proof to any number of variables iteratively (each of B and C can be developed with respect to y, for instance, and so on). He also notices that in general, the coefficients in the development can be obtained by the method of *LT,* 'assign[ing] to x, y, z [...] the particular interpretations 1, 0 *ie* Universe Nothing in every possible combination', but he avoids notations such as $f(0)$, $g(0, 1)$, etc., probably in order to limit mathematical symbolism as much as possible. In NT, as opposed to *LT,* Boole is careful to emphasize that the symbols 0 and 1 represent classes, not numbers.

1. (*Expression*) Express the premises by symbolical equations.

Boole did not address this step explicitly in NT. However, in line with his practice in *LT,* he assumed in the subsequent steps that all expressions in the premises were compositionally interpretable for all assignments of classes to the variables – hence both compositionally and formally interpretable. He did not, however, make the more onerous assumption discussed in Chapter X of *LT* that the premises are independent.

2. (*Transposition*) For each premise, bring the equation corresponding to it into the form $V = 0$.

This corresponds to Proposition 2 of NT, [34]:

> Proposition 2. *Every generally interpretable equation is reducible to a generally interpretable equation of the form* $U = 0$.

Recall that 'generally interpretable' by Boole's definition means, in the terminology I introduced above, compositionally interpretable for any assignment of classes to the variables.

This proposition is accomplished by way of the following theorem (NT, [35]):

> Theorem. *Every equation of the form* $u = v$ *in which* u *and* v *represent general conceptions may be replaced without loss of generality by the equation*
>
> $$u\bar{v} + \bar{u}v = 0.$$
>
> And this is an equation which is interpretable whatever the constitution of u and v as classes may be.

The equivalent formula is the same as in Chapter X of *LT,* but here Boole offers a very careful interpretable proof, deriving each equation from the other with no detour through (potentially not compositionally interpretable) subtractions.

3. (*Reduction*) Bring all equations $V_1 = 0, V_2 = 0, \ldots, V_n = 0$ together into a single one of the form $V = 0$.

This corresponds to Proposition 3 of NT, [35]:

> Proposition 3. Every system of generally interpretable equations is reducible to an equivalent single equation of the form $U = 0$.

Again, the expression Boole uses for U is the same as in Chapter X of *LT,* but in NT he carefully shows the equivalence of the initial system of equations with his single combined equation by deriving each from the other, with no reliance on uninterpretable steps (NT, [36]–[37]).

4. (*Elimination*) Eliminate any unwanted 'literal symbols' (class variables) from $V = 0$, obtaining a new equation $V' = 0$ in which the unwanted symbols no longer appear.

Elimination corresponds to Proposition 4 of NT, [38]. The method is essentially the same as in *LT,* though expressed without a functional notation: Boole developed $V = 0$ as $Ax + B\bar{x} = 0$, and obtained the elimination as $AB = 0$ (which corresponds to $f(0)f(1) = 0$ if writing $V = f(x)$, since in the development $B = f(0)$ and $A = f(1)$). We saw above that this method does not raise any interpretability issues, provided V is assumed to be both compositionally and formally interpretable, which Boole does throughout NT. The *justification* of this method in *LT,* though, heavily relied on formal reasoning, and in NT Boole replaces it with one that does not.

Boole's argument is rather terse and it may be useful to spell it out. The equations $A = 0$ and $B = 0$ 'express what the equation $V = 0$ becomes under the respective suppositions that x represents the Universe and that it represents Nothing', so that $AB = 0$ expresses 'what is involved in common in these suppositions' (NT, [38]). In other words, $AB = 0$ is the most we can conclude from $V = 0$ if we know that x is either the universe or nothing, but do not know which. This means that the result of the elimination of x is *at most as strong* as $AB = 0$. Indeed, if x was assigned some other fixed class s in between 0 and 1, $V = 0$ would lead to some further equation $C = 0$, so that the most we could conclude from $V = 0$ if we knew that x was either 0, 1 or s but did not know which would be $ABC = 0$ – an equation that, since ABC is included in or equal to AB, is at most as strong as $AB = 0$. In general, by considering further cases for the interpretation of x, we can only weaken our conclusion. Then, Boole shows that the equation $AB = 0$ follows from $V = 0$, so that the result of elimination is *at least as strong* as $AB = 0$. Thus, $AB = 0$ is the 'complete result of the elimination in question': it 'expresses *all* the truth that can be established independently' of the interpretation of x (NT, [38]).

5. (*Solution*) Develop $V' = 0$ as $Ax + B\bar{x} = 0$, then solve for x.

Remember that Boole's method in *LT* raised multiple issues: it involved an uninterpretable expression, which in the present notation is $B - A$; it required developing an expression containing a fraction, even though Boole had not discussed the meaning of division; and it led to a development containing such fractions as $\frac{0}{0}$ and $\frac{1}{0}$, whose interpretation by Boole appeared ad hoc to readers. In NT, Boole tackled all three.

First, instead of transforming the equation $Ax + B\bar{x} = 0$ into $(B - A)x = B$, which involves the uninterpretable expression $B - A$, Boole carefully proved (NT, [43]–[44]) that the equation $Ax + B\bar{x} = 0$ is equivalent to

$$(A\bar{B} + \bar{A}B)x = B,$$

which raises no interpretability issues.

Second, Boole defined division explicitly. This is in fact less of a shift than might initially appear. It is true that in *LT,* Boole did not say much about what 'mental operation' – or what words in ordinary language – the sign of division might correspond to; he just noted, quite tersely, that 'the mental operation which is represented by removing a symbol, z, from a combination zx' is 'identical with what is commonly termed Abstraction' (*LT,* 37). In various drafts, he elaborated upon this remark, sometimes describing the operation corresponding to the sign of division as 'Abstraction', sometimes as a more general operation of which abstraction is a special case (*SMLP,* 58; 75; 79–80; 93). Boole's treatment of division in NT does not pick up on this theme, but is in line with another remark he made in *LT,* and reiterated over and over in his manuscripts: division represents an *inverse operation*, and as such, 'its laws are dependent upon the laws already deduced' for the other operations (*LT,* 37) – as he put it in his long manuscript from 1856, 'abstraction is the inverse of composition and is fully defined by that relation', so that 'its laws do not [. . .] require to be separately determined' (*SMLP,* 79–80). This leads straightforwardly to the definition found in NT, [44]:

> Now let us adopt the inverse notation $\frac{M}{N}$ to denote the most general class[28] which possesses the property that the individuals common to it and to the class N will constitute the class M.

Following Shearman's 1906 history of symbolic logic, such a definition of division is often attributed to Venn;[29] however, it is made explicit in Boole's manuscripts, as *Panteki 1991,* 567–569 already noted, and one might even suggest that it is implicit in *LT*.

Third, Boole used his definition of division to interpret $\frac{0}{1}$, $\frac{0}{0}$, $\frac{1}{0}$, etc., as he did in many other late drafts. For instance, $\frac{0}{1}$ 'is a class such that the individuals common to it and the Universe make Nothing', hence it 'must itself be Nothing' (NT, [45]). The case of $\frac{1}{0}$ may seem more puzzling than the others: as Boole puts it (NT, [46]),

> [it] represents a class such that the individuals common to it and to Nothing constitutes the Universe. This is a contradiction – an impossibility.

Accordingly, when $\frac{1}{0}$ arises as a coefficient in the development of x, one has to assume, for the equation $(A\bar{B} + \bar{A}B)x = B$ to hold, that the corresponding constituent (in the instance, AB) vanishes. To check that this conclusion is correct, Boole offers a direct proof that $AB = 0$ is a consequence of the equation $Ax + B\bar{x} = 0$ (NT, [46]).

In contrast to *LT,* Boole is careful not to apply his theorem of development directly to the fractional form. Indeed, defining logical division as the inverse of logical multiplication

[28]In general, there will be many classes whose intersection with N is M, and Boole does not intend $\frac{M}{N}$ to denote a specific one. The reference to the 'most general' class is to be read in line with algebraic analysis, whose legacy is still present here: it means that $\frac{M}{N}$ will be equal to a general expression, containing an indeterminate term, under which all possible solutions fall, just like 'the' integral of a function has to contain an indeterminate term.

[29]Venn defined logical division in language strikingly similar to Boole's in NT, 'the expression $\frac{x}{y}$ stands for a class, viz. for the most general class which will, on imposition of the restriction denoted by y, just curtail itself to x' (*Venn 1881,* 74), and Shearman wrote that 'it needed Venn's careful analysis to bring to light the logical meaning underlying [fractional forms]' (*Shearman 1906,* 67).

does not, in and of itself, license the free manipulation of fractional expressions. Boole states cautiously that writing x as a fraction,

$$x = \frac{B}{A\bar{B} + \bar{A}B},$$

is 'but a giving of expression to that definition of x which is contained in $[(A\bar{B} + \bar{A}B)x = B]$ but so giving it as to show that the form of x as dependent upon the class symbols A, B is the object of search' (NT, [44]). Moreover, while in *LT* the process of development was presented from the outset as applying to any expression, including fractional ones, the theorem of development provided in NT is formulated and proved only for non-fractional expressions. Thus, in NT, Boole proceeds differently (NT, [44]–[46]). He writes

$$x = \frac{B}{A\bar{B} + \bar{A}B} = sAB + tA\bar{B} + u\bar{A}B + v\bar{A}\bar{B}$$

and then determines the indeterminate coefficients s, t, u, v in a way that requires no more than the *definition* of the fraction, i.e., the fact the development must be such that, if multiplied by $A\bar{B} + \bar{A}B$, it produces B (an approach similar to the one *Brown 2009,* 319 suggests is implicit in *LT*).

In broad outline, Boole's procedure here is consistent with his practice, elsewhere in his work, with inverse operations and series expansions (to which his method of development is closely related). In the chapter on 'Symbolical Methods' of his *Treatise on Differential Equations* (*Boole 1859*), he used language strikingly parallel to the passages on division in NT. Taking the example of an operation $(\frac{d}{dx} + a)$, he writes of the notation $(\frac{d}{dx} + a)^{-1}$ that it represents the inverse operation, but (*Boole 1859,* 376, his emphasis)

> only in its inverse character, i.e. conveying no information as to how it is to be performed, but only telling us that it must be such, that if, having performed it on v, we perform on the result the operation $\frac{d}{dx} + a$ to which it is inverse, we shall reproduce v. [. . .] The inverse procedure is thus presented as one, *the effect of which the direct operation simply annuls*. This is its *definition*.

Again, discussing the equation $(\frac{d^n}{dx^n} + A_1\frac{d^{n-1}}{dx^{n-1}} \ldots + A_n)u = v$, he wrote that 'on the above principle of notation we should have $u = (\frac{d^n}{dx^n} + A_1\frac{d^{n-1}}{dx^{n-1}} \ldots + A_n)^{-1}v$, the latter equation 'differing in interpretation from [the former], not at all as touching the *relation* between u and v, but only as more distinctly presenting u as the object of search' (*Boole 1859,* 377). Given such a definition, how should inverse operations be performed? Boole wrote that they are 'forms of interrogation, the answers to which are *to be tested* by the performance of the direct operations' (*Boole 1859,* 378, my emphasis). He gave the example of division in arithmetic (*Boole 1859,* 376, Boole's emphasis):

> What is meant by dividing a by b is the seeking of a third number c, which when multiplied by b will produce a. And the very procedure by which this is effected consists not in any new and distinct operation for determining the subject c, but in a series of guesses, suggested by our prior *general* knowledge of the results of multiplication, and tested by multiplication.

In practice, to find the result of an inverse operation in the context of differential equations, Boole's strategy tended to be to assume the result under the form of some kind of development (e.g., a power series) with indeterminate coefficients, and then to apply the direct operation to determine the coefficients in question. Seen in this light, Boole's treatment of fractional forms in NT is unsurprising.

In the end, however, there is a residual difficulty with Boole's procedure, even by the standards of NT. The trouble lies in his assuming that x (or the fraction equal to x) can be developed as $x = sAB + tA\bar{B} + u\bar{A}B + v\bar{A}\bar{B}$, with indeterminate coefficients s, t, u, and v that Boole plainly takes not to contain A and B (since, in order to determine them, he then sets A and B equal to 0 and 1 in all possible combinations). This would only follow from the theorem of development as he stated and proved it in NT if x was actually given as an expression of A and B, which is not the case. As it happens, however, the problematic assumption is not actually required for Boole's method to go through. Indeed, notice that if there is any x satisfying the equation, one can always write it $x = xAB + xA\bar{B} + x\bar{A}B + x\bar{A}\bar{B}$ (this is immediate from the fact that the sum of all constituents is always 1, and does not require the theorem of development, which would only be needed to obtain coefficients that do not contain x). We can then develop all parts of $(A\bar{B} + \bar{A}B)x = B$ separately as

$$(0AB + 1A\bar{B} + 1\bar{A}B + 0\bar{A}\bar{B})(xAB + xA\bar{B} + x\bar{A}B + x\bar{A}\bar{B}) = 1AB + 0A\bar{B} + 1\bar{A}B + 0\bar{A}\bar{B}.$$

The product of two distinct constituents among AB, $A\bar{B}$, $\bar{A}B$ and $\bar{A}\bar{B}$ is always zero, while the square of any of these constituents is the constituent itself (for instance, $A\bar{B} \times \bar{A}B = 0$ and $(A\bar{B})^2 = A\bar{B}$). Expanding the left-hand side, the preceding equality thus becomes

$$0 \times xAB + 1 \times xA\bar{B} + 1 \times x\bar{A}B + 0 \times x\bar{A}\bar{B} = 1AB + 0A\bar{B} + 1\bar{A}B + 0\bar{A}\bar{B}.$$

Multiplying both sides by each of the constituents in turn, we get:

$$\begin{aligned}
0 \times xAB &= 1AB && \text{which requires } AB = 0;\\
1 \times xA\bar{B} &= 0A\bar{B} && \text{which requires } xA\bar{B} = 0;\\
1 \times x\bar{A}B &= 1\bar{A}B && \text{which requires } x\bar{A}B = \bar{A}B;\\
0 \times x\bar{A}\bar{B} &= 0\bar{A}\bar{B} && \text{which places no constraint on } \bar{A}\bar{B} \text{ or } x\bar{A}\bar{B}.
\end{aligned}$$

As expected, we then get $AB = 0$ and $x = \bar{A}B + v\bar{A}\bar{B}$, where v is an arbitrary class. It is easy to check that, given that $AB = 0$, any such x satisfies the initial equation $(A\bar{B} + \bar{A}B)x = B$. Boole may or may not have been aware that no onerous assumption on the development of x was required. However that may be, it certainly is not made explicit in the text.

6 Conclusion

In Boole's logical calculus, the operations of aggregation, denoted by +, and of subtraction, denoted by −, can only be performed under some conditions on their operands: for instance, $x + y$ can only be performed if the classes denoted by x and y are disjoint. However, Boole's conception of algebra licenses the free use the formal laws of his calculus, such as $xy + xz = x(y + z)$ or $x + y - z = x - z + y$, without taking the conditions on the possibility of operations into account. This quickly leads, in the course of the logical methods presented in the *Investigation of the Laws of Thought*, to expressions that are not 'interpretable' in the sense that it may not be possible to assign a class to them compositionally, based on the definition of operations and on an assignment of classes to the individual variables occurring in them.

Some readers thus felt that Boole was offering 'dark and symbolic processes', in the words of *Jevons 1864,* 75*,* which were inscrutable in themselves and could only be justified because they led to conclusions that could be checked independently of them. This was never Boole's view. As his drafts testify, he believed to the end of his life in the intrinsic legitimacy of using formal reasoning beyond conditions of interpretability, in his algebra of logic just as much as in algebra in general. Moreover, the fact that certain expressions encountered in his logical methods were not 'interpretable' in the straightforward, compositional sense mentioned above did not make them meaningless. His calculus offered means – the 'method of development' – to reduce any expression, even if uninterpretable in the compositional sense, to a normal form which could allow 'interpreting' it indirectly. In this way, Boole maintained, it was possible to 'interpret' any equation between logical expressions, and hence any intermediate step in his logical methods.

Nevertheless, Boole was well aware that formal reasoning through (compositionally) uninterpretable steps proved hard to accept for some readers. In August 1863, the young Jevons sent him pointed criticisms; in particular, he suggested adopting the law $x + x = x$, thus essentially shifting to an inclusive reading of +, which, in contrast to Boole's, required no conditions. Boole, it appears, scrambled to show that no such radical reform was required, and that his logical methods could be rewritten so as to avoid passing through any expression that would be uninterpretable in what I called the compositional sense.

The late manuscript presented here, heretofore neglected, appears to be the result of Boole's late-1863 efforts. Entitled 'On the Nature of Thought', it displays a clear-eyed understanding of where, in the methods of the *Investigation of the Laws of Thought*, compositionally uninterpretable expressions might arise, and of how to avoid them – though Boole still maintained there was no need to do so. While terse and quite unpolished in places, the manuscript is complete and largely successful in his aims. Why Boole felt unsatisfied by the draft is unclear. Perhaps, as suggested in a letter to Jevons, he realized that his other commitments, in particular around the second edition of his *Treatise on Differential Equations*, would make it impossible to finalize 'On the Nature of Thought' in time to precede the publication of Jevons's criticisms and guarantee his claim to priority. Perhaps an early reader – a colleague or his wife, Mary Everest Boole – recommended against its publication (the manuscript bears a few marks in pencil that appear to be in another hand than Boole's). However that may be, there is no indication that Boole ever attempted to publish 'On the Nature of Thought'. Less than a year after his last letter to Jevons, he passed away.

Acknowledgements

For helpful comments on an earlier version, I wish to thank Marie-José Durand-Richard, David Makinson, and Dirk Schlimm.

References

LT Boole, G. 1854. *An Investigation of the Laws of Thought*. London: Walton and Maberly.

MAL Boole, G. 1847. *The Mathematical Analysis of Logic*. Cambridge: Macmillan, Barclay, & Macmillan.

SLP Boole, G. 1952. *Studies in Logic and Probability*. Ed. by R. Rhees. London: Watts & Co.

SMLP Boole, G. 1997. *Selected Manuscripts on Logic and its Philosophy*. Ed. by I. Grattan-Guinness and G. Bornet. Basel: Birkhäuser.

Boole, G. 1857. 'On the application of the theory of probabilities to the question of the combination of testimonies or judgments'. *Transactions of the Royal Society of Edinburgh*, 21 (4), 597–653.

— 1859. *A Treatise on Differential Equations*. Cambridge: Macmillan and Co.

— 1862. 'On the theory of probabilities'. *Transactions of the Royal Society*, 152, 225–252.

Brown, F. M. 2009. 'George Boole's deductive system'. *Notre Dame Journal of Formal Logic*, 50 (3), 303–30.

Burris, S. N. and Sankappanavar, H. P. 2013. 'The Horn theory of Boole's partial algebras'. *The Bulletin of Symbolic Logic*, 19 (1), 97–105.

Durand-Richard, M.-J. 1996. 'L'École algébrique anglaise : les conditions conceptuelles et institutionnelles d'un calcul symbolique comme fondement de la connaissance'. In: *L'Europe mathématique. Histoires, Mythes, Identités*. Ed. by C. Goldstein, J. Gray, and J. Ritter. Paris: Éditions de la Maison des sciences de l'homme, 447–77.

— 2022. 'Boole's symbolized laws of thought facing empiricism'. In: *Logic in Question*. Ed. by J.-Y. Béziau, J.-P. Desclés, A. Moktefi, and A. C. Pascu. Cham: Birkhäuser.

Ferraro, G. and Panza, M. 2003. 'Developing into series and returning from series. A note on the foundations of eighteenth-century analysis'. *Historia Mathematica*, 30 (1), 17–46.

Grattan-Guinness, I. 1991. 'The correspondence between George Boole and Stanley Jevons, 1863–1864'. *History and Philosophy of Logic*, 12, 15–35.

Gregory, D. F. 1840. 'On the real nature of symbolical algebra'. *Transactions of the Royal Society of Edinburgh*, 14, 208–16. Reprinted in *Gregory 1865,* 1–13.

— 1865. *Mathematical Writings*. Ed. by W. Walton. Cambridge: Deighton, Bell, and Co.

Haffner, E. 2024. 'Going to the source(s) of sources in mathematicians' drafts'. In: *Research in History and Philosophy of Mathematics. The CSHPM 2022 Volume*. Ed. by M. Zack and D. Waszek. Cham: Birkhäuser, 83–110.

Hailperin, T. 1986. *Boole's Logic and Probability*. 2nd ed. Amsterdam: North-Holland.

Hesse, M. B. 1952. 'Boole's philosophy of logic'. *Annals of Science*, 8 (1), 61–81.

Jevons, W. S. 1864. *Pure Logic, or the Logic of Quality Apart from Quantity. With Remarks on Boole's System and on the Relation of Logic and Mathematics*. London: Edward Stanford.

MacHale, D. 2014. *The Life and Work of George Boole. A Prelude to the Digital Age*. With a forew. by I. Stewart. Cork: Cork University Press.

MacHale, D. and Cohen, Y. 2018. *New Light on George Boole*. Cork: Atrium – Cork University Press.

Makinson, D. 2022. 'Boole's indefinite symbols re-examined'. *Australasian Journal of Logic*, 19 (5), 167–181.

McDougall-Waters, J. and Fyfe, A. 2022. 'Editing the journals, 1850s–1870s'. In: *A History of Scientific Journals. Publishing at the Royal Society, 1665–2015*. London: UCL Press. Chap. 9, 296–330.

Panteki, M. 1991. *Relationships between Algebra, Differential Equations and Logic in England, 1800–1860*. PhD thesis, Middlesex University.

Parshall, K. 2011. 'Victorian algebra: The freedom to create new mathematical entities'. In: *Mathematics in Victorian Britain*. Ed. by R. Flood, A. Rice, and R. Wilson. Oxford: Oxford University Press, 339–46.

Richards, J. L. 1980. 'The art and the science of British algebra'. *Historia Mathematica*, 7 (3), 343–65.

Schröder, E. 1877. *Der Operationskreis des Logikkalkuls*. Leipzig: B. G. Teubner.

Shearman, A. T. 1906. *The Development of Symbolic Logic. A Critical-Historical Study of the Logical Calculus*. London: Williams and Norgate.

Venn, J. 1881. *Symbolic Logic*. London: Macmillan.

Waszek, D. 2025. 'Interpreting the uninterpretable. Noncompositionality, normal forms, and philosophy of algebra in Boole'. Forthcoming.

Waszek, D. and Schlimm, D. 2021. 'Calculus as method or calculus as rules? Boole and Frege on the aims of a logical calculus'. *Synthese*, 199 (5–6), 11913–11943.

Appendix: George Boole, 'On the Nature of Thought'

[B.1] On the Nature of Thought[30]

Reflecting upon the processes of the mathematical form of Logic which I have developed in a special treatise, and which forms the ground of that theory of Probability the analytical characters of which together with their consequences have lately been discussed by me in the Transactions of this Society, I have been led to form certain conclusions concerning the general nature of Thought which appear to me to be not unlikely to interest others. These & the foundations upon which they rest will form the subject of the following paper.

The term Logic may be used in a twofold sense. Our highest conception of it is that which is implied in the derivation of the term. As λόγος signifies not only the inward thought but also its outward manifestation[,] Logic in its primary sense is the Science of the laws of Thought as *expressed*. Considered in this light Logic is conversant about all Thought which admits of expression, whether by the signs of common language, or in the symbolical forms of mathematics.[31]

[B.2] But in a secondary & narrower yet more usual sense Logic is the Science of the Laws of Thought *as expressed in the terms of ordinary Language*. In this sense it may be more precisely termed the Logic of Class. The relations of which it takes account are relations of Class eg those of genus and species, whole and part, identity and difference, & so on. It is this Logic of which particular forms, as some would say, but of which, as most educated persons believe, the universal canons have been preserved to us from ancient days in the writings of Aristotle.

To this Logic of Class the [same] positive conclusions of the present essay will relate although their suggested consequences will be of wider application. One remark I wish to premise. Except in so far as independent investigation upon such a subject is in its own nature controversial, I desire to avoid controversy. If I refer to received views and principles it will be rather in order to do justice to the elements of truth which they seem to me to contain,

[30]This edition obeys the following conventions. Boole emphasized passages by underlining them; these emphases have been rendered as italics. All footnotes are editorial (there are none in Boole's manuscript). Periods are regularly absent or hard to discern; as Boole systematically uses capitals at the beginning of sentences, their position is nevertheless unambiguous and they have been added wherever needed. All further editorial interventions in the text are typeset within square brackets. Unless there is an accompanying footnote clarifying the issue, such interventions can be taken to fall into one of three easily distinguishable categories: conjectural readings; suggested commas (Boole's text is lightly punctuated, and commas have sometimes been inserted, within square brackets, to make reading easier); and corrections to equation numbering. The reason for the latter is that Boole did not finalize his numbering, and often left empty parentheses in which he clearly intended to insert references to equations. For the sake of readability, equation numbers have been inserted and harmonized throughout. More precisely, in the manuscript equations are numbered continuously (1) through (15) on pp. [20]–[27]. Then the numbering stops; as Boole clearly intended to continue it, additional equation numbers have been provided. Equations numbers then reappear on p. [42], but the numbering is reset: five further equations are numbered (1)–(5) on pp. [42]–[47]. To avoid confusion, the latter equations have been renumbered in continuity with the preceding text.

[31]On the back of p. [1] is an annotation in pencil that reads: 'Explain on p 7′ the nature of *representative* thought as distinct from formal thought.' Underneath, also in pencil but apparently by another hand, is written 'v. B50', with an arrow pointing to the first annotation.

than to direct attention to what I may regard as their defects. Indeed to whatever degree the slow progress of knowledge may tend to modify views once held without reservation, there is much [B.3] of ancient doctrine that appears to be destined to endure. Thus the division of the faculties of Thought into Conception, Judgment & Reasoning seems nothwithstanding all attempts to reduce them to an intellectual unity, to be founded upon a real distinction in the nature of these faculties. Again the recognised order of their development appears to be the true one. In contemplating a group of objects we are perhaps impressed with the fact of their *likeness* to each other. We notice the several qualities in which that likeness consists, we combine these in a general conception, we express that conception by a name. Separated in thought from all other things the things which that name represents constitute a *class*. We compare this class with other classes the conceptions of which have been formed by a similar process of thought. We become conscious of class relations. We perceive that one class is contained in another as a part in a whole or as a species in a genus. Hence general propositions by which such relations are expressed. Hence finally *reasoning* by which from propositions so formed other propositions are deduced as conclusions.

But the specific doctrines which have gained widest acceptance upon such questions as the following: What is the real nature and what are the most general canons of deductive inference? What are the primary [B.4] laws of the operations of Thought? are not of a kind to preclude the hope of some exact knowledge upon these subjects. I will refer briefly to two of these doctrines.

1st. No opinion is more widely diffused than that which[,] regarding the chief function of Thought to be reasoning, presents as the one universal canon of inference, that what is true of the genus is true of the contained species; the principle which finds expression in a well-known syllogistic form.

Now while it is difficult to conceive that this opinion should have received an assent so nearly universal during so many ages without containing an element of the truth, it seems clear that the principle is one which affords little aid in the practical difficulties of those sciences which chiefly depend upon reasoning and in which the value of that principle, were it really entitled to the high place a traditional scholasticism claims for it, would be the most signally manifest. Yet in such sciences questions which certainly belong somewhere to the theory of inference e.g. questions arising from the employment of signs do arise, demand discussion and lead through discussion to results of great practical value. It seems that our theories of inference ought to take account of more than is commonly recognised as belonging to their province. And it would seem too that it can be only through them rising to a higher generality that they will even enable us to understand the true nature even of the particular fragments of truth which they really contain.

[B.5] 2ndly The opinion of those deserves notice who[,] studying the nature of Thought rather by the analysis of the operations of Conception and Judgment than that of reasoning[,] have been led to the conclusion that its ultimate laws are the three following viz

1st The Principle of Identity affirming that a thing is what it is

2ndly The Principle of Contradiction affirming with respect to any subject and any attribute that it is impossible for the subject to possess and not to possess that attribute

3rdly The Principle of Excluded Middle affirming with respect to any subject and any attribute that the subject either possess or does not possess the attribute – any middle supposition being excluded.

Now this doctrine[32] rests upon what must be regarded as a true foundation – upon the idea that laws of thought are manifested, and need only to be read aright, in the forms of possible conceptions and in the forms of necessary propositions ie of propositions which are true in virtue of their form. And no theory of the developed forms of Logic can possibly be true and general which does not contain the elements of this doctrine. But at the same time it is true that it is only in a developed theory in which laws are seen to form [B.6] the groundwork and the directing power of general methods that the true place & office of particular laws can be perceived. It is true as elsewhere so here that nothing is really isolated[,] that all truths must be studied in their organic connexion.

In my work above referred to, I have endeavoured to construct the organic forms of this Logic of Class by considerations of a peculiar kind founded upon what may with propriety be termed the doctrine of substituted relations – viz the doctrine that the expressed forms of thought depend only upon certain ultimate formal laws and can be transferred from one sphere of thought to another without regard [to][33] the nature of their respective *subjects*, provided that the laws in question are formally the same in the two. By an analysis of the fundamental operations of that faculty of the mind which in Logic is termed Conception I have established a formal agreement between the laws of such operations & those of the operations of a certain Algebra viz of an Algebra the subject of which is not the general relations of number but only those of the particular numbers 0 and 1. What the developed forms of such an Algebra should be is a question which it is quite within the province of Mathematicians to decide and the correctness of the determination of them contained in my work has not I believe been disputed. If the essential procedure of Thought in reasoning depend upon formal laws[,] the developed forms of this Algebra ought, the key to their interpretation being once found, to become developed forms of the Logic of Class. And in the work in question I have shown what the interpretation is and have translated the conclusions of the one into conclusions of the other.

[B.7][34] The result is the establishment of a system of Logic which rests as touching its primary laws upon an actual analysis of the intellectual operations but as touching the developed forms of the science upon a substitution. Now although this may cause but little perplexity to those who by the previous direction of their studies have been made familiar with those wonderful applications of the doctrine of substituted relations to which in the

[32] The two words ‘this doctrine’ are underlined in pencil, and there is a question mark in pencil in the margin.

[33] Boole corrected ‘wholly regardless’ into ‘without regard to’, but did not strike out ‘of’, so the manuscript actually reads ‘without regard to of’.

[34] The bottom of page [6] has been heavily corrected: part of the page has been cut out and replaced by a new piece of paper. This rewriting has created a redundancy with the top of page [7], which reads: ‘If then the essential procedure of thought so far as it falls under the province of the Science of Logic, depend upon formal laws[,] the developed forms of the above Algebra must also be the developed forms of Logic. And in the work in question I have interpreted them as such.’ As this passage was apparently meant to be replaced by the extended version at the bottom of page [6], it has been omitted here.

higher departments of Analysis we owe so much, to others it must seem to involve a most arbitrary assumption. It must appear as no ordinary paradox that *that* Logic which is held to govern all the Sciences should owe its development to the relation in which it stands to a particular science and[,] what is more[,] to a very limited and before unnoticed portion of such a science.

I will not attempt to weaken the great apparent force of this difficulty by anticipating here the answer to it which is as I believe contained in the results of the following investigations. The idea which I have sought to carry out in these may be [thus stated].[35]

[B.8] Although that identity of formal law which constitutes the relationship between the Logic of Class and a certain Algebra suffers no exception within the limits in which comparison is possible[,] ie so long as the respective operations of thought in these two provinces[,] which though different in themselves obey the same formal laws[,][36] are both possible[,] yet the limits of possibility are not the same in the one province as in the other. They are wider in the Algebra than in the Logic. If the same symbolic forms are employed for both then such forms will admit of interpretation into what may be termed representative thought in the Algebra in cases in which they are not so interpretable in the other. In the former case a freedom of operation is possible without going beyond the conditions under which the knowledge of the laws of such operations was acquired which is not possible in the latter. And thus it happens that although the final *results* of the method admit of being interpreted & in the work are actually interpreted into general theorems of the Logic of Class yet the intermediate *processes* by which these results are obtained are not always so interpretable. Now it is my object in this essay to show that it is possible [B.9] to pass from the same system of primary laws to the same final results without transgressing on the way the limits of that kind of Thought with which the Logic of Class is concerned. I shall show what that Logic is and what is the interpretation of its forms and methods when developed from within but still upon the same basis of formal laws as before. And I shall endeavour [to][37] draw from the results of this investigation some at least probable inferences concerning the nature of Thought generally.

Nature of signs

In describing the office of signs it may with a certain propriety be said that they *represent* things and *express* thought.

There is a stricter propriety in this language than at first sight appears. For Language is thought uttered. Thought[,] taking its rise in those impressions which[,] through the constitution of our perceiving faculties[,] external things produce upon us[,] advances by the operation of our other faculties of comparison & abstraction to the general conceptions of which signs are the immediate utterance. The order of procedure is manifest. Things first presented in perception are in a certain sense reproduced & presented to us a second time in the substituted forms of language.

[B.10] It follows hence that signs serve as instruments of Thought.

[35] The last two words on the last line of the page are almost illegible; 'thus stated' is a plausible conjecture.

[36] The clause 'which though different. . . formal laws' is enclosed in brackets in pencil.

[37] The manuscript reads 'to to'.

For signs are representative of things. They express our conception of things and then by a process of substitution stand for the things. And thus standing for things contemplated not as individuals but as falling under the general conception of *class* they represent them under that relation which makes deductive inference possible.

As instruments of thought signs are arbitrary in their outward character[,] fixed as to their interpretation and their laws.

The truth that signs are arbitrary in their outward character is manifested in the actual diversity of languages. That they are fixed as to their interpretation is a truth which is familiarly expressed in the rule that the meaning of a word or of any other sign must not be ambiguous. Whatever meaning is once given to it must continue to be associated with it if language is to be either definite as a medium of communication or exact as an instrument of thought. Again signs are fixed as concerns their laws. For the interpretation of a sign having been fixed its *use* as manifested in the nature of its combinations with other signs is fixed also. All intelligible language is [B.11] organic in its structure. It owes its significance not simply to the meaning of the signs employed but to their combinations. Now it is the general rules of such combinations[,] the rules determining the variety of forms under which such combinations are intelligible[,] which constitute the *laws* of signs. These laws are a visible expression of the laws of Thought.

The laws of signs are in a peculiar sense expressions of the *formal* laws of Thought. It may with the greatest propriety be said that the laws of signs express not the conditions under which the various intellectual operations involved in them are possible but the forms which when possible their expression assumes. Strictly speaking a formal law is one which determines the permitted variety of form in the expression of thought which arises from the nature of the thought itself. For instance in thinking of a whole as formed of parts we must[,] in order to picture that whole to ourselves[,] regard the parts as separate. This is a condition under which we think, a condition seated in the nature of that faculty by which we aggregate conceptions of parts together so as to form the conception of a whole. It is a condition which is of course realized in *things* when capable of being thought of as parts constituting a whole. But supposing this condition [B.12] satisfied the order in which the parts are aggregated together in thought so as to form the conception of the whole is indifferent. We must[,] it is true[,] think of the parts in some order but it matters not what that order is so far as concerns the operation of thought under consideration. If we express the mental operation by the conjunction *and* we should arrive at the formal law that in the expression of a whole the terms which it connects may be transposed. Animals & vegetables would express the same whole as vegetables and animals. There might be other reasons for preferring the one expression to the other but the two are equivalent with reference to the particular object supposed to be in view – the expression of a whole. [B.13] Again in thinking of an object which possesses & at the same time thinking of it as possessing two independent properties we think of these properties in *succession*, and in expressing our thought we represent them in language in a corresponding *order*. It may be that this order of thought has nothing responding to it in the order of nature. There the properties may be simply coexistent. Or if a ground of precedency exist it may be one with which our thought of them is wholly unconcerned. In either case the fact that in thinking of them we must think of them in a certain order while at the same time that order is indifferent constitutes a quality of the thought itself which finds expression in a formal law. If by x we represent one property, by y the other, & by these symbols written together their combination in actual existence, we shall have

$$xy = yx.$$

And this is the expression of a formal law of thought.

If we inquire into the ground of the existence of such laws as these[,] it will perhaps appear that it is founded in the nature of thought itself as an activity [B.14] which operates under the conditions of time and succession.

If we consider the mode in which this particular operation modifies the conception of Class we shall see that its office is to select as well as to attribute. A Universe or sphere of conception is in all discourse presupposed – it may be the actual Universe of all existences – it may be a particular province of it. Let us suppose the discourse to be of animated beings. Then, the use of any particular class term as 'fishes' limits that conception in a definite manner. Combine with this another class term as 'edible' & we have a further limitation, and so on in succession. The beginning of thought in this order is the conception of the Universe [–] existence, its final limit is that of nothing – nonexistence. But its progress *before* that limit is reached is always through conceptions successively narrower in extent but richer in comprehension.

These illustrations will serve to explain the object of the following section in which the general formal laws of conception are investigated.

[B.15] *Formal laws of conception*

For the general expression of the formal laws to which attention will be directed the following notation is convenient.

Let literal symbols as x, y, z be employed to represent classes of things or[,] according to the distinction explained in Art [38][,] to express our conceptions of classes.

Let xy denote that class which possess at once the characteristics of both the classes denoted by x & y.

Let $x + y$ represent the collection of things formed of the classes x and y together. The office of the symbol + is that of aggregation. It denotes that operation by which we connect parts into a whole.

The expression $x + y$ does not[,] it is evident[,] express a real conception unless x and y represent classes of things which have no members common. We may express this by saying that $x + y$ is not interpretable unless the conceptions expressed by x and y are mutually exclusive.

Let $x - y$ represent the remainder which is left when the whole of the class y is taken away from the class x.

This supposes the class y to be contained in the class x as a part in a whole.

[38] Boole left a blank space; he apparently intended to subdivide his text into numbered articles. Here, he likely meant to refer to the beginning of the section entitled 'Nature of signs', [9]–[10].

Let the enclosing of any expression in brackets denote that the group or class of things which it represents is to be regarded as a single class [B.16] which may as such be used in aggregation or composition with other classes.

Thus $x(y + z)$ would denote that class of things every member of which belongs at once to the class x and to the group formed by uniting by aggregation the mutually exclusive classes of y and z.

Let Nothing be represented by 0.

Let the Universe be represented by 1.

This mode of expressing the conceptions of Nothing & Universe is adopted from the 'Laws of Thought' where it is employed upon the ground of the identity which is there proved to exist between the formal laws of the conception Nothing in Logic and the number 0 in the science of Number and between the formal laws of the conception Universe in Logic and the number 1 in Arithmetic. Here[,] though we retain the notation[,] we dismiss for the present the analogy. No part of the following exposition would be affected if we represented the conceptions of Nothing and Universe by definite literal symbols just as we here express ordinary class conceptions, provided that those symbols were used in subjection to formal laws founded upon their peculiar interpretation – laws which would prove identical with those we [B.17] shall establish for the symbols 0 and 1.

As 1 represents the Universe it follows that if x represent a particular class of things, $1 - x$ will represent the class of things which remains when from the Universe the class x is in thought removed. Thus if x represented *men* $1 - x$ would represent in its primary signification the remainder of beings or existences left in the Universe when *men* had been in thought taken out of it – the class which in ordinary logical language is termed *not*-men. I believe however the use of this negative definition to rest upon a distinct intellectual act – to involve a substitution, lawful indeed, from another province of Thought; viz that in which by a reflex act the mind's own judgments are made the subjects of affirmation or denial.

The expression $1 - x$ as representing equally with x a class may in the same way [enter][39] in composition and the other operations of Thought by which conceptions are modified. In this way $x(1 - y)$ would represent the class of things which possesses the property denoted by x but wants that denoted by y.

The sign = will be used to denote identity[,] ie to denote that any two expressions between which it is placed represent classes of things which are identical as respects the individuals of which they consist[,] although these individuals are in the two expressions presented under different aspects of thought. [B.18][40] The members connected by the sign = must[,] in order that the equation may be interpretable[,] represent classes[,] but this is the only restriction. These classes may be, to any extent, complicated in form and composition.

Thus the equation

$$x = yz$$

[39] A word seems to be missing from this sentence; 'enter' is a suggestion for readability.

[40] The page starts with the line, redundant with the preceding page, 'of thought in the two members'. The latter part 'in the two members' is struck out in ink, probably by Boole himself, but 'of thought' is only struck out in pencil.

would express the identity of the members of the class x with the members of the class composed of all individuals which are contained at once in the classes y and z.

So too the equation

$$x = y\overline{z} + \overline{y}z$$

would express the identity of the members of the class x with those of the aggregate class formed of those members of the class y which are not contained in z and those of z which are not contained in y.

If for brevity we speak of the class x as the x's then the last equation might be interpreted more in accordance with common language by saying that the x's consist of all individuals that are either y's but not z's or z's but not y's.

But this introduction of the conjunction either appears to me like that of *not* to be an adoption from another part of Logic.

[B.19] In propositions in which[,] as in the vast majority of those of ordinary discourse[,] a predication is made[,] an identity is virtually affirmed to exist between the members of the subject class and an undefined portion of the members of the predicate class. We express then such propositions in the present scheme by introducing a class symbol v[41] denoting the undefined condition by which the predicate class is limited.

Thus the equation

$$y = vx$$

expresses that inclusion of the class y in the class x which is affirmed in the proposition All y's are x's. It is therefore the symbolic expression of this proposition.

The degree of indefiniteness of the class symbol v will depend upon the precise meaning which the proposition is intended to convey. Sometimes and especially in the expression of general principles[,] the form All y's are x's is meant to signify All y's if any exist are x's. Here the v is absolutely indefinite, it may mean any class not even the extreme limits of class extension Nothing Universe being excluded. But more usually and always in the statement of facts the *existence* of the subject All y's is assumed. In this case v represents a class which while otherwise indefinite must be supposed to include at least one individual of the class y.

[B.20][42] This may suffice as to the mode of expression of propositions.

And now let us consider to what formal laws the symbols are subject.

It is evident that we shall have

$$x + y = y + x \tag{1}$$

$$x + y - z = x - z + y. \tag{2}$$

These are particular expressions of a general law illustrated in Art [43] and which may be thus stated. The order in which class symbols connected by the signs + and − follow each other is indifferent.

[41] The class symbol v is underlined.

[42] The page starts with the line 'individual or individuals of the class y', redundant with the previous page; it is struck out in pencil and has been omitted here.

[43] Boole left a blank space; he is likely referring to his discussion that 'Animals & vegetables [...] express the same whole as vegetables and animals', [12].

Next we shall have

$$xy = yx. \tag{3}$$

This is only the expression of the law established in Art .[44] It may be stated by saying that in compositions the literal symbols x, y, ... are commutative.

Next it is evident that we shall have

$$x(y + z) = xy + xz \tag{4}$$

$$x(y - z) = xy - xz. \tag{5}$$

These are different expressions of the law that when we select all the members which possess a given quality x from a group or class formed whether by the parts into a whole or the removal of parts from a whole[,] the result is the same as if we had performed this act of selection upon the component parts or wholes first and then effected the combination in question. It may be stated by saying that the symbols x, y, z in composition are distributive.

[B.21] Lastly we shall have

$$xx = x \tag{6}$$

the expression of the law that when two conceptions are identical their composition does not produce a new conception but only reproduces the *one* which existed before. As the formal law last demonstrated holds true simply because x is the expression of a conception[,] it remains true for all the forms by which conceptions are expressed. Thus since $\overline{x}$ expresses a conception we should have

$$\overline{x}\overline{x} = \overline{x}$$

and we may assure ourselves that this is really the case in virtue of the formal constitution of $\overline{x}$ as the equivalent of $1 - x$. Thus

$$(1 - x)(1 - x) = 1 - x - x(1 - x).$$

Now $x(1 - x) = x - xx = 0$ by (6). Thus we have

$$(1 - x)(1 - x) = 1 - x$$

as was to be proved.

Thus also since xy represents a conception we shall have $(xy)(xy) = xy$. And this may be formally proved thus

$$\begin{aligned} (xy)(xy) &= xyxy \\ &= xxyy \quad \text{by (3)} \\ &= xy \quad \text{by (6)} \end{aligned}$$

We have now investigated the formal laws of the symbols and symbolical combinations by which the processes of thought in conception are expressed. Our [B.22] knowledge of

[44] Blank space, likely meant for a reference to p. [13].

them is derived from cases in which the elementary operations are *possible*. For instance the truth of the relation

$$x + y = y + x$$

is made known to us by reflecting upon the case in which x & y represent classes no members of which are common, & the truth of the equation

$$x + y - z = x - z + y$$

from the case in which[,] while x and y have no members in common[,] all the members of z are contained in x. These are conditions under which the forms themselves become interpretable. That the relations possess a truth beyond this will be shown hereafter. But I shall not assume here that such is the case. Here I seek only to develope the theory of the forms of Logic *under the conditions of interpretability*.

It will be observed that such conditions exist only in connexion with the operations denoted by the signs + and −. The operation of composition is always a possible one. If the symbols x and y represent classes known to have no members in common the combination xy becomes [B.23] equivalent to 0 and therefore does not pass *beyond* the limits of conception though it does reach one of them.

Let us now consider the formal laws which have reference to *equations*.

It is evident that if

$$x = y$$

represent a proposed equation then still under the conditions of interpretability[45]

$$x + z = y + z \tag{7}$$

$$x - z = y - z \tag{8}$$

and in the same way[,] only independently of conditions

$$zx = zy \tag{9}$$

These are expressions of the one general law that *if both members of an equation be affected by the same operation whether of aggregation or subtraction or composition, the equation remains true.*

If there be two equations as

$$x = y$$

$$w = z$$

then under the conditions of interpretability

$$x + w = y + z \tag{10}$$

$$x - w = y - z \tag{11}$$

[45]There is a question mark in pencil in the margin next to ‘the conditions of interpretability’.

and independently of conditions

$$xw = yz \tag{12}$$

If the equations have one member common so as to admit of expression in the form

$$x = y$$
$$x = z$$

then the result of composition becomes by (6)

$$x = yz \tag{13}$$

[B.24] These are expressions of the law that if the corresponding members of two equations be combined whether in the way of addition or subtraction or composition the resulting equation is valid.

It follows from this law that any term may be removed from one side of an equation to the other provided that its sign be changed.

Thus if

$$x + y = z$$

then subtracting y from both members

$$x = z - y \tag{14}$$

Thus y has been transposed with changed sign.

These are the primary laws of the modifications of equations dependent upon the operations of addition subtraction and composition.

It will be observed that the second of these operations is inverse to the first; and it might even [be][46] *defined* by this inverse relation. To subtract x from y might be defined as an act of thought by which we form a third conception z such that if to it we add the conception x we shall obtain the conception y. In the same way if the operation of subtraction were supposed to be known in itself[,] that of addition might be defined as its inverse. To add x to y would be to find a third conception z such that if from it [B.25] we subtracted x we should obtain y. And these definitions would fully suffice for the determination of the formal laws of that operation which is regarded as inverse. For such laws would be an immediate consequence of the formal laws of the operation which is supposed direct and primary.

It can however scarcely be doubted that in this instance the operation of addition is really the primary one. For in substracting x from y we must conceive of y as a whole containing x as a part, and the forming of this conception of x involves at least a virtual performance of the operation of addition.

There exists also an operation or more properly a procedure of thought inverse to that of composition. I say a procedure of thought because it is not of the nature of a simple elementary operation. It admits of no independent definition but can only be understood by

[46] This is a correction (the manuscript has ‘by’).

the relation in which its object stands to the operation of composition. And this object may be thus stated. What is that conception which by composition with a given conception x produces a given conception y? The laws of the inverse procedure here indicated will be investigated in the sequel.

It would seem therefore to be just to maintain that there are but two primary operations by which general conceptions[,] once formed[,] admit of being modified by combination viz the operations of addition & composition; while from the very constitution of the [B.26] mind in virtue of which operations modifying subjects are possible or, to adopt the language of analogy, from that in the nature of thought which is analogous to causal activity each operation suggests its inverse. If still adhering to this analogy without presupposing that it is more than analogy, we view a direct operation as a deduction of effect from cause[,] the corresponding inverse must be regarded as an inquiry what the cause from which a given effect may be deduced must be.

[B.27] *Consequences of the foregoing laws*

The most remarkable of the foregoing laws[,] both in itself and in its consequences[,] is that of which the formal expression is $x^2 = x$. Let us seek to unfold these consequences.

The equation after being reduced by transposition to the form

$$x - xx = 0$$

is seen to be equivalent to

$$x(1 - x) = 0 \tag{15}$$

of which the precise interpretation is *There are no things common to the class x and the class* $1 - x$.

If then we give to $1 - x$ its *primary* significance[,] interpreting [it][47] as the remainder left when the class x is taken away from the Universe of Thought[,] we see that the above is equivalent to *There are no things which belong at once to the class x and to the remainder of things left when the class x is taken away from the Universe.*

But if we give to $1 - x$ its secondary interpretation as *not-x's* we must interpret the equation by '*There are no things which are at once x's and not-x's*' or viewing this as the expression not of a fact but of a law *It is impossible that a thing should at the same time possess an attribute and not possess it.* This is the great principle of contradiction (Aristotle).

That the principle of contradiction is most inti-[B.28] mately connected with the formal law in question is then manifest. It is a direct consequence of that law if we assume the right to interpret $1 - x$ as *not-x's*. But as it admits of an interpretation without this assumption it would seem that such interpretation must be regarded as the primary expression of a principle which only becomes the principle of contradiction when[,] introducing an element not originally involved in conception *viz* the element of *negation*[,] we interpret $1 - x$ as *not x's*.

We are here brought to the threshold of a most important part of the subject *viz* the theory of the division of Logic into its two main provinces. For either we have to [do][48] directly

[47] Suggested insertion for readability.

[48] A word seems to be missing here; 'do' is a suggestion for readability.

with things their conceptions and relations as expressed in propositions[,] or by a reflex act of the mind the propositions themselves become the objects of thought and are themselves contemplated as subjects of relations viz the relations of truth and falsehood. It appears to me that in this reflex mode of thought such expressions as that of *not x's* must have arisen. The idea conveyed is not that of simple removal in exclusion but of denial. Forms of expression which have thus arisen may by a substitution find their place among others which have had no such reflex origin[,] and this I conceive to have been the case not only in the usual statement of the principle of contradiction but also in much of the ordinary language of mankind.

In their purely symbolical development [B.29] the theory of that part of logic which has reference to the relations of *things* and the theory of that part which has reference to the relations of *propositions* are absolutely parallel[,] governed by the same laws and differing only in interpretation. The language of ordinary thought recognises these two divisions in the main[,] but does not scruple for the sake of brevity or convenience sometimes to introduce into the one what strictly belongs to the other.

Let us examine the consequences of the law ([15]) as respects the expression of conceptions formed by composition.

Suppose there to be only two elementary conceptions involved viz x and y. Then as we can combine either x or $\overline{x}$ with either y or $\overline{y}$ but not x with $\overline{x}$ or y with $\overline{y}$ we have the four possible combinations

$$xy \qquad x\bar{y} \qquad \bar{x}y \qquad \bar{x}\bar{y}$$

Each of these represents a definite conception. And each obeys as such (Art [49]) the same formal law to which the elementary conceptions x & y are subject.

Again suppose three elementary conceptions x, y, z to be involved. Then as x or $\bar{x}$ may be combined with y or $\bar{y}$ and the result again with z or $\bar{z}$ we have in the whole eight possible combinations viz

$$\begin{array}{cccc} xyz & xy\bar{z} & x\bar{y}z & x\bar{y}\bar{z} \\ \bar{x}yz & \bar{x}y\bar{z} & \bar{x}\bar{y}z & \bar{x}\bar{y}\bar{z} \end{array}$$

and each of these represents a definite conception. Thus $x\bar{y}\bar{z}$ represents the class of things the members of which belong to the class represented by x but [B.30] are excluded from the classes represented by y and z.

I have in the Laws of Thought termed the expressions of conceptions which are thus formed constituents. And I shall adopt this language for the sake of uniformity here. It is important to notice that while in any set of such constituents, each expresses a class[,] those classes are mutually exclusive so that the composition of two of them is always equivalent to 0. Thus xyz and $xy\bar{z}$ represent classes which differ in that one of them consists of individuals which belong to the class z[,] the other consists of individuals all which belong to the class *not* z. If we represented xyz by u and $xy\bar{z}$ by v the combination uv would become equal to 0 because of its involving the combination $z\bar{z}$. And so in all other cases.

[49]Blank space, likely intended for a reference to p. [21].

Again all the classes represented by a set of constituents will[,] when taken together[,] form the Universe. This is evident from the law of their formation. For since $\bar{x} = 1 - x$ we have

$$x + \bar{x} = 1$$

Again the sum of the four constituents

$$xy \qquad x\bar{y} \qquad \bar{x}y \qquad \bar{x}\bar{y}$$

is equal to

$$\begin{aligned} &x(y + \bar{y}) + \bar{x}(y + \bar{y}) \\ &\quad x + \bar{x} \quad \text{since} \quad y + \bar{y} = 1 \\ &\quad = 1 \end{aligned}$$

Every thing thus falls under one or other of the mutually exclusive classes represented by a set of constituents. And this is the most general form of that [B.31] principle which is known under the name of 'Excluded Middle' and which consists in the affirmation *Every individual thing must either possess a given quality x or its opposite* $1 - x$.

We are now able to assign the one general form to which all expressions for conceptions are reducible. This however[,] and the other results which will follow in its train[,] it seems to [be][50] desirable on account of their importance to develope in distinct propositions.

Prop I. *Every conception formed by processes of addition subtraction and composition from given elementary conceptions x, y, z &c is expressible either as a constituent or as an aggregate of constituents formed from x, y, z . . .*

For let U represent the conception expressed in whatever way by means of the symbols x, y, z . . .

Considering first the constitution of [U][51] with respect to x[,] it is plain that but three kinds of terms can appear in its expression *viz* terms containing x, terms containing $\bar{x}$ and terms containing neither x nor $\bar{x}$.

But since $x + \bar{x} = 1$[,] any term A not containing x or $\bar{x}$ can be replaced by $A(x + \bar{x})$[,] ie by the aggregate $Ax + A\bar{x}$. Thus U can be brought to a form in which *all* the terms contain either x or $\bar{x}$.

Hence [U] can be reduced to the form

$$U = Bx + C\bar{x}$$

But U representing by hypothesis a *conception* we must have

$$UU = U$$

[50] 'Be' is inserted here in pencil; another possible correction is to delete the 'to'.

[51] Boole initially wrote v throughout the proof of Prop I, and later changed his notation to U, correcting most occurrences, but not all. For readability, all occurrences of v left uncorrected by Boole in the proof of Prop I have been replaced by U.

Therefore B and C must be such as to give

$$(Bx + C\bar{x})(Bx + C\bar{x}) = Bx + C\bar{x}.$$

[B.32] But

$$\begin{aligned}&(Bx + C\bar{x})(Bx + C\bar{x})\\ =&Bx(Bx + C\bar{x})\\ &+ C\bar{x}(Bx + C\bar{x})\\ &= BBx + CC\bar{x}\end{aligned}$$

since $x\bar{x} = 0$ and $xx = x$. That this expression may agree with $Bx + C\bar{x}$ we must have

$$BB = B \qquad CC = C$$

whence we see that B and C must represent conceptions. And these will be conceptions expressible without x & therefore by means of the other symbols y, z alone.

By the same reasoning then B and C must[,] when considered with respect to the symbol y[,] be of the forms

$$B = Dy + E\bar{y} \qquad C = Fy + G\bar{y}$$

in which D, E, F, G are themselves expressions of conceptions not involving x or y. Putting the above expressions for B and C in the previous expression for U we have

$$U = Dyx + E\bar{y}x + Fy\bar{x} + G\bar{y}\bar{x}.$$

Thus [U] considered with respect to the two symbols x & y must be reducible to a form in which at most the four constituents $yx \quad \bar{y}x \quad y\bar{x} \quad \bar{y}\bar{x}$ appear in composition with expressions which represent conceptions but which do not involve y or x.

The general law now becomes apparent. However far the resolution be carried[,] the expression for [U] will be composed of constituents in composition with expressions which are themselves expressions of conceptions but which do not contain any of the symbols involved in the constituents.

[B.33] Ultimately then[,] when the resolution has been effected with respect to all the symbols[,] we shall obtain for [U] an expression consisting of constituents affected by factors which themselves represent conceptions but do not contain any of the given symbols. Any such factors can thus only be either 0 or 1. The constituents affected by 1 form a simple aggregate of constituents unless there exist only one such. Any affected by 0 disappear. The ultimate expression of [U] is then a constituent or an aggregate of constituents. [As] was to be proved.

The form of this aggregate may also be determined in another way which deserves attention on account of the great importance of the principle involved. We are permitted to express [U] whatever its given form as an aggregate of all possible constituents affected by unknown factors. What these factors must be will depend solely upon the given form of [U], not upon the interpretation of the symbols x, y, z, [&c]. Assign to x, y, z then the particular interpretations 1, 0 *ie* Universe Nothing in every possible combination. It will be found that each such combination will determine one of the factors; and thus all will be determined in succession.

We will exemplify both methods by the same instance.[52]

[B.34] Proposition 2. *Every generally interpretable equation is reducible to a generally interpretable equation of the form* $U = 0$.

By a generally interpretable equation is meant one in which the members connected by the sign of equality = express general conceptions being formed of elementary conceptions $x, y, z, \ldots$ by processes which are possible independently of the particular interpretation of x, y, z.

Such an equation will therefore be of the form

$$u = v$$

in which u and v express general conceptions.

For compound each member with $\bar{v}$ and we have since $v\bar{v} = 0$

$$u\bar{v} = 0.$$

Again compound each member with $\bar{u}$ and we have since $u\bar{u} = 0$

$$\bar{u}v = 0.$$

Both these are interpretable equations. Add them together and we shall have

$$u\bar{v} + \bar{u}v = 0$$

which is also interpretable[,] its first member expressing the conception of the class formed by whatever is contained in the class u but not in v with whatever is contained in v and not in u; and the equation itself expressing that this is the conception of a class of things which does not exist.

Thus we we have obtained from the given equation an equation of the required form. Now is this equation as general as the one it is derived from?

[B.35] To prove that it is so we shall derive the given equation from it.

Compound each side of the equation last obtained with u and we have

$$u\bar{v} = 0$$

Operate in the same way with v and we have

$$\bar{u}v = 0$$

[52] Boole left the bottom of the page blank, likely in order to insert an example.

Hence

$$u\bar{v} = \bar{u}v.$$

To each side add the same class uv and we have

$$u\bar{v} + uv = \bar{u}v + uv$$

or

$$u(\bar{v} + v) = v(\bar{u} + u).$$

Therefore since $\bar{v} + v = 1$ and $\bar{u} + u =$ [1][53]

$$u = v$$

which is the original equation reproduced.

We are thus led to the following theorem.

Theorem. *Every equation of the form $u = v$ in which u and v represent general conceptions may be replaced without loss of generality by the equation*

$$u\bar{v} + \bar{u}v = 0.$$

And this is an equation which is interpretable whatever the constitution of u and v as classes may be.

Proposition 3. Every system of generally interpretable equations is reducible to an equivalent single equation of the form $U = 0$.

For by the last proposition the system is reducible to an equivalent system of the form

$$u = 0 \quad v = 0 \quad w = 0 \quad \&c. \qquad [(16)]$$

u, v, w &c being expressions of general conceptions.

[B.36] Operate on the second of these equations with $\bar{u}$[,] on the third with $\bar{u}\bar{v}$[,] and so on. The system becomes

$$\begin{aligned} &u = 0 \\ &v\bar{u} = 0 \\ &w\bar{u}\bar{v} = 0 \\ &\&c \end{aligned}$$

and under its present form the classes represented by the first members are mutually exclusive. The separate equations indeed after the first possess no longer the same degree of generality as before but they are valid equations.

Adding them together we have

$$u + v\bar{u} + w\bar{u}\bar{v} \ldots = 0 \qquad [(17)]$$

[53]The manuscript has '$\bar{u} + u = 0$', which is incorrect. An annotation in pencil next to this equation indeed reads '(shld be 1)'.

an equation the first member of which expresses a general conception. If the first member of that equation were represented by U the formal law

$$UU = U$$

would be identically satisfied.

The equation obtained is then an equation of the required form & it is derived from the system. I shall show that it virtually includes the system.

1st Operate on the equation with u and we have

$$u(u + v\bar{u} + w\bar{u}\bar{v} + \ldots) = 0$$

which reduces to

$$u = 0$$

since $uu = u$ and $u\bar{u} = 0$. Thus the first equation of [the][54] [B.37] original system has been reproduced.

2ndly Operate by composition on both sides of the equation with v and we have

$$v(u + v\bar{u} + w\bar{u}\bar{v}) = 0$$

which reduces to

$$vu + v\bar{u} = 0$$

$$\text{or} \quad v(u + \bar{u}) = 0$$

$$\text{or} \quad v = 0.$$

Thus the second equation has been reproduced.

3rdly Operate in like manner on both sides of the equation with w and we have

$$w(u + v\bar{u} + w\bar{u}\bar{v} \ldots) = 0$$

$$\text{or} \quad wu + wv\bar{u} + w\bar{u}\bar{v} = 0$$

$$\text{or} \quad wu + w\bar{u}(v + \bar{v}) = 0$$

or since $u + \bar{u} = 1$

$$w(u + \bar{u}) = 0$$

or

$$w = 0.$$

Thus the third equation has been reproduced.

And that this order must continue will become evident if we consider the meaning both of the system ([16]) and of the equation ([17]).

The former asserts the nonexistence of individuals of the classes u, v, w, &c. The latter asserts the nonexistence 1st of all things of the class u then of all things of the class v not

[54] The word ‘the’ appears to be missing at the page boundary and has been restored here.

already included in the class u, then of all things of the class w not [B.38] already included in the classes u and v and so on. This amounts to a denial of existence to all things included in any of the classes u, v, w, ...

Proposition 4. The elimination of any class symbol x from any equation or system of equations will be effected by reducing the given equation or system by the foregoing propositions to the form $V = 0$[,] reducing this equation by development with respect to x to the form

$$Ax + B\bar{x} = 0[,] \tag{[(18)]}$$

and then forming the equation

$$AB = 0.$$

And this equation will be the complete result of the elimination in question.

For by the law of development A is what V becomes when x is made equal to 1 and B is what U becomes when $x = 0$. Hence the equations

$$A = 0 \qquad B = 0$$

express what the equation $V = 0$ becomes under the respective suppositions that x represents the Universe and that it represents Nothing. Therefore the equation

$$AB = 0$$

expresses therefore what is involved in common in these suppositions. But the object of elimination is to find what is true and all that is true independently of the interpretation of x. If then we can show that the equation $AB = 0$ is true independently of the interpretation of x it will follow that it expresses *all* the truth that can be established independently of such interpretation. Else, that which was not true under particular assumptions would be true under [B.39] all assumptions independently of their particular nature.

Now if we compound the equation ([18]) with Bx we get

$$ABx = 0$$

and if we compound it with $A\bar{x}$ we get

$$AB\bar{x} = 0$$

and these give on addition

$$AB = 0.$$

Hence this equation is true; and it is the *general* result of the elimination of x from the given system.

It may be that in the given system of equations the symbol x which it is sought to eliminate appears in only certain of the equations. In this case it is best to reduce this portion of the system to an equivalent single equation[,] to eliminate x thence by the foregoing propositions[,] and then to combine the resulting equation with the remaining portion of the system.

But it makes no final difference whether we do this or reduce first the entire system to a single equation and then eliminate x from that system. This may be proved in the following manner.

Let

$$\begin{aligned} Ax + B\bar{x} &= 0 \\ C &= 0 \end{aligned} \qquad [(19)]$$

be the single equations which are respectively equivalent to that portion of the given system which contains x and to that portion which does not contain x. [B.40] Eliminating x from the former we have

$$AB = 0$$

and this in combination with the equation

$$C = 0$$

forms a system of two equations the single equivalent of which is by Prop 3

$$C + AB\bar{C} = 0$$

and this is the single final result of the elimination.

Now if we combine into a single equation the equations ([19]) before eliminating x we have

$$C + (Ax + B\bar{x})\bar{C} = 0$$

But C not containing x we have

$$C = Cx + C\bar{x}$$

Substituting the above equation developed with respect to x becomes

$$(C + A\bar{C})x + (C + B\bar{C})\bar{x} = 0$$

whence eliminating x

$$\begin{aligned} & (C + A\bar{C})(C + B\bar{C}) = 0 \\ \text{or} \quad & C(C + B\bar{C}) \\ & + A\bar{C}(C + B\bar{C}) = 0 \end{aligned}$$

or since $CC = C$ and $C\bar{C} = 0$

$$C + AB\bar{C} = 0$$

as before.

If we have two symbols x, y to eliminate we can either do it in succession, by the foregoing method[,] or simultaneously by the following which is an obvious extension of it:

Forming that equation $V = 0$ which is the equivalent of the portion of the por-[B.41] tion of the system in which the two symbols are contained develope it in the form

$$Axy + Bx\bar{y} + C\bar{x}y + D\bar{x}\bar{y} = 0.$$

Then will the equation

$$ABCD = 0$$

be the result of the elimination of x & y from that portion of the system. And this must be combined with the remaining portion of the system which does not contain x or y.

In general the complete result of the elimination of any number of symbols from a proposed system is obtained by forming the single equation which is equivalent to that portion of the system in which such variables are contained[,] developing that equation fully and combining by composition the factors of all the constituents. If any constituent do[es] not appear in the development it must be regarded as appearing but with 0 for its [coefficient].[55] The result of this elimination must be associated with the remaining equations of the system.

[B.42] Prop 5. From any system of equations, to deduce the complete expression of any one of the symbols x in terms either of all the others or of any portion of them and to interpret the result.

In the first case viz when one symbol is to be expressed in terms of all the other symbols let that single equation be formed which is equivalent to all the equations of the system.

In the second case viz when one symbol is to be expressed in terms of a portion of the others let all the others except that portion be eliminated and the single equation which is equivalent to the system after such elimination be formed.

In either case we have a single finite equation

$$V = 0$$

which developed in terms of x takes the form

$$Ax + B\bar{x} = 0 \qquad ([20])$$

in which A & B represent *conceptions*.

Now we are proceeding under the express condition that no forms are to be employed which are not interpretable. The effect of this condition[,] which as I shall afterwards show is not a necessary one[,] is to limit the freedom of our operations. In the present instance that effect consists in the necessitating of a transformation by which the equation will be presented under a different but equivalent form before that process can be applied upon which the solution [B.43][56] of the problem really depends – the process of development. It must therefore be understood that the next step is one which in the purely formal procedure of thought is wholly unnecessary and which is made here only in order to enable us to present

[55] This is a correction: in a slip of the pen, Boole wrote ‘constituent’ here.

[56] The words ‘the solution’ appear both before and after the page break. The second occurrence has been eliminated here.

under a certain condition of interpretability forms of Thought which as to their essence are independent of such conditions.

The transformation of the last equation which it is necessary to employ in order to secure the fulfillment of the above mentioned condition is the following

$$(A\bar{B} + \bar{A}B)x = B. \qquad ([21])$$

Its validity may be proved as follows.

The equation just written is equivalent to

$$A\bar{B}x + \bar{A}Bx = B$$

and it is evident that both its members represent conceptions. By the general theorem of Prop 2 the equation is therefore reducible to the form

$$(A\bar{B}x + \bar{A}Bx)\bar{B} + B(1 - A\bar{B}x - \bar{A}Bx) = 0$$

or

$$A\bar{B}x + B - \bar{A}Bx = 0$$

since on effecting the compositions the terms which would contain $B\bar{B}$ vanish.

[B.44] Now if we develope this with respect to A, B, and x as symbols of class, it assumes the form

$$ABx + A\bar{B}x + B\bar{A}\bar{x} + BA\bar{x} = 0$$

which is the same as

$$Ax + B\bar{x} = 0[,]$$

the equation ([20]). The validity of the transformation is therefore proved.

The transformed equation ([21]) implies that x is a class such that the individuals common to it and to the class signified by $A\bar{B} + \bar{A}B$ will constitute the class B.

Now let us adopt the inverse notation $\frac{M}{N}$ to denote the most general class which possesses the property that the individuals common to it and to the class N will constitute the class M.

Then ([21]) gives

$$x = \frac{B}{A\bar{B} + \bar{A}B} \qquad ([22])$$

This in fact is but a giving of expression to that definition of x which is contained in ([21]) but so giving it as to show that the form of x as dependent upon the class symbols A, B is the object of search.

Now let us apply to this expression the principles of that method of development which is illustrated in [57]. According to this the sought form of x in terms of A and B must consist of the constituents of A and B under relations which must emerge out of the definition involved in ([22]). We may therefore write [B.45]

$$x = \frac{B}{A\bar{B} + \bar{A}B} = sAB + tA\bar{B} + u\bar{A}B + v\bar{A}\bar{B} \qquad ([23])$$

[57] Blank space, likely intended for a reference to the discussion of development, [29]–[33].

s, t, u, v denoting the as yet unknown conditions under which the constituents AB, &c are to be placed. To determine these conditions let us proceed as follows.

1st Let $A = 0$ and $B = 1$, then since $\bar{A} = 1$, $\bar{B} = 0$ the second member of ([23]) becomes $\frac{1}{1}$ and the third u, & we have

$$u = \frac{1}{1}.$$

Hence u denotes a class such that the individuals common to it and the Universe constitute the Universe. It is clear then that u must itself be the Universe. We have therefore $u = 1$.

2ndly Let $A = 1 \quad B = 0$ then since $\bar{A} = 0 \quad \bar{B} = 1$ the equating of the second and third members of ([23]) will give

$$t = \frac{0}{1}.$$

Thus t is a class such that the individuals common to it and the Universe make Nothing. Hence t must itself be Nothing; we may then write $t = 0$.

3rdly Let $A = 0 \quad B = [0]$.[58] Then since $\bar{A} = 1 \quad \bar{B} = [1]$ we find by a procedure similar to the last

$$v = \frac{0}{0}.$$

[B.46] It follows that v represents a class such that the individuals common to it and to Nothing constitute Nothing – in other words a class such that if we take of it *none* we get as the result *Nothing*. But this is true of all classes whatever. Hence v represents *any* class. Beyond the condition that it does represent a class its interpretation is *indefinite*. We shall replace v by the form $\frac{0}{0}$ but with the interpretation now assigned.

4thly Let $A = 1 \quad B = 1$ then as $\bar{A} = 0 \quad \bar{B} = 0$ the equation becomes

$$\frac{1}{0} = s.$$

s then represents a class such [that][59] the individuals common to it and to Nothing constitutes the Universe. This is a contradiction – an impossibility. It implies not simply the nonexistence of any individuals of the class AB in the class x but the impossibility of conceiving the class AB as existing at all compatibly with the relations expressed in the actual data.

We can verify this result by ordinary reasoning. For the given equation

$$Ax + B\bar{x} = 0$$

breaks up into

$$Ax = 0 \qquad B\bar{x} = 0$$

the first of which expresses that there are no individuals common to the classes A & x[,] the second that there are none common to the class B and the [class $\bar{x}$,][60] [B.47] from which it is seen that there cannot be any individuals common to the classes A and B.

[58] In a slip of the pen, Boole wrote $B = 1$ and (in the next sentence) $\bar{B} = 0$, which would replicate the first case and is not consistent with what follows. This has been corrected here for readability.

[59] In the manuscript, the word 'that' is repeated as a result of a correction by Boole.

[60] Words seem to be missing at the page boundary here; instead, the word 'the' is repeated twice. The suggested insertion 'class $\bar{x}$' makes good sense in context.

The developed expression for x is

$$x = \frac{1}{0}AB + 0A\bar{B} + \bar{A}B + \frac{0}{0}\bar{A}\bar{B} \qquad ([24])$$

and this implies 1st that the class x consists of all members of the class $\bar{A}B$, & of an indefinite part, some none or all, of the members of the class $\bar{A}\bar{B}$[,] 2ndly that it contains no members of the class $A\bar{B}$[,] 3rdly that the class AB does not exist.

As the classes A and B are themselves not simple in expression but supposed to be formed by addition composition &c from the classes represented by these literal symbols of the original equation in terms of which x is ultimately to be expressed it will be necessary in applying the above theorem to develope AB, $A\bar{B}$, $\bar{A}B$ and $\bar{A}\bar{B}$ in terms of these symbols. Thus according to the above method we have a double process of development to perform before the desired final expression for x can be evolved out of the transformed equation ([21]). We might however obtain that result by a single process of development and this would be in accordance with the [B.48] procedure adopted in the 'Laws of Thought'. For this purpose instead of developing the expression for x in ([23]) viz

$$\frac{B}{A\bar{B} + \bar{A}B}$$

with respect to A, B as class symbols we should have to develope it on the same principle in terms of the literal symbols which are contained in A and B.

Prop 6. *Given any system of equations it is required to express any class defined by a portion of the symbols involved in these equations in terms of any other symbols of the same equations.*

This may be easily reduced to the last case. For if we represent the [B.49] class of which the definition in terms of one set of symbols is given and the expression in terms of another set required by t[,] and then conjoin with the given equations a new equation expressing the definition of t[,] we shall have a completed system consisting of the given equations with one new equation containing t. From this system we can eliminate those symbols which it is intended should not appear in the final result from the single equation which is equivalent to the reduced system[,] and then develope t in terms of the symbols which remain.

These propositions enable us to accomplish every object which lies within the scope of the formal Logic of class. And they enable us to do this without transgressing the conditions of interpretability.

But it is seen that the freedom of our procedure is restrained by these conditions. Before two equations can be added together they must undergo a previous preparation[61] unless they already satisfy the condition which that preparation is designed to produce. So the solution of the equation

$$Ax + B\bar{x} = 0$$

[61] There is a blank in the manuscript here (filled with a hard-to-decipher pencil annotation), which was probably intended for a reference to a previous passage. The obvious reference would be to Proposition 3, pp. [35]–[37].

[*lacuna*][62] to determine x[63] requires that the equation [be reduc]ed[64] as in ([21]). And the entire procedure [B.50] of thought as manifested in the foregoing propositions is one in which while the result of each step is determined by formal laws the *order* of the steps is restricted by the condition that each result as it arises shall be interpretable into actual representative thought.[65]

[62] The corner of the page is torn here and words are missing.

[63] The variable x is underlined.

[64] Again, there is a lacuna at the beginning of the line; 'be reduc' is suggested as a plausible insertion.

[65] Next to the last word is 'p 7′' written in pencil – an annotation Boole might have made himself. Underneath is a second pencil annotation, linked by a line to the first, which reads 'v B1 back' (referring to the annotation discussed above in footnote 31, p. 36) and was most likely made by a reader after Boole's death.