# NATURAL LOGIC

Michael Sussna

August 1990

# Introduction

Natural logic is the result of reviewing logic for naturalness, and of renewing logic by revising and extending it where such changes make sense. Logic as a subject need not be something strictly artificial, divorced from everyday life. Rather, it can be treated as an integral part of all human thought whether or not we understand how it works. Faithfully describing logic in an intuitive way can aid both human thought and artificial intelligence.

# The Fallacy of TFTT

One starting point for seeing what natural logic contains is the traditional treatment of implication. Traditionally, given two statements p and q, and a third statement of the form "p implies q," the truth or falsity of the third statement is determined solely from the truth or falsity of p and q. The only time that "p implies q" (written $p \supset q$) is false is when p is true and q is false. The rest of the time, $p \supset q$ is true. This well-established treatment is not intuitive. It can be considered a purely syntactic approach. In those other three cases, how can we be sure that p implies q when we don't know anything about p or q except their truth or falsity? Surely what p and q *say* must matter. Traditional explanations (rationalizations) shrug off this counterintuitive interpretation as acceptable, since formal logic is simply mathematical and is not intended to be perfectly intuitive.

Can we reconcile this counterintuitive traditional interpretation with what is intuitive and yet preserve at least some of the power, elegance, and usability of logic? Yes. It actually does make sense that given a true p and a false q, we can conclude that $p \supset q$ is false without knowing anything further about p or q. Just think about what is being said. "P implies q" means q follows, given p. (Exactly what "follows" means will be discussed in due time.) How can this statement be true for a specific p and q where p is true but q is not? Thus, in this one case out of the four possible cases where we look at all combinations of the truth values of p and q, the overall traditional conclusion makes sense. When we try to think about the other three cases it is not as easy to decide the overall outcome when we only look at the truth values of p and q. In fact it cannot be decided at all! However, the traditional approach assigns outcomes anyway: in all three cases the truth value of $p \supset q$ is decreed to be "true."

There is an explanation for this result. By only allowing for two truth values, "true" and "false," we are boxed into concluding that if a statement isn't false then it must be true. This is exactly what is happening with those three other cases of implication. In the one case (p true and q false) where we can safely and sensibly figure out the result knowing just the truth values of p and q, "p implies q" is false. In the other three cases we cannot say that $p \supset q$ is false. So, traditionally this has meant that the result in those three cases is $p \supset q$ is true. Traditional logic adheres to the "Law of the Excluded Middle," which decrees that there shall be no intermediate truth value between "true" and "false." A statement is either true or false but cannot be both.

In order to make the traditional approach to implication more intuitive we need to begin by repealing the Law of the Excluded Middle. Much of the substance of natural logic has to do with exploring the ramifications of allowing for more than two truth values. The quick answer to "Then what *is* the truth value of $p \supset q$ in those other three cases?" is "indeterminate" for now. As we shall see in time, there is more to it than that. First we need to develop the machinery with which to talk and think about these issues in a natural way. We will eventually return to the topic of implication and see what the natural logic treatment of it produces, including a truth table for implication.

# *Beyond T and F*

It would seem from our discussion of implication that there may be motivation for looking into the possibility of there being truth values besides T and F. We intimated that "indeterminate" or "I" might be another truth value. Just what would an extended set of truth values contain? Certainly "I" makes sense. Often the truth or falsity of a statement is unknown given current information. As one can see, time and state enter into things. Perhaps it would be more accurate to refer to the truth *state*, rather than the truth *value*, as indeterminate. We can say that the truth state of a statement is either determinate or indeterminate. Within the determinate truth state, we can have truth values of true and false. Are other truth values possible? Does the notion of truth value apply to the indeterminate truth state?

We often encounter statements that can be true sometimes and false at other times. There are several ways in which such statements arise. Such statements are not to be confused with statements of indeterminate truth state -- here we are referring to statements where we have enough information to decide their truth or falsity, it's just that sometimes they're true and sometimes they're not. Let us call them "only sometimes true" or "only partly true." We will distinguish three truth values for statements with determinate truth value: entirely true, only partly true, and not at all true. We can use symbols for these as follows.

Let us distinguish three "qualifiers" which together exhaust all possible values in a range of values and which also are mutually exclusive. They thus "partition" the range of values. The three qualifiers are: "entirely" or "all," "only partly" or "some but not all," and "not at all" or "none." These are handy items useful in many contexts where some expressiveness is desired and yet full detail is unnecessary. We will borrow the "for all" and "there exists" quantifier symbols from traditional logic so that we symbolize "entirely" and "all" as $\forall$, and we will symbolize "only partly" and "some but not all" as $\exists$. Strictly speaking, we should distinguish "inclusive some" from "exclusive some"; that is, "some or all" versus "some but not all," respectively, just as we distinguish "inclusive or" from "exclusive or." This disambiguation is important when we consider that many of the acknowledged logical fallacies are caused by ambiguity. Thus, we shall keep ready the symbols $\exists_{in}$ and $\exists_{ex}$. So, our second qualifier, $\exists$, is actually $\exists_{ex}$. The third qualifier is symbolized as 0 (zero).

Now we are ready to symbolize the three truth values. "Entirely true" is $T_{\forall}$, "only partly true" is $T_{\exists}$, and "not at all true" is $T_0$. Of course the perfect symmetry here allows us to equally validly say that "entirely true" is $F_0$, "only partly true" is $F_{\exists}$, and "not at all true" is $F_{\forall}$. These last three expressions can alternatively be stated as "not at all false," "only partly false," and "entirely false" respectively.

Well, then, in what ways can statements arise whose truth value is $T_{\exists}$? Before we answer that question, we need a bit more machinery.

Let us consider that statements in general state a relationship, describing either an event in the physical world or a connection in the logical world. For example, "The jet streaked across the sky" describes a physical event, while "Squares have four sides" describes a logical connection. In general, it can be argued, statements concern the relationship between two things. (One might disagree and present intransitive sentences as counterexamples. These can be considered to contain an implicit second party, "the world." For example, "The computer hummed" or "The woman dreamed" on the surface only deal with one participant, not two. Here, the implicit second party, the world, can be appended explicity although awkwardly to complete the pairing as in "The computer hummed (in the world)" and "The woman dreamed (in the world)." The point here is not to contrive sentences but to indicate that in some sense there really is a second participant implicit. One could also counter with complex sentence constructions in which three or more participants are involved. For example, "The boy gave the apple to the horse" or "Sam startled Bob and Ray with his laughter." Even here we can regard a statement as having two main participants related in some way. The participants may be composite (compound) or have structure in some other way, or the second participant may be implicit.) Statements can be characterized as having the form "x r y." That is, x is in relationship r with y. X corresponds to the subject of a sentence, r to the verb, and y to the object if any. In the previous two examples, the subjects are "the boy" and "Sam," the verbs (relations) are "to give an apple to" and "to startle with laughter," and the objects are "the horse" and "Bob and Ray" respectively. (By the way, these verb/relation phrases are treated in natural logic not just as verb phrases but as verbs, per se. Thus, multi-term verbs are recognized (as infinitives), which constitutes a departure from traditional linguistics. Awareness of such "superverbs" can help with our understanding of natural language phenomena, for example why it is sometimes alright to end a sentence with a preposition.)

So, let us consider statements to have the form x r y. Truth values arise from the interaction of statements and the subject matter that the statements discuss. If a statement accurately describes its subject matter, then it is said to be true. Conversely, if a statement misrepresents reality, then it is said to be false. Thus, a statement's truth value is the result of comparing the stated with the actual. The relationship between the stated and the actual need not be as simple as portraying it as either black or white with no shades of grey possible. When performing our comparison, or matching, between the stated and the actual, we can allow for at least three kinds of black-to-white ranges. Working with our statement format of "x r y," we can discuss

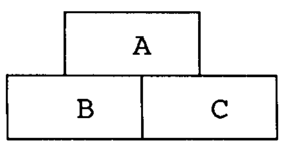
1) prevalence of match
2) degree of match
and 3) duration of match

as topics whose analysis will show us how statements can end up with a truth value of "only partly true."

## *Partial Truth*

Prevalence of match concerns the proportion of cases where an asserted relationship holds. For example, suppose we have the statement "Birds fly." We know that some birds do not fly, though most do. It is inaccurate to say that all birds fly, as it is inaccurate to say that no birds fly. The truth lies somewhere in between: some but not all birds fly. So this is an example of a range of cases (the x's in our "x r y" format), where the proportion of cases that the relationship is true of ranges from none to all. Thus, in our simple yet somewhat expressive approach, the prevalence of match ranges from 0 through $\exists$ to $\forall$. We can thus modify the x in "x r y" with a qualifier of $\forall$, $\exists$, or 0 if we want to be somewhat specific.

Degree of match has to do with how completely or incompletely the x's are in the relationship with the y's. For example, we could say "Zebras are black" or "Zebras are white." In both cases the statements are only partly true because zebras are only partly black and only partly white. To take another example, consider the statement "A is on B." If the following picture presents their true relationship, then the statement is only partly true:




We can thus apply a three-valued qualifier to the "r" in "x r y" as well.

Finally, duration of match involves the proportion of the time when a statement's asserted relationship holds. An example is "Jill's sick" versus "Jill's sick at the moment." To be symmetrical, we could qualify the "y" in "x r y" as we have done with the "x" and the "r" to place our time qualifier, but we need to be careful here. The time element of a statement does not relate tightly to the object of the sentence, but rather to the relationship as a whole. Thus, we place the time qualifier at the end of the statement, after the y rather than in front of the y. This way the symmetry is broken, we are reminded that this third qualifier modifies the whole statement and not the y specifically, and the statement reads naturally in English. We will see this illustrated momentarily.

We thus further characterize statements as "x r y" plus three proportion or degree indicators. We can use the somewhat suggestive notation $\%_c$, $\%_r$, and $\%_t$ for these qualifiers, for "proportion of the cases," "degree of relation," and "proportion of the time," respectively. Our statement format becomes $\%_c$ x $\%_r$ r y $\%_t$. This allows us to express such assertions as "Only some tennis racquets are entirely wooden these days." The x and r components and the $\%_c$ and $\%_r$ qualifiers are clear here, but what is y and what is the $\%_t$ qualifier? Y here is implicit (we have an intransitive sentence). The time qualifier is not one of our three discrete values, $\forall$, $\exists$, and 0, but a phrase filling in the time interval involved. This is not a problem -- our three discrete qualifier values are useful for summarizing and give minimal specificity -- there is no limit on how much more specific we can get. Sometimes it is enough to simply summarize the actual when comparing it with the stated. Quantifying or otherwise being more specific may not be necessary, possible, or safe. How could this be unsafe? We can sometimes over-quantify when our knowledge doesn't warrant such precision -- this can lead to erroneous assertion.

A final touch, for now, in elaborating a statement's format is to also allow specifying contexts for $\%_c$ and $\%_t$ as follows. We can specify the cases *under discussion* and the times *under discussion.* Then we specify the proportion of cases and the proportion of the time with respect to the cases and times under discussion, respectively. For example, we might say "All of the ball bearings *in this sample* meet our tolerances for sphericity" which illustrates using a proportion of cases in the context of a set of cases under discussion. Similarly with time, we can say "The ball bearings this month have usually met specs." Here, the proportion of the time is "usually" and the times under discussion are "this month."

The statement format now becomes $\%_c$/cud x $\%_r$ r y $\%_t$/tud, where "cud" and "tud" are "cases under discussion" and "times under discussion," respectively. As promised, the way that this statement format reads as natural English can be illustrated with the following example:

```
%c   ....  "Only some
cud  ....   of all
x    ....   mushrooms
     ....   are
%r   ....  entirely
r    ....   safe enough to be eaten by
y    ....   people
%t   ....  on all
tud  ....   those occasions when they see mushrooms."
```

Actually there is one further embellishment that we can add to our statement format at this point. That is to acknowledge a "truth value qualifier" at the beginning of the statement. Again, we can employ our three-valued qualifier for a bit of specificity. We can use % notation here as well, to reflect "degree of truth." Since we already have a $\%_t$, rather than call our new qualifier $\%_{tr}$ or something, we can go to a modified naming scheme. Let us rename $\%_c$ to $\%_1$, $\%_r$ to $\%_3$ (yes, $\%_3$, not $\%_2$), and $\%_t$ to $\%_2$. One reason for this switch between r and t is that we can easily think in terms of cases-and-times giving overall context and then "relation" being specified. This is also linguistically natural, as in "Only some mushrooms, on all those occasions when people see mushrooms, are entirely ripe enough to be eaten." A second reason for reordering our qualifiers concerns a fascinating result in natural logic, which we will see later, referred to as the negation theorem.

We shall call our truth value qualifier $\%_0$. For now let us say that it can take on the three values $T_\forall$, $T_\exists$, and $T_0$. Statements modified by this qualifier would begin with "It is entirely true that..." or "It is only partly true that..." or "It is not at all true that..." for example. Thus our statement now looks like this: $\%_0$ $\%_1$/cud x $\%_3$ r y $\%_2$/tud. We can abstract the four qualifiers out from the statement and list them as follows: $\%_0$ $\%_1$ $\%_2$ $\%_3$. This abstraction in no way undermines the roles of the x, r, and y components of a statement; it merely isolates some of the interesting features of a fairly well specified assertion. We can work with these isolated features alone for some purposes. Before we look at the negation theorem, and see what working with the abstraction can do, we need to investigate what natural logic has to say about negation.

## *Negation*

In natural logic we distinguish two kinds of negation. One is the "other than" or complement operation. The other is the "opposite of" operation. We need to make this distinction in multi-valued logic. In traditional two-valued logic, where the only truth value choices are T and F, both complementation and taking the opposite of a truth value yield the same result. The opposite of true is false and vice versa. The complement of true is false and vice versa. The situation changes when we go to three or more truth values. Suppose we have three truth values, as we have with $T_\forall$, $T_\exists$, and $T_0$. What is the meaning of "not $T_\forall$"? Well, that depends on whether we have "other than" or "opposite of" in mind. The opposite of $T_\forall$ is $T_0$ and vice versa, since we can intuitively define "opposite of" as follows. Listing a range of discrete values, there is a central value or pair of values, depending on whether the list of values has odd or even length, respectively. Two values are said to be "opposite" each other when they are on opposite sides of the central value and equidistant from it. If the list has odd length, it is easy to see how this works. If the list has even length, the two central values are the opposites of each other. Any other opposites can be seen as equidistant and on opposite sides of the central pair. So, for our three-valued range, $T_\forall$ and $T_0$ are at opposite ends of the range and are thus opposites. The single central value, $T_\exists$, is its own opposite. This makes sense when one considers that "partly true and partly false" is equivalent to "partly false and partly true."

Complementation works here as it does in other contexts. For a three-valued range, the complement of any particular value is the set containing the other two values. So, the complement of $T_\forall$ is $\{T_\exists, T_0\}$. Similarly, the complement of $T_0$ is $\{T_\forall, T_\exists\}$ and the complement of $T_\exists$ is $\{T_\forall, T_0\}$. In general, for an n-valued range, the complement of any one value is the set of all n-1 other values. The complement of a set of values is also meaningful, whereas it does not make much sense to speak of the opposite of a set of values (although this could be defined as the set of its members' individual opposites). We can give an intuitive meaning to certain subsets of ranges, and then to their complements. For example, we mentioned earlier that "some" can and should be disambiguated into "exclusive some" and "inclusive some." Exclusive some is familiar by now as "some but not all" and is used as the subscript for "only partly true." Inclusive some captures "some or all." This concept can also be expressed by formally taking the exclusive OR of the two values $\forall$ and $\exists_{ex}$. Literally this says "all or some-but-not-all but not both." We can denote this as $\forall \vee_{ex} \exists_{ex}$. The complement of "some or all" is "none" and can be expressed as $\neg(\forall \vee_{ex} \exists_{ex}) \equiv 0$. Similarly, "not all" can be expressed as either the complement of "all" or as the OR of "some but not all" and "none": $\neg\forall \equiv (\exists_{ex} \vee_{ex} 0)$.

One might question the use of the "not" symbol as potentially ambiguous. How are we to know when "opposite of" is meant, versus when "other than" is meant? In actual usage, "not" usually means "other than" rather than "opposite of." In fact it is hard to think of examples of "not" meaning "opposite of" in natural language. "Opposite of" comes up when we discuss qualities such as "hot" and "cold", "black" and "white," and so on. By "not hot" we wouldn't necessarily mean "cold." The same is true with other quality terms that do have opposite terms. The only time the "opposite of" a value is equivalent to its complement is when there are only the two values. That is why in traditional logic, where there are just two truth values, "other than" TRUE is the same as the "opposite of" TRUE. "Other than" collapses to "opposite of" in the two-valued case. So, let us reserve the "not" symbol for meaning "other than." For "opposite of" we can use a minus sign.

## *The Negation Theorem*

Allowing that we might somehow apply "opposite of" to any of the four % qualifiers, each of which we can constrain to take on one of the three values ∀, ∃, or 0, we can derive a curious invariant. Recall that the qualifiers regard respectively the truth value of a statement, the proportion of cases participating in the assertion, the proportion of the time for which the assertion is claimed to hold, and the degree with which the subject of the sentence is in the relation with the object. These qualifiers are symbolized as $\%_0$, $\%_1$, $\%_2$, and $\%_3$ respectively. The negation theorem is based on working with just these abstracted properties of a statement, completely disregarding the content of x, r, and y. It concerns the preservation of meaning under certain prescribed applications of "other than" and "opposite of." Working with just these abstracted properties and varying the status of each qualifier systematically with regard to whether it is negated in one sense or the other, or in both senses or in neither, we can prove the theorem by the inelegant but tractable and effective method of enumerating and testing all possible variations against the common sense meaning of the result.

The theorem goes as follows. Suppose we have a statement whose qualifiers are $\%_0$ $\%_1$ $\%_2$ $\%_3$. Fixing x, r, and y, suppose further that we wish to preserve the meaning of the statement but will apply the negation operations to one or more of the four qualifiers. When is meaning preserved? Let us depict equivalence as follows. In the tautologous case:

$$\%_0\ \%_1\ \%_2\ \%_3 \equiv \%_0\ \%_1\ \%_2\ \%_3 \,.$$

Of course each side is short for $\%_0$ $\%_1$/cud x $\%_3$ r y $\%_2$/tud. Where can we safely place ¬'s and -'s on the two sides of the equivalence and still have equivalence? The answer is: whenever there is a - applied to $\%_n$ there must be a ¬ applied to $\%_{n+1}$. The ¬ may appear either on the same side of the equivalence as the - or on the other side. Let us illustrate this:

$$\%_0\ \%_1\ \%_2\ \%_3 \equiv -\%_0\ \neg\%_1\ \%_2\ \%_3$$

$$\%_0\ \%_1\ \%_2\ \%_3 \equiv \%_0\ -\%_1\ \neg\%_2\ \%_3$$

$$\%_0\ \%_1\ \%_2\ \%_3 \equiv \%_0\ \%_1\ -\%_2\ \neg\%_3$$

$$\%_0\ \%_1\ \%_2\ \%_3 \equiv -\%_0\ -\neg\%_1\ \neg\%_2\ \%_3$$

$$\%_0\ \%_1\ \%_2\ \%_3 \equiv -\%_0\ \neg\%_1\ -\%_2\ \neg\%_3$$

$$\%_0\ \%_1\ \%_2\ \%_3 \equiv \%_0\ -\%_1\ -\neg\%_2\ \neg\%_3$$

$\%_0\ \%_1\ \%_2\ \%_3 \equiv -\%_0\ -\neg\%_1\ -\neg\%_2\ \neg\%_3$

$-\%_0\ \%_1\ \%_2\ \%_3 \equiv \%_0\ \neg\%_1\ \%_2\ \%_3$

$\%_0\ -\%_1\ \%_2\ \%_3 \equiv \%_0\ \%_1\ \neg\%_2\ \%_3$

$-\%_0\ \%_1\ \%_2\ \neg\%_3 \equiv \%_0\ -\neg\%_1\ -\neg\%_2\ \%_3$

and so on.

Before we look at what such an equivalence might be saying, let us deal with questions that may arise concerning applying both negation operators to the same qualifier. First, what do we get when we apply both operators to a qualifier? Second, does it matter in which order we apply them? Suppose the value of the qualifier is $\forall$. Let's first apply - and then $\neg$. Then let's start over and apply $\neg$ first and then -. $-\forall = 0$. $\neg 0 = \{\forall, \exists\}$. So, the first way we get $-\neg\forall = \{\forall, \exists\}$. Starting over, $\neg\forall = \{\exists, 0\}$. $-\{\exists, 0\} = \{\exists, \forall\}$ (remember that we said earlier that the opposite of a set can be defined as the set of the opposites of its elements). Interestingly, then, we get the same answer either way: $-\neg\forall = \neg-\forall = \{\forall, \exists\} = \{\exists, \forall\}$. What about when we apply the two operations to a central value, $\exists$, rather than a polar value such as $\forall$ or 0? Does the order matter then? $-\exists = \exists$. $\neg\exists = \{\forall, 0\}$. Applying the operations in the reverse order, $\neg\exists = \{\forall, 0\}$. $-\{\forall, 0\} = \{0, \forall\}$. The results are the same once again. Thus we can see that the order that the two operations are applied in does not matter.

Now, what would an example of one of the above sentence equivalences be that might make the theorem a bit more concrete? Let us take the following instantiation of the qualifiers and negation operators:

$-T_\forall\ \forall\ \forall\ \neg\forall \equiv T_\forall\ -\neg\forall\ -\neg\forall\ \forall$

and let us manipulate this to somewhat simplify its intelligibility:

$T_0\ \forall\ \forall\ \neg\forall \equiv T_\forall\ \neg 0\ \neg 0\ \forall.$

All that we have done is to replace "the opposite of $T_\forall$" with $T_0$ and replace $-\neg\forall$ with the equivalent "other than 0." "Other than 0" ($\neg 0$), is $\{\forall, \exists\}$, and as we just saw above, $-\neg\forall = \{\forall, \exists\}$. So what do our left and right hand sides say, now that we've made them a little simpler?

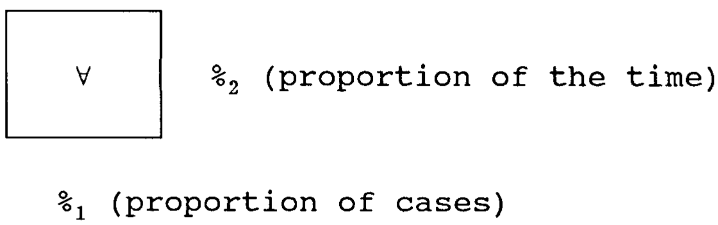
"It is not at all true that all are at all times at most partly" ≡
"It is entirely true that at least some are at least some times entirely"

All what are? At most partly what? It doesn't matter. Although we are working entirely in the abstract, this can be done succesfully with no regard for what the x, r, and y constituents of the statements are.

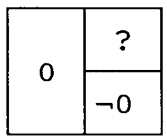
As it may appear difficult to juggle so many variables (the four qualifiers and any variations on negation), there is a pictorial method in natural logic which helps one envision what is being stated. The picture for one statement can be easily compared visually with the picture for another statement to see if the statements are equivalent. The picture summarizes the situation concerning x r y for the cases and times under discussion. Again, we work strictly in the abstract, so it doesn't matter exactly what x, r, and y are. Rather we care here only about what the values of the four qualifiers are, and apply any negation operations specified to arrive at an overall picture.

The technique works as follows. Each picture is rectangular. The horizontal (x) axis is for cases and the vertical (y) axis is for times. What goes inside the border is the value for $\%_3$, the degree of relation. For example, if $\%_3$ is $\forall$, that is, if the x's of the statement are entirely r y, then $\forall$ goes inside the picture. Being self-contained domain/range summaries, the pictures are called "domanges." Each possible configuration of or variation on the qualifiers and negations can be mapped to a unique domange picture. Let us first illustrate the technique for drawing a domange picture and then apply the technique to show the equivalence of two statements.

Suppose that we want to depict $\%_0\ \%_1\ \%_2\ \%_3 = T_\forall\ \forall\ \forall\ \forall$, i.e. that it is entirely true that all x under discussion are at all times entirely r y, or for short, that all are at all times entirely. The picture for this is:

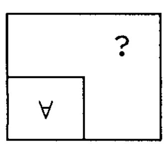



From now on, we'll dispense with the axis labels. Let's look at another example that will help convey how these pictures express information. Suppose we wish to depict $T_\forall\ \exists\ \forall\ 0$ (only some are at all times not-at-all:




This curious notation says that some cases are at all times not at all (0), but other cases are at least some times at least partly (non-0). The question mark indicates that for those cases which are at least some times non-0, the rest of the time those cases can be any value.

Now let us look at those two statements stated above to be equivalent by the negation theorem and see if their domange pictures coincide. The first expression was $\neg T_\forall\ \forall\ \forall\ \neg\forall$. This can be paraphrased as "It is not at all true that all are at all times not-entirely." To depict this we can think of it as saying that, in effect, at least some are at some times other than not-entirely. The picture for this is




In other words, at least some are at some times entirely. But this is, of course, exactly what our translation of the equivalent statement was. The equivalent statement was $T_\forall\ \neg\forall\ \neg\neg\forall\ \forall$ and we simplified its complexity to get $T_\forall$ ¬0 ¬0 ∀. The picture for the latter expression is exactly the picture that we just drew.

In general, this technique produces equivalent domange pictures for qualifier patterns with equivalent underlying meaning.

## *Implication Revisited*

We have acknowledged that an implication is false when its antecedent is true and its conclusion is false, even when we know nothing about their relationship, if any. We have also rejected the traditional treatment of the other three combinations of antecedent vs. conclusion truth and falsity, that is, that in the other three cases the implication is automatically true. That being the case, when *is* an implication true?

Intuitively, we are asking "When is the statement 'If p then q' true?" On closer analysis, we need to distinguish some key points. First, we could be talking about a state or event p in the physical world, or we could be talking about a statement p being true. Similarly, q could be a physical state or event, or a logical statement. And, in the same implication p could be physical and q logical, or vice versa. These aspects of the situation may or may not be important. We shall see.

Another key point for departure is the temporal relation between p and q. We can say "if p is the case then q was the case," or "If p is the case then q is the case," or "If p is the case then q will be the case."

It turns out that the causal or definitional relationship between p and q have the utmost importance for determining whether "if p then q" is true. The physical versus logical nature of the two participants p and q is not so much a deciding factor as it is an interesting feature of the situation. In general, we can cut across the physical vs. logical boundary when we talk about necessary and sufficient conditions. We can think of the necessary conditions for some physical situation or event in the same way that we think of the defining properties of some logical concept or relation.

Let us define a set of five relations which are all that we will need in order to address the relevant relationships between p and q for implication. The relations are mutually exclusive and together cover all possibilities. The five relations thus partition the relevant set of possible causal relationships between p and q. The five relations are:

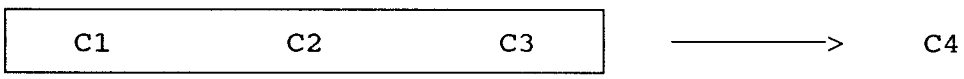
1) the set of sufficient conditions for
2) one set of sufficient conditions for
3) a globally necessary condition for
4) a locally necessary condition for
5) not causally related to

We can symbolize these as:

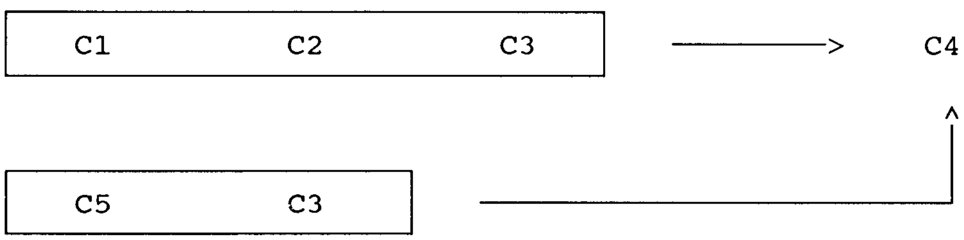
1) $x \ \mathrm{suf}_{\mathrm{the}} \ y$
2) $x \ \mathrm{suf}_{\mathrm{one}} \ y$
3) $x \ \mathrm{nec}_{\mathrm{glo}} \ y$
4) $x \ \mathrm{nec}_{\mathrm{loc}} \ y$
5) $x \ \neg\mathrm{rel}_{\mathrm{cau}} \ y$

Please note that $nec_{glo}$ and $nec_{loc}$ are considered mutually exclusive. Let us briefly explain the relations. An outcome can have exactly one set of conditions that brings it about, or it can have more than one way of coming into effect. In the former case, we say that the sole set of conditions that causes the outcome is "the set of sufficient conditions for" that result. In the latter case, no one set of conditions has the monopoly on causing the outcome, so each set of conditions able to bring about the result is "one set of sufficient conditions for" it. A certain condition can be necessary in order for some outcome to occur. If that necessary condition is *always* necessary then it is called "a globally necessary condition for" the outcome. If the necessary condition is only sometimes necessary for the result to occur, then it is "a locally necessary condition for" the result.

If we imagine a set of sufficient conditions for some outcome as containing no superfluous or irrelevant members, then we can say that each member of the set is a locally necessary condition for the outcome produced by the set. If there is only one set of sufficient conditions for an outcome, then each of the participating relevant conditions can be considered globally necessary as well. In this case, the global participation takes precedence over the local. If there are multiple sets of sufficient conditions for an outcome, then local to each set are its particular members. Any necessary conditions common to all sufficient sets are globally necessary, and in this case are the only global necessary conditions. Let us illustrate this:




Example 1. One set of sufficient conditions.




Example 2. Two sufficient sets. C3 is globally necessary.

We now have the tools with which to discuss implication. When is "If p then q" true, and what is its truth value or truth state otherwise? Let's first take the cases where the implication is true. In the case where p is preceded by q, one way for the implication to be true is for q to be the set of sufficient conditions for p. If p is the case then q was the case. Another relationship between p and q with p preceded by q where the implication is true is when q is a globally necessary condition for p.

Looking next at when p and q are present at the same time, the implication is true when either p is the only set of sufficient conditions for q, or when it is merely one of many such sets. Interestingly, when we look at the third scenario, that is, when p precedes q, the same two relationships are the relevant ones. When p precedes q, and q follows from p, it is because p is sufficient for q (either solely or as one of many ways to get q).

What about all of the other relationships that p and q can have, even in our circumscribed world of just five causal relations and their temporal relations? In some of those other cases, the result is that the implication is false. But, in the rest of those other cases, the implication's truth value is "sometimes true and sometimes false." For example, take the case where we are looking back in time from p, and q is only one of multiple sets of sufficient conditions for p. In this example, if p is the case then q may or may not have been the case. The truth state, given that we know p is the case but don't know whether q was in effect or not, is "indeterminate." If we look at the matter in a different way, that is, from a "determinate" point of view, then it is definite that p *may* have been caused by q. We give this summary the truth value $T_\exists$, i.e. the implication is definitely true only sometimes and the rest of the time it is false.

We can summarize the truth value assignment for implication in the following table. "$p \supset_{\leftarrow} q$" means "if p is then q was," "$p \supset_{=} q$" means "if p is then q is," and "$p \supset_{\rightarrow} q$" means "if p is then q will be."

| p r q | $p \supset_{\leftarrow} q$ | $p \supset_{=} q$ | $p \supset_{\rightarrow} q$ |
|---|---|---|---|
| q $suf_{the}$ p | $T_\forall$ | $T_0$ | $T_0$ |
| q $suf_{one}$ p | $T_\exists$ | $T_0$ | $T_0$ |
| q $nec_{glo}$ p | $T_\forall$ | $T_0$ | $T_0$ |
| q $nec_{loc}$ p | $T_\exists$ | $T_0$ | $T_0$ |
| q $\neg rel_{cau}$ p | $T_0$ | $T_0$ | $T_0$ |
| p $suf_{the}$ q | $T_0$ | $T_\forall$ | $T_\forall$ |
| p $suf_{one}$ q | $T_0$ | $T_\forall$ | $T_\forall$ |
| p $nec_{glo}$ q | $T_0$ | $T_\exists$ | $T_\exists$ |
| p $nec_{loc}$ q | $T_0$ | $T_\exists$ | $T_\exists$ |
| p $\neg rel_{cau}$ q | $T_0$ | $T_0$ | $T_0$ |

The strategy for determining implication truth or falsity is now: first, make the traditional syntactic check for true premise and false conclusion. If this is the case, we're done; the implication is false. If this is not the case, then we must dig further and use the above table once we know something about the semantic relationship between p and q.

## *Afterword*

This paper has been a first attempt to document some of the findings arrived at within the framework of natural logic. Many of the features presented here have been investigated in greater depth than time and space allow discussion of at this point. Other topics have been explored as well.

Conspicuously absent to the serious scholar are any references to the literature. On the one hand, only one work really contributed to the development of natural logic, the introductory logic text, *Exercises in Logic* by Terrell and Baker (Holt, Rinehart and Winston, New York 1967) in the undergraduate logic course attended by the author in 1969. This introduction to symbolic logic started the author off on the analysis that has led to the results presented here. On the other hand, the ideas here have for practical purposes been developed in isolation and in a self-contained fashion. Through the years the author has seen discussion from time to time of related topics and issues. This has been heartening. Rather than influence this work, however, the work of others has seemed somewhat orthogonal. Even work on multi-valued logic, and especially three-valued logic, has seemed divergent from what the author has been arriving at.

There are many issues that could be pursued in connection with the ideas and presentation here, and perhaps they shall at some later time. Related to the points just raised, how does this work tie into the literature? Specifically, one could see looking into comparing and contrasting natural logic with fuzzy logic, multi-valued logic, temporal logic, modal logic, intuitionistic logic, deviant logic, relevant logic, and so on. How do proof theory and model theory apply? What are the relations with AI work in nonmonotonic reasoning in general, default logic, circumscription, subsumption, truth maintenance, logic programming, qualitative reasoning, etc.? If any of these natural logic results are actually new and turn out to be valid, what can we do with them? Would software embodying these concepts be of use or interest? No attempt has been made here to constrain the theory to computational tractability. Rather, what has been emphasized has been tractability of analysis -- trying to keep things so that they make sense.

Natural logic touches upon many areas. To mention a few, besides logic there are: linguisitics and natural language studies, epistemology, ontology, psychology and cognitive science, computer science, the scientific method, and even education. How do we internationalize the details here that are so far English-based without distancing them too far from natural language and turning them into an unintelligible tangle of symbols and jargon? How do we make logic easily mastered, when it is one of the great tools of humanity and a facility whose use is ubiquitous among peoples day-in and day-out whether or not individuals know that they use it or how they use it? One thing is certain -- logic must be made intuitive. It is, after all, not an artificial construct but a natural faculty.

$\equiv$

$p \equiv q$ $\qquad$ $p \supset q \;\&\; q \supset p$

1. HAVE SAME MEANING (e.g. USING NEGATION THEOREM)

2. NO GOOD FOR FEEDBACK SITUATION ($p$ CAUSES $q$ & $q$ CAUSES $p$ BUT $p \not\equiv q$)

PERHAPS $p \supset_= q$ & $q \supset_= p$ ?

3. EXISTING $\supset$ IN NATURALOGIC HAS NO WAY FOR $\supset_=$ TO BE AT LEAST PARTLY TRUE IN BOTH DIRECTIONS. SO, MAYBE A 6th REL IS NEEDED: $p \equiv q$! (BESIDES $p$ $suf_{THE}$ $q$, etc.)!

4. DEGREES OF EQUIVALENCE:

$\{a, b, c\} = \{b, c, a\}$

BUT IF THE ELEMENTS OF THE SETS ARE <u>STRUCTURED</u> DIFFERENTLY THEN THE CONSTRUCTS ARE <u>NOT</u> EQUIVALENT

IS ICE $\equiv$ WATER $\equiv$ STEAM?

IS DIAMOND $\equiv$ GRAPHITE $\equiv$ CARBON?

SO, COMPONENT-EQUIVALENCE (OR CONTENT-EQ.)

STRUCTURE-EQUIVALENCE (OR FORM-EQ.)

COMPLETE EQUIVALENCE (BOTH OF ABOVE)