

The Propositional Content of Data

Dave S. Henley

Abstract

Our online interaction with information-systems may well provide the largest arena of formal logical reasoning in the world today. Presented here is a critique of the foundations of Logic, in which the metaphysical assumptions underlying traditional logic are contrasted with those of such 'closed world' reasoning. Closed worlds mostly employ a syntactic alternative to formal language namely, recording data in files. Whilst this may be unfamiliar as logical syntax, it is argued here that propositions are expressed by data stored in files which are essentially *non-linguistic* and so cannot be expressed by simple formulae $F(a)$, with the inference-rules normally used in Logic. Hence, the syntax of data may be said to define a fundamentally new kind of logical form for simple propositions. In this way, the logic of closed systems is shown to be non-classical, differing from traditional logic in its truth-conditions, inferences and metaphysics. This paper will be concerned mainly with how the reference and certain inferences in such a closed system differ metaphysically from classical logic.

1. Introduction

Closed information-systems employ a common 'closed world' logic which is just as universal as classical logic but differs significantly from it in terms of reference, truth and inference. Yet our use of these systems provides what is arguably the largest arena of formal logical reasoning in the world today. Such systems include hospitals, government departments, warehouses, airports, universities. As an example of the different logic being used, consider that if 'Jones' was not listed in the passenger manifest for a flight, we are entitled to logically infer that Jones was not on that flight. But in classical logic however, from the mere absence of a proposition, no proposition whatsoever can be inferred. The manifest describes a closed world, and no passenger is allowed to exist

outside the manifest. In other words, by the *logic of closed systems*, the world description provided by the passenger manifest is required to be complete. The fundamental metaphysical implications of this closed world logic that are presented here have been largely overlooked by philosophers, despite the global impact of such systems. The present paper will examine this metaphysics regarding mainly the referential and certain inferential differences; a future paper will analyse the metaphysics of the different truth-conditions implicit in this kind of discourse.

The simplest propositions employed in such logical discourse are normally formalised as data stored in files. These are treated as a closed system in the sense that the system ontology is presumed to be completely listed within the files; this assumption is called *domain closure*. To explore the metaphysics of such discourse, we adopt the terminology of Wittgenstein's *Tractatus*, in which the world was described as a totality of facts. If we apply this to a local world as given by files, then notationally, the syntax of the files is treated as a boundary to whatever exists within the system; anything not recorded by the files is said to be 'outside the system' and not within the universe of discourse. No such syntactic device was available to Wittgenstein however, in the Frege/Russell formalism that he used in the *Tractatus* to express the totality of facts. By imposing domain closure, such logical syntax can nevertheless now be seen to have great practical significance.

The users of such a closed system form a speech community, and it is the deductive inferences permitted by this speech that we shall here demonstrate to possess a non-classical logic. In a future paper, the truth-conditions of this speech will be found to also require a non-Tarski semantics. Here however, the focus will be on reference, and unlike the referring use of names, definite or indefinite descriptions, demonstratives or indexicals familiar from Fregean logic, it will be shown rather, that in the logic being studied here, objects are referred to *anonymously*, and that this is a new kind of reference which, as far as I am aware, has not previously been identified by philosophers.

Likewise when no record is stored ascribing a simple predicate G to an object, the user of a closed system is logically entitled to infer that G is not true of that object; this logical inference is called the *closed world assumption*. Classical logic however, assumes an open world since it is based upon the interpretation of linguistic formulae, formalising various sentences of natural language. And in this classical logic of sentences, the mere absence of a simple predicate G in a set of sentences believed to be true of an object does not in itself entitle one to infer anything about whether G is true or false of the object. Consequently, a file of records is unlike a set of sentences, since the deductive inferences possible from a file of data differ from those possible from a set of axioms. Yet despite our lives being dominated by such reasoning from data, it has been largely overlooked by philosophers as a valid mode of logical inference, in favour of reasoning from sentences.

This series of papers undertakes to systematically compare and contrast the propositional content of these two kinds of world-description: the nonlinguistic data in files, versus the linguistic formulae $F(a)$ of traditional logic. This difference in content is revealed by their different truth-conditions and modes of logical inference, together with their differing metaphysical implications. The argument will be structured with reference to how the philosophy of Wittgenstein developed. For it is suggested here that Wittgenstein was motivated by similar considerations concerning the limits of expression within different symbolic systems. We consider in particular, how easily numerical propositions can be expressed in the logic of a closed world, in the light of Wittgenstein's dissatisfaction with the complexity of their expression by classical logic. The subsequent syntactic development of Wittgenstein's thoughts on numerical expression, starting with the logic of *The Tractatus*, through his criticism of this logic in his middle period, to the *Philosophical Investigations* are all fruitfully tracked here from the unexpected perspective of domain closure and related considerations.

We begin by considering what Wittgenstein himself required of a logical notation; firstly, in describing totalities of facts and numbers of objects, and secondly in expressing what I call their ‘anonymity’.

To demonstrate how data in files are suited to meet these requirements, I break down the syntax of files into their primitive elements. Examining, in their simplest form, the underlying syntactic principles distinguishing the reference performed by data from reference by formal sentences, then enables the syntax of files to be viewed as a rudimentary kind of logical form. It is found that the primitive syntactic elements of files may then be adjudged as providing revised logical categories. For, just as variables, argument-places and quantifiers helped formalise earlier metaphysical ideas of substance, attributes, existence etc, so I show how the same may now be said of records, files, data-values. More specifically, the two above-mentioned issues that the logical calculus of Frege and Russell, made problematic for Wittgenstein, are now resolved if elementary propositions in a world description are conceived non-linguistically, as data in files.

My reason for proposing to treat the syntax of tabular data as a logical form for propositions, is that there are truth-conditions needed by closed systems and provided by physical operations on the syntax of data, which cannot be provided by physical operations on the syntax of simple formal sentences. And yet every logical operation performable on sentences of a formal language can also be performed equivalently, upon data in files. The additional truth-conditions to be discussed here are the two previously mentioned referential characteristics, totality and anonymity, that I believe serve to

distinguish data fundamentally from formal sentences. As a result, we may be entitled to say that for many elementary or non truth-functional propositions describing closed systems the syntax of data provides a better logical form than the syntax of formal sentences. To now specify these two referential characteristics further, we introduce each of them in the context in which it was first encountered by Wittgenstein, as follows.

Firstly, in the Tractatus Wittgenstein was very concerned, in a manner not unlike a modern database designer, with the logical possibility of giving a complete description of the world. At 4.26 he said that the world can be completely described by giving all true and false elementary propositions:

4.26 If all true elementary propositions are given, the result is a complete description of the world. The world is completely described by giving all elementary propositions and adding which of them are true and which false.

I shall argue that, for this to be the case then, by elementary propositions, Wittgenstein cannot just have meant interpreted simple (non truth-functional) sentences $F(a)$, in the notation of Frege and Russell. For, since the simple singular terms a, b, \dots in such sentences refer independently of one another, they cannot jointly ensure that *all objects in the world have been described*. Rather, such global reference was intended by Wittgenstein separately. Without this additional constraint, no universal or numerical facts about even a finite world can be deduced from the elementary sentences alone, and so such elementary propositions cannot be guaranteed to describe the world completely.

In this way, it is confirmed that Wittgenstein meant more by elementary propositions than just simple sentences, as ordinarily interpreted. I then argue the case for the *non-linguistic*

symbolism of stored data, used nowadays for this very purpose of world description. Files of data can be seen as systems of elementary propositions precisely meeting Wittgenstein's requirement, at least for a particular domain. In keeping with T4.26, files of data have the advantage that they impose the above referential constraint (domain closure) automatically, ie as part of their own interpretation. For a particular domain, it is clear that, unlike Fregean sentences, this enables elementary propositions (data) to express (Tr 3.01) 'all that is the case'. As well as all universal facts, this also includes all numerical facts since, when enumerating objects, we always assume we are counting their totality.

The second referential characteristic, anonymity, of files of data, was encountered by Wittgenstein in a purely numerical context. For, to know the number three of say, green apples, need not be to know *which* apples are green. It seems that to know the number of green apples, these apples do not need to be individually identified and can be *anonymous*. This was expressed linguistically by Wittgenstein in the Tractatus [spelled out eg in 1932]¹ by means of variables, using quantifiers, to say that the green apples can be assigned one-one to x_1, x_2, x_3 respectively, in such a way that there are no green apples left over, to assign to x_4 , where none of these apples need be named. However, in 1932 he expressed dissatisfaction with this quantificational account of number, as misrepresenting our actual use of numerals in counting. I shall argue that a formalisation of *records in a file* enables the same anonymity to be logically expressed for numerals, but avoiding, as Wittgenstein seemed to require, the apparently unnatural use of variables

¹ *Wittgenstein's Lectures Cambridge 1932-1935* Ed Ambrose

or quantifiers. And so the truths of Arithmetic will be deducible simply and naturally from Logic, once we allow files of data as a new non-classical form of logic.

Metaphysically, I aim to show that, whereas in his former definition of number the imposed use of a formal logical language required us to assume that, 'To be is to be the value of a bound variable' [Quine², Tractatus 4.1272], the propositions in a database require us to assume only, that to be, is to be recorded in a file without necessarily being identified by name³. Like the bound variable, this may be said to grant expression, by means of a formal symbolism, to the metaphysical idea of substance. More generally, when the very concept of a file of data is analysed into its primitive syntactic elements, these elements will sometimes be found to possess a logical significance conforming more closely with traditional metaphysical categories, than do the corresponding syntactic categories of modern Logic.

We now explore each of these linguistic/non-linguistic issues: totality, anonymity, in greater detail. The approach will be to start by examining each of these criteria of truth and reference with respect to only the most elementary constituents of which files are built. We will then gradually combine the semantic effects of these syntactic elements, arriving only at the end, at the full philosophical idea of records in a file.

2. Totalities of Facts

At Tractatus 1 Wittgenstein famously said 'the world is everything that is the case', from which we may infer that if the list of elementary propositions at T4.26 'describes the

² W V O Quine 'On What There Is' in *from a Logical Point of View* Harvard 1953

³ As with statistical data.

world completely' it must describe every fact about the objects in the world. However, if propositions are interpreted sentences of a formalised object language, ie interpreted logical forms, where the elementary propositions are those that are not truth-functions of other propositions, then many facts about the world cannot be described by such a list.

Consider a concrete example: let a local world be given by some family-tree. Then the corresponding propositions describe all the family relationships ie truth-functions of the parent-child relation, together with male, female gender predicates. Let the elementary propositions be expressed therefore, by sentences of the form: *a is a parent of b, a is male, b is female*. Then no set of such sentences can describe the family completely, for consider a male parent John whose only children are James and William, then the world description would contain the sentences *John is a parent of James, John is a parent of William, John is male, James is male, William is male*. Now our inferences about this family cannot be given by these sentences alone because we are also entitled to infer in this world that William has no sisters, *but this fact is not deducible from any of these sentences, or even all of them*. It is a universal proposition depending on John having no daughters, but nothing in these sentences tells us that John has no daughters, only that, if he has daughters, then they are not listed here. In short, we have no way of knowing that the list is complete, that it describes the totality of facts.

However epistemically, if we consider William's knowledge, we do find it is complete, since he clearly knows he has no sisters. Thus William's knowledge is not adequately expressed by the list of sentences. We might say William's knowledge more resembles a

mental model, than a mental list of sentences⁴. Nor could the linguistic deficiency be remedied by adding (elementary) sentences to the list, for none of these can tell us that the list is complete, ie that what is missing from the list does not exist. We seem to have identified a deficiency in the descriptive power of formal languages, due to each simple singular term in the sentences referring independently of the others. However, like a mental model, it seems clear also that Wittgenstein at T4.26 fully intended his elementary propositions to refer to every object in the world, and not to succumb to this deficiency of formal languages.

Contrast the set of sentences with the case of the family-tree. The family-tree correctly allows us to infer, not only that James and William are both sons of John, but also that John has no other children. For what is absent from the tree is treated as nonexistent, and since in the tree daughters are absent from John. In this (closed world logic) we can validly infer from the tree that John has no daughters. Epistemically the tree also adequately expresses William's knowledge, where the list of sentences does not. William really knows the totality of facts about this limited world, and the tree meets a local version of Wittgenstein's requirement that 'a complete description of the world' (T4.26) should entail 'everything that is the case' (T1) and that it should assert 'the totality of facts' (T1.1). But it seems to depend on the representation being a model rather than a linguistic description. In other words, elementary logical forms in the Frege/Russell notation are unable fully to express William's knowledge. It does indeed seem as if Wittgenstein was describing the logic of what we would now call a closed world.

⁴ Ie William's knowledge is not given by any deductive theory that uses only these sentences as its axioms.

We see that, apart from the Tractatus requirement at 4.26 that the truth/falsity of all possible such sentences should be listed, there was also the requirement that the description be *known to be complete*, ie refer to all the objects in the world. However, the description need not be a formal structure such as a tree, in order to achieve this. For consider another example: the register for a class at a school. The assertion, deduced from the register, that every child is present, is not simply inferred from the equivalent list *Anne is present, Bob is present, ...* etc for this is only the case if the register lists all the children, and nothing in the register, viewed as a list of sentences, tells us this is so. So the register can only be interpreted as a set of sentences if these are understood under the additional semantic condition that their totality refers to every child in the class, ie that the sentences are jointly to be treated *as a model* or picture of the world. I shall show that what really satisfies Wittgenstein's requirement is something that is neither a formal model nor a set of formal sentences, but a propositional sign that subtly combines the two namely, a formal file or database.

A formal model like the family-tree employs a one-one correspondence between simple elements (nodes) in the world-description and objects in the world. Formal databases however, seem to be models which also *express propositions* –by allowing their logical structure to be provided by formal symbols alone. Indeed, Wittgenstein said at T3.01 that the totality of true thoughts (ie the world-description) must itself be a picture (or model) of the world. Yet we have seen that the requisite one-one correspondence with objects entails that the set of such thoughts cannot be expressed just by interpreting the syntax of

Fregean sentences. Rather, the formal structure of a picture or model might lead us to a new syntax for thoughts, namely as formal data in a file.

Only with the additional rule that every child in the class is referred to by its name in some sentence (domain-closure) does the set of sentences in the register entail ‘everything that is the case’, as I understand Wittgenstein to have required, since only then can we be sure there is a one-one correspondence set up between the children and the names. But this rule is not a linguistic rule, it is not a condition for the truth of individual sentences, rather it is a condition for a world description to be complete. It means that when all the individual elementary sentences (axioms) are listed according to Wittgenstein’s requirements at T4.26, an additional condition for a world-description (and hence for T4.26 itself) to be true is that to every object in the domain there must be a simple singular term in some sentence, which refers to that object. This condition evidently cannot be expressed in the object language, but is a rule for its interpretation. However, in a database the same condition *is expressed by a syntactic part of the object-notation itself*, viz: by the boundaries of the files.

By referring at T4.26 to ‘all elementary propositions’ Wittgenstein was requiring that every object has been named; in doing this he was making an additional stipulation that was not an ordinary truth-condition of the particular language of logical forms due to Frege and Russell, and used by the Tractatus. However, without this assumption of domain-closure, such a putative world-description cannot entail any universal generalisations since it is possible for the domain to contain unnamed objects which do

not fit the generalisation. It is because of this that from singular propositions, universal propositions cannot normally be deductively inferred.

Wittgenstein never defined elementary propositions exactly⁵, but domain-closure does appear to have the significant consequence that no set of Fregean formulae is capable, at least on the standard interpretation, of fully expressing what he meant by elementary propositions. It seems, to make T4.26 true, that a suitable formalism for elementary propositions would need to incorporate some version of domain-closure into its standard truth-conditions. However, not only in the two examples above but in files generally, the data do normally meet this requirement, for this is the meaning of enclosing the data within a tabular boundary.

3. Arithmetic and Logic

As well as universal propositions, ‘Everything that is the case’ also includes numerical propositions, and so T4.26 would seem to also require numerical propositions about the world to be entailed by the elementary propositions. However, we shall see that this is not possible for Fregean sentences, because numerical propositions also assume domain closure. Thus, when counting objects, e.g. apple-1, apple-2, ... we naturally assume that when we finish counting we will have described the totality of apples. But because of the possible lack of domain-closure, this assumption cannot be made about any Fregean world-description, and numerical quantity cannot be deduced from elementary sentences alone. However, this is not a problem for files, and I show how, in this case, numerical

⁵ Although at T4.24 he permits them Fregean expression, apparently using variables x, y, z , as names.

quantity can be deduced from elementary propositions, if elementary propositions are viewed as data in a file.

In the absence of domain-closure for sentences, the Tractatus [T4.1272], instead of naming every object, resorted to bound variables. However, while still accepting that arithmetic is fully expressible in Logic, Wittgenstein came to feel nevertheless [1932]⁶ that this use of bound variables misrepresents our use of numerals (although not their truth-conditions). As a proposed solution, I develop the view that we may regard files of data as propositional signs referring by the same principle of one-one correspondence as used in counting objects. By then deriving numerical statements more naturally from such ‘elementary propositions’, we aim to also explain Wittgenstein’s misgivings about the use of quantifiers to express arithmetic by means of Fregean logic.

To this end, we first examine Wittgenstein’s early account in the Tractatus endorsing a quantificational definition of number. At T4.1272 he expressed the view that ‘There are two objects which ...’ is expressed by ‘ $(\exists x,y) \dots$ ’. The full formula may be found in [1932]⁷, where however Wittgenstein now says of Frege’s functional notation that it is queer ‘that we never use it when we are asked to reckon how many apples we have’. Rather, we prefer to use the familiar language of one-one correspondence with numerals.

In nevertheless accepting the unity of arithmetic and logic, it seems from this comment that Wittgenstein still required the number of apples to be a *logical* consequence of which

⁶ Wittgenstein’s Lectures Cambridge 1932-1935 Ed Ambrose

⁷ Ibid

particular objects are apples, but that this should happen in a natural way, not involving complex quantifiers. One way of examining the apparent tension between the relative expressive powers for the number of apples of arithmetic and of Fregean logic, might be to ask whether, given a world-description, say of particular children in a class, it is possible to deduce from the world-description how many children we have⁸, purely by logical inference.

This can be seen to be similar to the previous problem about the totality of facts, since the number of children also, on the Tractatus analysis, is found to depend upon an implied universal quantification whose inference from the class-register would assume all the children have been named. To see this, consider the following example, referring to a domain of apples. The statement that exactly three apples are green would be written⁹,
 $(E3x)G(x)=$

$$(1)\dots \exists x1,x2,x3(Gx1 \& Gx2 \& Gx3) \ \& \ \sim \exists x1,x2,x3,x4(Gx1 \& Gx2 \& Gx3 \& Gx4)$$

where the second clause says that in every quadruple of apples, at least one of the apples is not green. Wittgenstein implies the main reason we would not write this ‘to reckon how many apples we have’ is because it misrepresents how we think. He also notes that it still employs the principle of correlation or one-one correspondence used in arithmetic. As Wittgenstein says rather shortly, this expression of arithmetic in Fregean notation requires us to correlate the variables of one clause in the formula with the variables in the other clause, just as we do in counting objects; a circular definition. We appear to have reduced arithmetic only to another version of itself.

⁸ Recalling his earlier belief at T4.26

⁹ Here, to avoid using an identity predicate, Wittgenstein requires different variables to denote different objects.

I shall try to show that the unity between arithmetic and logic is exhibited, not by attempting to define arithmetic in a logical notation foreign to it, but rather by altering the notation of logic to a form (files of data) more congenial to arithmetic, viz: to one employing one-one correspondence itself as a mode of reference in Logic.

Firstly, to establish that (1) cannot be inferred from an elementary world description in the Frege/Russell calculus, consider some such description, as perhaps envisaged at T4.26, given by the set S of elementary formulae,

$\{G(a), G(b), G(c), \sim R(a), \sim R(b), \sim R(c), R(d), R(e), \sim G(d), \sim G(e)\}$. Then S can describe any domain $\{a, b, c, d, e, \dots\}$ containing at least three green and two red, apples.

Therefore, the above formula (1) stating that there are exactly three green apples cannot be deduced from this set of sentences, because, although from $G(a)$, one can infer $\exists x1G(x1)$, [and likewise for b, c and $x2, x3$], there is no way to infer the $\sim\exists$ clause in (1) since, in the absence of domain-closure, there may well be some other x in the domain of apples which is not named in the set S, and which is green. Hence we cannot logically infer the number of green apples from such a set of elementary sentences about apples. In other words, the Fregean notation does not allow numerical facts about the world to be logically deduced from *elementary* descriptions of the world alone.

However, the fact that this consequence of elementary world-descriptions is counterintuitive seems to show we do tacitly wish¹⁰ to assume domain-closure for them. Since linguistic sentences (axioms) need not satisfy this assumption, we might be led to

¹⁰ As Wittgenstein himself revealed at T3.01

seek a notation that does, so that our world-description really does entail ‘everything that is the case’, including, in particular, all numerical facts. Moreover, if we could indeed deduce Number from our logically simple propositions, then our definition of number might then be more intuitive than (1). In particular, it may eliminate the awkwardness in (1) of requiring certain objects not to exist.

Such a symbolism for propositions may not be inconsistent with Wittgenstein’s belief at this time [1932], since it seems clear that this belief was not a commitment to any single notation. Rather, it was the general concern that numerical truths must result from the fundamental nature of the proposition, and that a correct symbolism for propositions would reveal this in a natural way. His problem however, was that the current symbolism for propositions appeared not to do this. For, in the terminology of the *Tractatus*, if propositions took a purely linguistic or more specifically, Fregean, form, then the number of objects of a certain kind did not follow from the ‘elementary’ propositions about objects of that kind.¹¹

I plan to show that if elementary propositions are reconstrued as data in a file, then numerical propositions do become deducible from elementary propositions in a natural way, thus facilitating the expression of Arithmetic in Logic. This is because the data in such a file will be found to perform the same kind of referring that we perform when counting.

¹¹ This can be compensated if we allow other formulae in our world-description apart from elementary propositions viz: an axiom ensuring domain-closure. Such an axiom is not elementary as it is universally quantified.

4. Anonymous objects

Consider counting five apples: apple-1, apple-2, ..., apple-5. Here, 'the second apple', or 'apple-2' is certainly a singular term, but it is not a definite description of the apple, nor is it a name¹², but rather it is just a temporary designation by counting. It is clear that if the counting is ad hoc, not following any method, then it may later be impossible to infer, from the numeral alone, the identity of the apple; the numerals may thus be said to refer *anonymously*.

However, by the anonymity of numerals, we do not mean the kind of anonymous reference performed by free or bound variables, since each numeral refers to a particular individual. Admittedly, a variable $x1$ does not identify an object either, and so is anonymous, but neither does it refer only to *one particular object*: it is an indeterminate and can refer to any object in the domain. Thus if apples a, b, c are green, then even if $G(x2)$ is true of apple b, it is equally the case that $G(x2)$ is true of apples a, c also and, unlike the applied numerals 'apple-1', 'apple-2', 'apple-3', ... , there is no particular apple exclusively assigned to $x2$. Rather, we may say the applied numerals are treated as *constants*; thus $G(\text{apple-2})$ is a sentence, having a unique truth-value. Nevertheless these constants refer anonymously, because the sign *apple-2* may alone be insufficient to identify its referent, (unless there was some method to the counting).

Thus, to say that apple-2 is green is to say simply that the apple counted as 'apple-2' is green. This is a definite description referring not only to apples, but also to the reference

¹² It is not a rigid designator since, counterfactually, the apples could have been counted in a different order.

(by counting) of a sign; and it is this that constitutes the anonymity of numerals. Hence we see that such a proposition is not expressible in any formal object language, but requires a metalanguage. It seems that a language must, to enumerate objects, contain constants ('apple-2') which refer reflexively to 'the referent of this sign' *under some given referring relation*, ie implicitly refer to themselves and the relation, as well as to objects. This says, they must be constants that refer anonymously.

Thus, applied arithmetic appears to use a novel kind of reference, *counting*, not available to formal object languages. In order to formalise the special propositions of enumeration¹³ therefore, we need a formal notation that yields both the one-one reference and the anonymity, of numerals. It would be ideal for this purpose to abstract out the numerical meaning¹⁴ from the set of singular terms {'apple-1', ..., 'apple-5'}, leaving only their anonymous correspondence with objects. Unfortunately this is not possible, since the numerals refer only as part of the very process of counting.

However, there is an ancient non-numerical symbolism that that refers to objects in exactly the same way as numerals, namely



This is the simple symbolism of keeping a tally, whereby an identical dot or mark is made for each new object being tallied, without enumerating them. The dot signifies a particular object but, like a numeral, does not reveal its identity. This surely must be the most ancient and primitive symbolism for depicting *objects in general*. Nevertheless, the

¹³ Such as 'apple-2 is green'

¹⁴ Ie their ordinal or cardinal meaning.

tally operates according to its own strict rules, for it means the dots are to be understood as being in exact one-one correspondence with *all* the objects in the domain, without identifying any of these objects. If this is not the case, then the tally was kept improperly. It is not possible for the individual dots (like symbols) to identify their referents by their own distinctive shapes, as the dots are all indistinguishable. It might be possible for them to identify their referents by their position, but equally these positions may be random and have no referential significance. Thus the dots or tally-marks refer, but they do so anonymously.

If the dots are used to keep a tally of apples, it seems they perform the same purely referential function as do the ‘anonymous constants’ ‘apple-1’, ‘apple-2’, ‘apple-3’, ... but without the function of ascribing numerical order or quantity. Moreover, they do this without being symbols in any language. We now consider whether these dots could nevertheless form the basis of a logical notation. Although these dots are not as yet singular terms in a propositional sign, and are so far only elements in a tally¹⁵, let us first review the mode of reference by such elements, as compared with the mode of reference of variables in logic, where the domain comprises exactly the objects being tallied.

1. *Anonymity*. Like a variable, each dot refers to a unique unspecified object but, unlike a variable, the dot stands for one *particular* object and, once used, the dot cannot be used to refer to any other object.

¹⁵ The tally is not even a model, since no predicates or relationships between the dots is presented, their sequence may have no significance.

2. *One-one*. Further, unlike the normal Tarskian variables, different dots must refer to different objects.
3. *Domain-closure*. And finally, like a bound variable, the dots jointly refer to all the objects in the domain.

In other words, the mapping from dots to objects is an anonymous one-one function satisfying domain-closure. This is what defines an overall one-one correspondence between the dots and the domain. But the big contrast with variables, of course, is that each dot is unrepeatable and so cannot be treated as a symbol at all. Nevertheless, each of the three conditions remains true of certain symbols, eg if for ‘dot’ we substitute ‘applied numeral’ or ‘apple-2’. By contrast, a Tarski assignment of objects to free variables is a *many-one* function from variables into the domain [infringing 2, 3].

Curiously, these three elementary properties seem to counter some of the complaints or discomforts experienced by Wittgenstein with Fregean notation. [For example his modification of variables to one-one reference, while intended purely to avoid using an identity predicate, also makes them conform to (2), and hence more like tally-marks].

In summary, the dot refers to a particular but unidentified object¹⁶, and has the kind of reference we perform when counting a set of objects. Here we keep track of each object counted, ensuring (3) that every object is counted, and (2) that we do not count it again [but (1) we need not identify it as an individual]. However, a tally is more basic than

¹⁶ The objects referred to by dots somewhat resemble the medieval scholastic notion of substances as *bare particulars*. This was the metaphysical idea of objects possessing both individuality (essence) and existence (substance), but which could not be identified by any (or all) of their properties.

counting since I can keep a tally of say, passing cars, or days in prison, without attaching any numerals to individual cars or days, ie without recording their order or quantity.

Tallying is always referential, but not necessarily numerical.

5. Data as non-linguistic propositions

Now the way to make these dots not merely referential, but propositional, is to add predicates to our referring-terms. This can be done by the simple expedient of hollowing out our dots to provide empty cells:



And then, by inserting in each cell a symbol to signify a predicate, we can generate a list of propositions such as the following,

G G R G R

to describe green and red apples. The cells themselves have now been made invisible, and need to be imagined; each cell is now a *place* in which exactly one symbol may be written, and is what we formerly regarded as a tally-mark. Thus, for our new kind of singular term, the empty cell, I shall argue that we also have a new kind of logically simple proposition.

This tally of apples might be called a *file*, and the elementary propositional signs of which it is composed, might be called *data*. Such a file naturally inherits for its cells, and

hence for its data, the three referential properties just listed for tallies, defining anonymous one-one correspondence. Unlike a pure model (or picture), a file of data is interpreted *assertorically*, ie it expresses propositions. This is because the data do appear to combine singular and general terms, but in a non-linguistic manner.

However, the reference of the singular terms is known only transiently (while tallying), and so such a file is not a permanent record, unless symbols are added to identify the apples, ie they are no longer anonymous.

While the singular terms are physical tokens (unrepeatable cells), as required for arithmetic, the predicates or general terms are abstract types (symbols), as required by language. Thus a file of data may be regarded referentially as a model and predicatively as linguistic (since the predicates in the file merely denote but do not *resemble* properties of the objects). However, as in a model, the mere absence of a predicate from a cell in which it is eligible is sufficient to signify its negation¹⁷ for that object. Apparently unique therefore, among modes of representation, a *file* seems to be a hybrid both of a physical model, and of a linguistic description, combining advantages of both. We may call it, an assertoric model¹⁸.

Of course, the very limited kind of file presented here can express only a restricted range of elementary propositions, viz: those referring anonymously, and ascribing only single

¹⁷ Database theorists call this property of files: the Closed World Assumption.

¹⁸ However, a Relational Model used as a sign (comprising lists instead of sets) cannot count as a file, because its *singular* terms are, not physical objects (cells), but abstract symbols (names) and hence, while they achieve domain-closure, they fail to refer anonymously. Thus no relational model can express the meaning of the file GGRGR.

monadic predicates such as red, green that are mutually exclusive. This is because each singular term is not a symbol but a physical *place* where only one symbol can be written, so no two such predicates can be ascribed to the same object at the same time¹⁹.

Nevertheless, even such simple tallies can be useful as ‘world’ descriptions, as we shall see.

It may be noted that this syntactic convention is the exact opposite of Frege’s convention of saturated and unsaturated signs, for here it is the singular term that is unsaturated, and the general term that completes it. This is because the cell is interpreted here not as an argument-place, but as an object²⁰, and is consistent with Aristotle’s talk of properties ‘inhering’ in substances. Here, the logical relation of subject to predicate meant by ‘inhering’, or ‘informing’, can no longer be regarded with Frege and Russell as linguistic. Predication, in this language/model hybrid, is now better expressed as *the occupancy of a tabular cell by a symbol*.

Moreover, all the singular terms are united in a totality, reminiscent of scholastic talk viewing substance as a plastic medium, in which no two impressions can be made in the same place. Notice that the notation also accommodates Wittgenstein’s rejection of an identity predicate [T5.531] since every cell refers to a different object, and so two cells cannot be identified with each other.

¹⁹ However, this severe limitation can ultimately be extended to the full generality of a database admitting each object into all possible properties and relations.

²⁰ However, in complex relational files, some cells do express Fregean argument-places.

Such elementary propositions are genuinely new in the sense of being inexpressible in Fregean logical form, thus granting at least some of the extra flexibility that the later Wittgenstein wanted from language-games. For, not only does domain-closure prevent the above file of data being translated into elementary formulae $F(a)$ in Frege's notation, it cannot be translated into any formulae whatsoever in a formalised *object*-language, since the truth-conditions of individual anonymity can only be provided by naming cells and the reference relation itself, as well as the objects to which they refer²¹. Data should be regarded therefore, as expressing propositions that are, in formal terms, strictly non-linguistic, and therefore not expressible in classical logic. The logical form of such propositions is provided instead by formal symbols occupying cells of records in an abstract file.

If cells do indeed provide a new formal mode of reference, not possible in the current symbolism of logic, then it would seem that Wittgenstein's claim,

T4.1272 'Wherever the word 'object' ('thing' etc) is correctly used, it is expressed in conceptual notation by a variable name.'

is no longer quite true. Similarly, for Quine's dictum that 'To be is to be the value of a bound variable'. For we have found that the word 'object' can also be expressed by the *cells in a file*, and that this is not logically equivalent to the use of variables. For, while neither cells nor variables identify their referents, a cell has only one such referent and, like the referent of a bound variable, this may also be described as an 'object'. Thus, speaking of a non-empty file, one may indeed say eg 'There are objects', contrary to T4.1272. Apart from formal language, there are clearly also *non-linguistic* ways of

²¹ But they can, like all truth-conditions, be defined in an informal metalanguage.

making precise reference to objects, eg by systematic record-keeping, using files. It appears that data in cells conform to a non-linguistic logic, enabling us to describe and refer to particular anonymous objects in a new way. And so, perhaps sometimes, ‘whereof one cannot speak’ (eg ‘objects’), thereof one may be able, nevertheless, to meaningfully tabulate.

6. Wittgenstein’s shopkeeper

We now reconsider how this non-classical symbolism might help us express numerical quantity and thereby assist the expression of Arithmetic by pure Logic. In his later work, Wittgenstein disregarded his earlier quantificational definition of Number, and instead examined the logical nature of enumeration. At §1 in *Philosophical Investigations (PI)*, Wittgenstein presents an example where the objects are again apples, in which a shopkeeper receives the request “five red apples” from a customer. He responds by selecting red apples one at a time from a drawer and pronouncing a numeral for each one, until ‘five’ is reached.

In this language-game the apples are never mentioned individually by name; rather, the shopkeeper just counts out apples without identifying them individually. Anonymous reference is presupposed in this language-game by shopkeeper and customer alike. What is this anonymous reference? As each apple is counted, both parties understand the numeral to refer to its apple by, as previously discussed, being a constant, not by being a variable. If the numerals were being used as bound variables: *apple-1*, *apple-2*, ... then the customer and shopkeeper would be alike agreeing about the non-existence of any

apple-6, in accordance with the quantificational definition (1). But clearly, no such agreement takes place, since no such ‘apple-6’ is ever mentioned by either party. So the linguistic practice expressed by this enumeration cannot adequately be described in Fregean notation.

However, even if the process of enumeration lacks a connection with the notation of Fregean logic, it does have an intimate connection with the ‘logical’ notation of data in a file. For, suppose the shopkeeper keeps a record of his stock of apples by means of the above file GGRGR. Every time he receives an apple its colour and the fact of its existence are recorded by adding ‘R’ or ‘G’ to the file, and every time he sells an apple an ‘R’ or ‘G’ is deleted appropriately. If his stock is currently given by the above file GGRGR then, without inspecting the apples themselves, he is able to *logically infer* from this set of propositions that he has insufficient apples to meet the order ‘five red apples’. Instead of opening the drawer, the shopkeeper can perform upon the symbols ‘R’ in the list, exactly the procedure with cardinal numbers that Wittgenstein describes for the apples themselves, ie he can count them²².

We see that if the list of elementary propositions is a file, ie constitutes an assertoric world-model, then *counting* may be considered a form of logical inference from those propositions. From his file he can deduce the proposition ‘There are fewer than five red apples’. This would not be the case if the elementary propositions were presented as a set

²² A computerised system for a supermarket works in essentially the same way, updating an electronically accessible stock file in response to goods scanned at the checkout. Thus, as well as their metaphysical significance for the foundations of Logic, systems of files are used to formalise much of our modern everyday knowledge.

of axioms even if the apples were individually named, since classical logic always allows for the possible existence of unnamed red apples. Nevertheless, there is an interaction between the linguistic and non-linguistic notations, with Fregean formulae such as $(\exists x)G(x)$ being derivable as truth-functions from the data in the above file. While neither notation is fully translatable into the other, the linguistic and non-linguistic formats are complementary, as exhibited by a database and its query language.

A reason for this interaction is that, for a finite domain, instead of defining existence only by the use of an existential quantifier, we are now assuming that *to be, is to be recorded in a file*, and *to be destroyed is to be deleted from a file*, where by a file, we mean a list of occupied cells referring to all the objects in the domain, ie a list of elementary propositions exhibiting domain-closure. In this sense an empty file containing no records would be a representation of the metaphysical idea of nothing existing. [PI §55].

Similarly, an occupied file means *everything* that exists, and a single record means a unique *something* that exists different from the other somethings (records) that exist.

Finally, the metaphysical idea of a bare particular, an object with no properties, would be expressed anonymously by a tally-mark, ie an empty record within a file.

It is all of this that enables quantified formulae of classical logic to be inferred from files, an equivalence between tabulation and language universally presupposed by the implementation of database systems. Moreover, properties of the world as a whole, held by the early Wittgenstein to be inexpressible as propositions, are now seen to be

properties of the filing system itself, independent of the data. These tabular properties yield a new logic for the propositions of any closed world.

One part of this non-classical logic is that a proposition is expressed by a cell only if it is occupied and so, as with quantifiers, a predicate is still required for the data to assert an existential proposition. It is just that here it is the subject that is unsaturated, instead of the predicate²³.

7. Deducing Arithmetic

The foregoing logical analysis of enumeration can now help improve the expression of Arithmetic as Logic. By regarding a file as a new kind of propositional sign, then the defining of number by logic (T4.1272) reconsidered by Wittgenstein in 1932 as seeming unnatural, can now be undertaken in a more natural way. Take GGG as a simplest possible world-description, and suppose it fully describes a world W of three green apples a, b, c respectively. Firstly, we derive Wittgenstein's earlier definition of 'three green apples' from this file. We can easily see by assigning a to $x1$, b to $x2$, c to $x3$, that in this case, Wittgenstein's first clause

$$(2)\text{-----} \quad (\exists x1,x2,x3) (G(x1) \& G(x2) \& G(x3))$$

is true in W. Next, since language also allows for the existence of unmentioned green apples, such a possibility had to be eliminated by Wittgenstein's second clause. This

²³ A simple tally of dots is not an affirmation of existence, since it is not an affirmation at all. Recall that we considered a tally of dots or strokes to be qualitatively different from propositions, viz: a simple list of *objects*. Similarly for a tally of empty records in a file.

clause was effectively needed to convert the ordinary linguistic reference of $x1, x2, x3$ into one-one correspondence.

However in the case of files, GGG itself (unlike formula (2)) is able, because of *domain closure*, to say there is no d in W , additional to a, b , and c . Hence there is no Tarski assignment of a to $x1$, b to $x2$, c to $x3$ and d to $x4$, such that

(3)----- $G(x1) \& G(x2) \& G(x3) \& G(x4)$ is true in W at $[a,b,c,d]$

Consequently, Wittgenstein's second clause,

(4)----- $(\sim \exists x1,x2,x3,x4)(G(x1) \& G(x2) \& G(x3) \& G(x4))$

also follows from GGG. And from (2), (4), $(E3x)G(x)$ follows by his (1932) definition (1). In this way, we see that $(E3x)G(x)$ can be inferred from the (anonymous) elementary propositions GGG and not just from quantified clauses.

Thus, both the clauses (2), (4) follow logically from the one-one correspondence (implying domain-closure) used by GGG, *as a mode of reference*. Finally, this proof of $(E3x)G(x)$ from GGG now justifies a rule for inferring $(E//x)G(x)$ *directly* from²⁴ the file GGG, viz: by *counting*. In doing this we simply employ the fact that the unary digits of the numeral correspond one-one with those cells of the file which contain the predicate G . By thus proving that the linguistic complexity of (1) is equivalent to humble counting in a file, we avoid the tortuous route of having to infer the existence of triples (2) and the nonexistence of quadruples (4).

²⁴ This is the unary notation used by Wittgenstein in 1932 (*Ibid*)

An argument therefore, in favour of GGG now being considered a *logical form*, is surely that by matching it with unary digits we are now able to deduce number from logic in a natural way, as required by Wittgenstein in 1932. This not only formalizes Wittgenstein’s later example of a shopkeeper, but is effectively the form of numerical inference used by modern databases²⁵ when counting the entities in a file. The unnaturalness of the earlier Wittgenstein’s second clause (4) was due to the absence of domain-closure in classical logic, and the resulting attempt to define one-one correspondence in terms purely of linguistic reference by variables. For, even when Wittgenstein makes these variables one-one by special decree [1932], the first clause (2) cannot alone guarantee that it refers to everything in the domain.

8. More Typical Files

Finally, if we take our original anonymous file GGRGR and eliminate the anonymity by adding symbols to identify the objects referred to by the cells, we obtain a table, thus

<i>a</i>	<i>b</i>	<i>d</i>	<i>c</i>	<i>e</i>
G	G	R	G	R

Here the surrounding box sets a boundary to the universe of discourse (The ‘World’), subdivided into cells, which are now rendered visible once again. Strictly, we must now say that an object is referred to, not by a cell, but by a pair of cells, a column, containing two kinds of data-value: a predicate, and a singular term. Each column is called a *record*, the record of a particular object. Thus what started out just as a simple tally-mark has

²⁵ They do not infer quantity by separately inferring clauses (2), (4) from the data. While this is of course theoretically possible, it is computationally inefficient.

gradually become elaborated into a record in a file. Furthermore, by adding more rows to the table we permit ourselves to record more predicates for each object.

The above table expresses not only all the propositions listed in the set S: $G(a), R(d), \dots$ including $\sim R(a), \sim G(d), \dots$ but also generalisations such as $\forall x(G(x) \vee R(x))$ and numerical statements like $E\exists xG(x)$. Thus the table expresses strictly more content than the Fregean formulae $G(a), R(d), \dots$ and their negations, and so does more than merely express Fregean logical form. Rather it provides *a new kind of logical form* for elementary propositions, satisfying Tractatus 4.26, and founded on tabulation instead of language. If propositions are what we know or believe, then examples such as the family tree, the class-register and the shopkeeper show that often the facts that we believe about the world are seen as a bounded whole or system. This means they are best expressed by *data*, ie better formalised by occupied cells of a file than by sentences of a formal language. For even our simplest knowledge of any world can include universal and numerical propositions about it.

The foregoing analysis of elementary propositions as non-linguistic data has of course been confined only to the simplest of files. However, the versatility and pervasiveness of databases suggests that it can be extended to all human knowledge of closed systems. Unlike the elementary formulae of the Frege/Russell formalism, the data contained in a file meet the exact requirements of our informal knowledge for any closed world, including the requirements of modern information-processing. What has been shown here

is that such a file of data is capable of entailing, at least for a finite domain and monadic predicates, everything that is the case, ie The World.