6

THEORIES (1929)

Let us try to describe a theory simply as a language for discussing the facts the theory is said to explain. This need not commit us on the philosophical question of whether a theory is only a language, but rather if we knew what sort of language it would be if it were one at all, we might be further towards discovering if it is one. We must try to make our account as general as possible, but we cannot be sure that we have in fact reached the most general type of theory, since the possible complication is infinite.

First, let us consider the facts to be explained. These occur in a universe of discourse which we will call the *primary system*, this system being composed of all the terms [1] and propositions (true or false) in the universe in question. We must suppose the primary system in some way given to us so that we have a notation capable of expressing every proposition in it. Of what sort must this notation be?

It might in the first case consist of names of different types any two or more of which conjoined together gave an atomic proposition; for instance, the names *a, b . . . z*, ' red ', ' before '. But I think the systems we try to explain are rarely of this kind; if for instance we are concerned with a series of experiences, we do not try to explain their time order (which we could not explain by anything simpler) or

[1] The ' universe ' of the primary system might contain ' blue or red ' but not ' blue ' or ' red '; i.e. we might be out to explain when a thing was ' blue or red ' as opposed to ' green or yellow ', but not which it was, blue or red. ' Blue-or-red ' would then be a term: ' blue ', ' red ' nonsense for our present purpose.

even, assuming *an* order, whether it is a or b that comes first; we take for granted that they are in order and that a comes before b, etc., and try to explain which is red, which blue, etc. a is essentially one before b, and 'a', 'b', etc., are not really names but descriptions except in the case of the present. We take it for granted that these descriptions describe uniquely, and instead of 'a was red' we have e.g. 'The 3rd one ago was red'. The symbols we want are not names but numbers: the 0th (i.e. the present), 1st, –1th, etc., in general the nth, and we can use red (n) to mean the nth is red counting forward or backward from a particular place. If the series terminates at say 100, we could write $N(101)$, and generally $N(m)$ if $m > 100$, meaning 'There is no mth'; or else simply regard e.g. red (m) as nonsense if $m > 100$, whereas if we wrote $N(m)$ we should say red (m) was false. I am not sure this is necessary, but it seems to me always so in practice; i.e. the terms of our primary system have a structure, and any structure can be represented by numbers (or pairs or other combinations of numbers).

It may be possible to go further than this, for of the terms in our primary system not merely some but even all may be best symbolized by numbers. For instance, colours have a structure, in which any given colour may be assigned a place by three numbers, and so on. Even smells may be so treated: the presence of the smell being denoted by 1, the absence by 0 (or all total smell qualities may be given numbers). Of course, we cannot make a proposition out of numbers without some link. Moment 3 has colour 1 and smell 2 must be written $\chi(3) = 1$ and $\phi(3) = 2$, χ and ϕ corresponding to the general forms of colour and smell, and possibly being functions with a limited number of values, so that e.g. $\phi(3) = 55$ might be nonsense, since there was no 55th smell.

Whether or no this is possible, it is not so advantageous where we have relatively few terms (e.g. a few smells) to deal with. Where we have a multitude as e.g. with times, we cannot name them, and our theory will not explain a primary system in which they have names, for it will take no account of their individuality but only of their position. In general nothing is gained and clarity may be lost by using numbers when the order, etc., of the numbers corresponds to nothing in the nature of the terms.

If all terms were represented by numbers, the propositions of the primary system would all take the form of assertions about the values taken by certain one-valued numerical functions. These would not be mathematical functions in the ordinary sense; for that such a function had such-and-such a value would always be a matter of fact, not a matter of mathematics.

We have spoken as if the numbers involved were always integers, and if the finitists are right this must indeed be so in the ultimate primary system, though the integers may, of course, take the form of rationals. This means we may be concerned with pairs (m, n) with $(\lambda m, \lambda n)$ always identical to (m, n). If, however, our primary system is already a secondary system from some other theory, real numbers may well occur.

So much for the primary system; now for the theoretical construction.

We will begin by taking a typical form of theory, and consider later whether or not this form is the most general. Suppose the atomic propositions of our primary system are such as $A(n)$, $B(m, n)$ . . . where m, n, etc., take positive or negative integral values subject to any restrictions, e.g. that in $B(m, n)$ m may only take the values 1, 2.

Then we introduce new propositional functions $\alpha(n)$, $\beta(n)$, $\gamma(m, n)$, etc., and by propositions of the *secondary system* we shall mean any truth-functions of the values of α, β, γ, etc. We shall also lay down propositions about these values, e.g. $(n).\ \overline{\alpha(n).\ \beta(n)}$ which we shall call *axioms*, and whatever propositions of the secondary system can be deduced from the axioms we shall call *theorems*.

Besides this we shall make a *dictionary* which takes the form of a series of definitions of the functions of the primary system $A, B, C. \ . \ . \ .$ in terms of those of the secondary system α, β, γ, e.g. $A(n) = \alpha(n) . \vee . \gamma(O, n^2)$. By taking these 'definitions' as equivalences and adding them to the axioms we may be able to deduce propositions in the primary system which we shall call *laws* if they are general propositions, *consequences* if they are singular. The totality of laws and consequences will be the eliminant when α, β, γ . ., etc., are eliminated from the dictionary and axioms, and it is this totality of laws and consequences which our theory asserts to be true.

We may make this clearer by an example[1]; let us interpret numbers n, n_1, n_2, etc.. as instants of time and suppose the primary system to contain the following functions :—

$A(n)$ = I see blue at n.

$B(n)$ = I see red at n.

[$\bar{A}(n) . \bar{B}(n)$ = I see nothing at n].

$C(n)$ = Between n–1 and n I feel my eyes open.

$D(n)$ = Between n–1 and n I feel my eyes shut.

$E(n)$ = I move forward a step at n.

$F(n)$ = I move backward a step at n.

and that we construct a theory in the following way :

[1] [The example *seems* futile, therefore try to invent a better ; but it in fact brings out several good points, which it would be difficult otherwise to bring out. It may however miss some points which we will consider later. A defect in all Nicod's examples is that they do not give an external world in which anything *happens*.—F. P. R.]

First m will be understood to take only the values 1, 2, 3,

and $f(m)$ is defined by $\begin{cases} f(1) = 2 \\ f(2) = 3 \\ f(3) = 1 \end{cases}$

Then we introduce

$$\alpha(n, m) = \text{At time } n \text{ I am at place } m.$$

$$\beta(n, m) = \text{At time } n \text{ place } m \text{ is blue.}$$

$$\gamma(n) \quad = \text{At time } n \text{ my eyes are open.}$$

And the axioms

$$(n, m, m') : \quad \alpha(n, m) \,.\, \alpha(n, m') \,.\, \supset \,.\, m = m'.$$

$$(n). \quad (\exists m). \quad \alpha(n, m).$$

$$(n). \quad \beta(n, 1).$$

$$(n) : \quad \beta(n, 2) \,.\, \equiv \,.\, \tilde{\beta}(n + 1, 2).$$

And the dictionary

$$A(n) = (\exists m) \,.\, \alpha(n, m) \,.\, \beta(n, m) \,.\, \gamma(n).$$

$$B(n) = (\exists m) \,.\, \alpha(n, m) \,.\, \tilde{\beta}(n, m) \,.\, \gamma(n).$$

$$C(n) = \tilde{\gamma}(n - 1) \,.\, \gamma(n).$$

$$D(n) = \gamma(n - 1) \,.\, \tilde{\gamma}(n).$$

$$E(n) = (\exists m) \,.\, \alpha(n - 1, m) \,.\, \alpha\{n, f(m)\}.$$

$$F(n) = (\exists m) \,.\, \alpha\{n - 1, f(m)\} \,.\, \alpha(n, m).$$

This theory can be said to represent me as moving among 3 places, 'forwards' being in the sense $ABCA$, 'backwards' $ACBA$. Place A is always blue, place B alternately blue and red, place C blue or red according to a law I have not discovered. If my eyes are open I see the colour of the place

I am in, if they are shut I see no colour. The laws resulting from the theory can be expressed as follows:

(1) $(n) . \{\bar{A}(n) \vee \bar{B}(n)\} : \{\bar{C}(n) \vee \bar{D}(n)\} . \{\bar{E}(n) \vee \bar{F}(n)\}$

(2) $(n_1, n_2) \{n_1 > n_2 . C(n_1) . C(n_2) . \supset . (\exists n_3) . n_1 > n_3 > n_2 . D(n_3)\}$

(2[1]) (2) with the C's and D's interchanged

Let us define $0 \; (n_1, n_2)$ to mean
1
2

$$Nc' \hat{\nu} \{n_1 < \nu \leqslant n_2 . E(\nu)\}$$

$$-Nc' \hat{\nu} \{n_1 < \nu \leqslant n_2 . F(\nu)\} \equiv \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} \pmod 3$$

(3) $[(\exists n_1) . C(n_1) . n_1 \leqslant n . n \geqslant \nu > n_1 . \supset_\nu . \bar{D}(\nu)] \supset_n : A(n) \vee B(n)$

(4) $[(\exists n_1) . D(n_1) . n_1 \leqslant n . n \geqslant \nu \geqslant n_1 . \supset_\nu . \bar{C}(\nu)] \supset_n : \bar{A}(n) . \bar{B}(n)$.

(5) $(n) :. (\exists m) : m = 0, 1, \text{ or } 2 : m(\nu, n) \supset_\nu \bar{B}(\nu)$

$$: (m - 1) \;\; (\nu_1, n) . (m - 1) \;\; (\nu_2, n) . \nu_1 \not\equiv \nu_2 \pmod 2$$

$$. \supset_{\nu_1, \nu_2} . \bar{A}(\nu_1) \vee \bar{A}(\nu_2) . \bar{B}(\nu_1) \vee \bar{B}(\nu_2).$$

[Where $0 - 1 = 2$ for this purpose.]

These can then be compared with the axioms and dictionary, and there is no doubt that to the normal mind the axioms and dictionary give the laws in a more manageable form.

Let us now put it all into mathematics by writing

$$A(n) \qquad \text{as } \phi(n) = 1$$

$$B(n) \qquad \text{as } \phi(n) = -1$$

$$\bar{A}(n) \,.\, \bar{B}(n) \text{ as } \phi(n) = 0$$
$$C(n) \quad \text{as } \chi(n) = 1$$
$$D(n) \quad \text{as } \chi(n) = -1$$
$$\bar{C}(n) \,.\, \bar{D}(n) \text{ as } \chi(n) = 0$$
$$E(n) \quad \text{as } \psi(n) = 1$$
$$F(n) \quad \text{as } \psi(n) = -1$$
$$\bar{E}(n)\ \bar{F}(n) \quad \text{as } \psi(n) = 0.$$

Instead of $\alpha(n, m)$ have $\alpha(n)$ a function taking values 1, 2, 3
$\beta(n, m)$,, $\beta(n, m)$,, ,, 1, -1
$\gamma(n)$ $\gamma(n)$ 1, 0

Our axioms are just

$$(1)\quad (n) \,.\, \alpha(n) = 1 \vee 2 \vee 3$$
$$(2)\quad (n) \,.\, \beta(n, 1) = 1$$
$$(3)\quad (n) \,.\, \beta(n, 2) \neq \beta(n + 1, 2)$$
$$(4)\quad (n, m) \,.\, \beta(n, m) = 1 \vee -1$$
$$(5)\quad (n) \,.\, \gamma(n) = 0 \vee 1$$

Of these (1) (4) (5) hardly count since they merely say what values the functions are capable of taking.

Our definitions become.

$$\text{(i)}\quad \phi(n) = \gamma(n) \times \beta\{n, \alpha(n)\}$$
$$\text{(ii)}\quad \chi(n) = \gamma(n) - \gamma(n - 1)$$
$$\text{(iii)}\quad \psi(n) = \text{Remainder mod 3 of } \alpha(n) - \alpha(n - 1)$$

Our laws are of course that ϕ, χ, ψ must be such that α, β, γ can be found to satisfy 1–5, i–iii. Going through the old laws we have instead of

(1) $\phi(n) = -1 \vee 0 \vee 1$, $\chi(n) = -1 \vee 0 \vee 1$, $\psi(n) = -1 \vee 0 \vee 1$ [understood].

(2) $(n, m) . \left| \sum_{r=n}^{m} \chi(\nu) \right| \leqslant 1.$

(3) $(\exists m) . \sum_{r=m}^{n} \chi(r) = 1 : \supset_n . \phi(n) \neq 0.$

(4) $(\exists m) . \sum_{r=m}^{n} \chi(r) = -1 : \supset_n . \phi(n) = 0.$

(5) $(n) :. (\exists m) : \sum_{r=n}^{n'} \psi(r) \equiv m \ (\text{mod } 3) \ . \supset_{n'} \ . \ \phi(n') \neq -1$

$: \sum_{r=n}^{n'} \psi(r) \equiv \sum_{r=n}^{n''} \psi(r) \equiv m-1 \ (\text{mod } 3). \ n' \equiv n'' + 1 \ (\text{mod } 2)$

$. \supset_{n', n''} . \phi(n') \, \phi(n'') = 0 \lor -1.$

So far we have only shown the genesis of *laws* ; *consequences* arise when we add to the axioms a proposition involving e.g. a particular value of n, from which we can deduce propositions in the primary system not of the form (n) . . . These we call the *consequences*.

If we take it in its mathematical form we can explain the idea of a theory as follows : Instead of saying simply what we know about the values of the functions with which we are concerned, we say that they can be constructed in a definite way given by the dictionary out of functions satisfying certain conditions given by the axioms.

Such then is an example of a theory ; before we go on to discuss systematically the different features of the example and whether they occur in any theory, let us take some questions that might be asked about theories and see how they would be answered in the present case.

1. Can we say anything in the language of this theory that we could not say without it ?

Obviously not ; for we can easily eliminate the functions of the second system and so say in the primary system all that the theory gives us.

2. Can we reproduce the structure of our theory by means of explicit definitions within the primary system?

[This question is important because Russell, Whitehead, Nicod and Carnap all seem to suppose that we can and must do this.[1]]

Here there are some distinctions to make. We might, for instance, argue as follows. Supposing the laws and consequences to be true, the facts of the primary system must be such as to allow functions to be defined with all the properties of those of the secondary system, and these give the solution of our problem. But the trouble is that the laws and consequences can be made true by a number of different sets of facts, corresponding to each of which we might have different definitions. So that our problem of finding a single set of definitions which will make the dictionary and axioms true whenever the laws and consequences are true, is still unsolved. We can, however, at once solve it formally, by disjoining the sets of definitions previously obtained; i.e. if the different sets of facts satisfying the laws and consequences are P_1, P_2, P_3, and the corresponding definitions of $\alpha(n, m)$ are

$$\alpha(n, m) = \mathrm{L}_1 \{A, B, C \ldots, n, m\}$$

$$\mathrm{L}_2 \{A, B, C, \ldots, n, m\} \text{ etc.}$$

we make the definition

$$\alpha(n, m) = P_1 \supset L_1 \{A, B, C \ldots n, m\}.$$

$$P_2 \supset L_2 \{A, B, C \ldots n, m\}.$$

$$\text{etc.}$$

Such a definition is formally valid and evidently fulfils our requirements.

[1] Jean Nicod, *La Géométrie dans le Monde Sensible* (1924), translated in his *Problems of Geometry and Induction* (1930): Rudolf Carnap, *Der Logische Aufbau der Welt* (1928).

What can be objected to it is complexity and arbitrariness, since L_1, L_2 . . . can probably be chosen each in many ways.

Also it explicitly assumes that our primary system is finite and contains a definite number of assignable atomic propositions.

Let us therefore see what other ways there are of proceeding.

We might at first sight suppose that the key lay simply in the dictionary; this gives definitions of A, B, C . . . in terms of α, β, γ . . . Can we not invert it to get definitions of α, β, γ . . . in terms of A, B, C . . . ? Or, in the mathematical form, can we not solve the equations for α, β, γ . . . in terms of ϕ, χ, ψ . . ., at any rate if we add to the dictionary, as we legitimately can, those laws and axioms which merely state what values the functions are capable of taking?

When, however, we look at these equations (i), (ii), (iii) what we find is this: If we neglect the limitations on the values of the functions they possess an integral solution provided $\gamma(n)$ can be found from (ii) so as always to be a factor of $\phi(n)$, i.e. in general always to be ± 1 or 0 and never to vanish unless $\phi(n)$ vanishes. This is, of course, only true in virtue of the conditions laid on ϕ and χ by the laws; assuming these laws and the limitation on values, we get the solution

$$\alpha(n) \equiv \sum_{o}^{n} \psi(n) + C_1 \pmod 3$$

$$\gamma(n) = \sum_{o}^{u} \chi(n) + C_2$$

for α and γ.

And for $\beta(n, m)$ no definite solution but e.g. the trivial one $\beta(n, m) = \phi(n)$ (assuming $\gamma(n) = 1$ or 0).

Here C_2 must be chosen so as to make $\gamma(n)$ always 1 or 0; and the value necessary for this purpose depends on the facts of the primary system and cannot be deduced simply from the laws. It must in fact be one or nought:

(*a*) If there is a least positive or zero n for which $\chi(n) \neq 0$, according as $\chi(n)$ for that n is -1 or $+1$.

(*b*) If there is a least negative n for which $\chi(n) \neq 0$, according as $\chi(n)$ for that n is $+1$ or -1.

(*c*) If for no n $\chi(n) \neq 0$ it does not matter whether C_2 is $+1$ or -1.

We thus have a disjunctive definition of C_2 and so of $\gamma(n)$. Again although any value of C_1 will satisfy the limitations on the value of $\alpha(n)$, probably only one such will satisfy the *axioms*, and this value will again have to be disjunctively defined. And, thirdly, $\beta(n, m)$ is not at all fixed by the equations, and it will be a complicated matter in which we shall again have to distinguish cases, to say which of the many possible solutions for $\beta(n, m)$ will satisfy the axioms.

We conclude, therefore, that there is neither in this case nor in general any simple way of inverting the dictionary so as to get either a unique or an obviously preeminent solution which will also satisfy the axioms, the reason for this lying partly in difficulties of detail in the solution of the equations, partly in the fact that the secondary system has a higher multiplicity, i.e. more degrees of freedom, than the primary. In our case the primary system contains three one valued functions, the secondary virtually five [$\beta(n, 1)$, $\beta(n, 2)$, $\beta(n, 3)$, $\alpha(n)$, $\gamma(n)$] each taking 2 or 3 values, and such an increase of multiplicity is, I think, a *universal* characteristic of useful theories.

Since, therefore, the dictionary alone does not suffice, the next hopeful method is to use both dictionary and axioms in a way which is referred to in many popular discussions of theories when it is said that the meaning of a proposition about the external world is what we should ordinarily regard as the *criterion* or *test* of its truth. This suggests that we should define propositions in the secondary system by their criteria in the primary.

In following this method we have first to distinguish the *sufficient* criterion of a proposition from its *necessary* criterion. If p is a proposition of the secondary system, we shall mean by its sufficient criterion, $\sigma(p)$, the disjunction of all propositions q of the primary system such that p is a logical consequence of q together with the dictionary and axioms, and such that $\sim q$ is not a consequence of the dictionary and axioms.[1] On the other hand, by the necessary criterion of p, $\tau(p)$ we shall mean the conjunction of all those propositions of the primary system which follow from p together with the dictionary and axioms.

We can elucidate the connection of $\sigma(p)$ and $\tau(p)$ as follows. Consider all truth-possibilities of atomic propositions in the primary system which are compatible with the dictionary and axioms. Denote such a truth-possibility by r, the dictionary and axioms by a. Then $\sigma(p)$ is the disjunction of every r such that

$$r\,\bar{p}\,a \text{ is a contradiction,}$$

$\tau(p)$ the disjunction of every r such that

$$r\,p\,a \text{ is not a contradiction.}$$

If we denote by L the totality of laws and consequences, i.e. the disjunction of every r here in question, then we have evidently

$$\sigma(p) : \equiv : L \,.\, \sim \tau(\sim p), \qquad \text{(i)}$$

[1] The laws and consequences need not be added, since they follow from the dictionary and axioms. It might be thought, however, that we should take them *instead* of the axioms, but it is easy to see that this would merely increase the divergence between sufficient and necessary criteria and in general the difficulties of the method. The last clause could be put as that $\sim q$ must not follow from or be a law or consequence.

$$\tau(p) :\equiv: L . \sim \sigma(\sim p), \tag{ii}$$

$$\sigma(p) \vee \tau(\sim p) . \equiv . L. \tag{iii}$$

We have also

$$\sigma(p_1 . p_2) :\equiv: \sigma(p_1) . \sigma(p_2), \tag{iv}$$

for $p_1 . p_2$ follows from q when and only when p_1 and p_2 both follow.

Whence, or similarly, we get the dual

$$\tau(p_1 \vee p_2) . \equiv . \tau(p_1) \vee \tau(p_2). \tag{v}$$

We also have

$$\sigma(p) \supset \tau(p), \tag{vi}$$

(Consider the r's above.)

$$\sigma(p) \vee \sigma(\sim p) . \supset . L . \supset . \tau(p) \vee \tau(\sim p), \tag{vii}$$

(from iii)

and from (vi), (ii), (iii).

$$\sigma(p) . \supset . \sim \sigma(\sim p) . L, \tag{viii}$$

$$L . \sim \tau(\sim p) . \supset . \tau(p). \tag{ix}$$

Lastly we have

$$\sigma(p_1) \vee \sigma(p_2) . \supset . \sigma(p_1 \vee p_2). \tag{x}$$

Since if q follows either from p_1 or from p_2 it follows from $p_1 \vee p_2$; and the dual

$$\tau(p_1 . p_2) . \supset . \tau(p_1) . \tau(p_2). \tag{xi}$$

On the other hand, and this is a very important point, the converses of (vi)–(xi) are not in general true. Let us illustrate this by taking (x) and considering this 'r':

$$B(0) . \bar{A}(0) : n \neq 0 . \supset_n . \bar{A}(n) . \bar{B}(n).$$

$$C(n) \,.\, \equiv_n . \, n = 0 : D(n) \,.\, \equiv_n . \, n = 1.$$

$$(n) \,.\, \bar{E}n \,.\, \bar{F}n,$$

i.e. that the man's eyes are only open once when he sees blue.

From this we can deduce $\alpha(0, 2) \vee \alpha(0, 3)$

$\therefore$ This $r \supset \sigma\{\alpha(0, 2) \vee \alpha(0, 3)\}$.

But we cannot deduce from it $\alpha(0, 2)$ or $\alpha(0, 3)$, since it is equally compatible with either. Hence neither $\sigma\{\alpha(0, 2)\}$ nor $\sigma\{\alpha(0, 3)\}$ is true. Hence we do not have

$$\sigma\ \{\alpha(0,\ 2) \vee \alpha(0,\ 3)\} \supset \sigma\ \{\alpha(0,\ 2)\} \vee \sigma\ \{\alpha(0,\ 3)\}.$$

It follows from this that we cannot give definitions such that, if p is any proposition of the secondary system, p will in virtue of the definitions mean $\sigma(p)$ [or alternatively $\tau\ (p)$], for if p_1 is defined to mean $\sigma(p_1)$, p_2 to mean $\sigma(p_2)$, $p_1 \vee p_2$ will mean $\sigma(p_1) \vee \sigma(p_2)$, which is not, in general, the same as $\sigma(p_1 \vee p_2)$. We can therefore only use σ to define some of the propositions of the secondary systems, what we might call *atomic* secondary propositions, from which the meanings of the others would follow.

For instance, taking our functions α, β, γ we could proceed as follows:

$\gamma(n)$ is defined as $A(n) \vee B(n)$, where there is no difficulty as $A(n) \vee B(n) \equiv \sigma\{\gamma(n)\} \equiv \tau\ \{\gamma(n)\}$.

$\beta(n, m)$ could be defined to mean $\sigma\{\beta(n, m)\}$, i.e. we should say place m was 'blue' at time n, only if there were proof that it was. Otherwise we should say it was not 'blue' ('red' in common parlance).

$\bar{\beta}(n, m)$ would then mean $\bar{\sigma}\{\beta(n, m)\}$ not $\sigma\{\bar{\beta}(n, m)\}$.

Alternatively we could use τ, and define

$$\beta(n, m) \text{ to be } \tau\ \{\beta(n, m)\},$$

$$\text{and } \bar{\beta}(n, m) \text{ would be } \bar{\tau}\ \{\beta(n, m)\}$$

In this case we should say m was 'blue' whenever there was no proof that it was not; this could, however, have been achieved by means of σ if we had defined $\beta(n, m)$ to be $\sim \beta'(n, m)$, and $\beta'(n, m)$ to be $\sigma\{\beta'(n, m)\}$, i.e. applied σ to $\bar{\beta}$ instead of β.

In general it is clear that τ always gives what could be got from applying σ to the contradictory, and we may confine our attention to σ.

It makes, however, a real difference whether we define β or $\bar{\beta}$ by means of σ, especially in connection with place 3. For we have no law as to the values of $\beta(n, 3)$, nor any way of deducing one except when $\alpha(n, 3)$ is true and $A(n)$ or $B(n)$ is true.

If we define $\beta(n, 3)$ to be $\sigma\{\beta(n, 3)\}$, we shall say that 3 is never blue except when we observe it to be; if we define $\bar{\beta}(n, 3)$ to be $\sigma\{\bar{\beta}(n, 3)\}$ we shall say it is always blue except when we observe it not to be.

Coming now to $\alpha(n, m)$ we could define

$$\alpha(n, 1) = \sigma\{\alpha(n, 1)\},$$

$$\alpha(n, 2) = \sigma\{\alpha(n, 1) \vee \alpha(n, 2)\} . \bar{\sigma}\{\alpha(n, 1)\},$$

$$\alpha(n, 3) = \bar{\sigma}\{\alpha(n, 1) \vee \alpha(n, 2)\};$$

and we should for any n have one and only one of $\alpha(n, 1)$, $\alpha(n, 2)$, $\alpha(n, 3)$ true: whereas if we simply put

$$\alpha(n, m) = \sigma\{\alpha(n, m)\},$$

this would not follow, since

$$\sigma\{\alpha(n, 1)\}, \sigma\{\alpha(n, 2)\}, \sigma\{\alpha(n, 3)\}$$

could quite well all be false.

[E.g. if $(n) . \bar{A}(n) . \bar{B}(n)$]

Of course in all these definitions we must suppose $\sigma\{\alpha(n, m)\}$ etc., replaced by what on calculation we find them to be. As they stand the definitions look circular, but are not when interpreted in this way.

For instance $\sigma\{\alpha(n, 1)\}$ is L, i.e. laws (1)–(5) together with

$$(\exists n_1, n_2) \,.\, 2(n_1, n) \,.\, 2(n_2, n) \,.\, n_1 \not\equiv n_2 \pmod 2 \,.\, Bn_1 \,.\, Bn_2 \,.$$

$$\vee. \quad (\exists n_1, n_2, n_3) \,.\, 2(n_1, n) \,.\, 2(n_2, n) \,.\, 2(n_3, n) \,.\, n_1 \not\equiv n_2 \pmod 2 . \; An_1 \,.\, An_2 \,.\, Bn_3 \,.$$

$$\vee. \quad (\exists n_1, n_2) \,.\, 1(n_1, n) \,.\, 2(n_2, n) \,.\, Bn_1 \,.\, Bn_2.$$

Such then seem to be the definitions to which we are led by the popular phrase that the meaning of a statement in the second system is given by its criterion in the first. Are they such as we require ?

What we want is that, using these definitions, the axioms and dictionary should be true whenever the theory is applicable, i.e. whenever the laws and consequences are true ; i.e. that interpreted by means of these definitions, the axioms and dictionary should follow from the laws and consequences.

It is easy to see that they do not so follow. Take for instance the last axiom on p. 216 :

$$(n) : \beta(n, 2) \,.\equiv.\, \bar{\beta}(n + 1,\ 2),$$

which means according to our definitions

$$(n) : \sigma\{\beta(n, 2)\} \,.\equiv.\, \bar{\sigma}\{\beta(n + 1,\ 2)\},$$

which is plainly false, since if, as is perfectly possible, the man has never opened his eyes at place 2, both $\sigma\{\beta(n, 2)\}$ and $\sigma\{\beta(n + 1,\ 2)\}$ will be false.

[The definition by τ is no better, since $\tau\,\{\beta\ (n,\ 2)\}$ and $\tau\ (\beta\ (n + 1, 2))$ would both be true.]

This line of argument is, however, exposed to an objection of the following sort: If we adopt these definitions it is true that the axioms will not follow from the laws and consequences, but it is not really necessary that they should. For the laws and consequences cannot represent the whole empirical (i.e. primary system) basis of the theory. It is, for instance, compatible with the laws and consequences that the man should never have had his eyes open at place 2; but how could he then have ever formulated this theory with the peculiar law of alternation which he ascribes to place 2? What we want in order to construct our theory by means of explicit definitions is not that the axioms should follow from the laws and consequences alone, but from them together with certain existential propositions of the primary system representing experiences the man must have had in order to be able with any show of reason to formulate the theory.

Reasonable though this objection is in the present case, it can be seen by taking a slightly more complicated theory to provide us with no general solution of the difficulty; that is to say, such propositions as could in this way be added to the laws and consequences would not always provide a sufficient basis for the axioms. For instance, suppose the theory provided for a whole system of places identified by the movement sequences necessary to get from one to another, and it was found and embodied in the theory that the colour of each place followed a complicated cycle, the same for each place, but that the places differed from one another as to the phase of this cycle according to no ascertainable law. Clearly such a theory could be reasonably formed by a man who had not had his eyes open at each place, and had no grounds for thinking that he ever would open his eyes at all the places or even visit them at all. Suppose then m is a place he never goes to, and that $\beta(n, m)$ is a function of the second system, signifying that m is blue at n; then unless he knows the phase of m, we can never have $\sigma\{\beta(n, m)\}$, but if e.g. the cycle gives a blue colour once in six, we must have from an axiom $\beta(0, m) \vee \beta(1, m) \vee \ldots \vee \beta(6, m)$. We have, therefore, just the same difficulty as before.

If, therefore, our theory is to be constructed by explicit definitions, these cannot be simple definitions by means of σ (or τ), but must be more complicated. For instance, in regard to place 2 in our original example we can define

$$\beta(0, 2) \text{ as } \sigma\{\beta(0, 2)\},$$

$$\beta(n, 2) \text{ as } \sigma\{\beta(0, 2)\} \text{ if } n \text{ is even},$$

$$\sim \sigma\{\beta(0, 2)\} \text{ if } n \text{ is odd}.$$

I.e. if we do not know which phase it is, we assume it to be a certain one, including that 'assumption' in our definition. E.g. by saying the phase is blue-even, red-odd, we mean that we have reason to think it is; by saying the phase is blue-odd, red-even, we mean not that we have reason to think it is but merely that we have no reason to think the contrary.

But in general the definitions will have to be very complicated; we shall have, in order to verify that they are complete, to go through all the cases that satisfy the laws and consequences (together with any other propositions of the primary system we think right to assume) and see that in each case the definitions satisfy the axioms, so that in the end we shall come to something very like the general disjunctive definitions with which we started this discussion (p. 220). At best we shall have disjunctions with fewer terms and more coherence and unity in their construction; how much will depend on the particular case.

We could see straight off that (in a finite scheme) such definitions were always possible, and by means of σ and τ we have reached no real simplification.

3. We have seen that we *can* always reproduce the structure of our theory by means of explicit definitions. Our next question is 'Is this *necessary* for the legitimate use of the theory?'

To this the answer seems clear that it cannot be necessary, or a theory would be no use at all. Rather than give all these definitions it would be simpler to leave the facts, laws and consequences in the language of the primary system. Also the arbitrariness of the definitions makes it impossible for them to be adequate to the theory as something in process of growth. For instance, our theory does not give any law for the colour of place 3 ; we should, therefore, in embodying our theory in explicit definition, define place 3 to be red unless it was observed to be blue (or else *vice versa*). Further observation might now lead us to add to our theory a new axiom about the colour of place 3 giving say a cycle which it followed ; this would appear simply as an addition to the axioms, the other axioms and the dictionary being unaltered.

But if our theory had been constructed by explicit definitions, this new axiom would not be true unless we changed the definitions, for it would depend on quite a different assignment of colours to place 3 at times when it was unobserved from our old one (which always made it red at such times), or indeed from *any* old one, except exactly that prescribed by our new axiom, which we should never have hit on to use in our definitions unless we knew the new axiom already. That is to say, if we proceed by explicit definition we cannot add to our theory without changing the definitions, and so the meaning of the whole.

[But though the use of explicit definitions cannot be necessary, it is, I think, instructive to consider (as we have done) how such definitions could be constructed, and upon what the possibility of giving them simply depends. Indeed I think this is essential to a complete understanding of the subject.]

4. Taking it then that explicit definitions are not necessary, how are we to explain the functioning of our theory without them ?

Clearly in such a theory judgment is involved, and the judgments in question could be given by the laws and consequences, the theory being simply a language in which they are clothed, and which we can use without working out the laws and consequences.

The best way to write our theory seems to be this $(\exists\alpha, \beta, \gamma)$: dictionary . axioms.

The dictionary being in the form of equivalences.

Here it is evident that α, β, γ are to be taken purely extensionally. Their extensions may be filled with intensions or not, but this is irrelevant to what can be deduced in the primary system.

Any additions to the theory, whether in the form of new axioms or particular assertions like $\alpha(0, 3)$, are to be made within the scope of the original α, β, γ. They are not, therefore, strictly propositions by themselves just as the different sentences in a story beginning ' Once upon a time ' have not complete meanings and so are not propositions by themselves.

This makes both a theoretical and a practical difference :

(*a*) When we ask for the meaning of e.g. $\alpha(0, 3)$ it can only be given when we know to what stock of ' propositions ' of the *first and second* systems $\alpha(0, 3)$ is to be added. Then the meaning is the difference in the first system between $(\exists\alpha, \beta, \gamma)$: stock . $\alpha(0, 3)$, and $(\exists\alpha, \beta, \gamma)$. stock. (We include propositions of the primary system in our stock although these do not contain α, β, γ.)

This account makes $\alpha(0, 3)$ mean something like what we called above $\tau\{\alpha(0, 3)\}$, but it is really the difference between $\tau\{\alpha(0, 3) + \text{stock}\}$ and τ (stock).

(*b*) In practice, if we ask ourselves the question " Is $\alpha(0, 3)$ true ? ", we have to adopt an attitude rather different from that which we should adopt to a genuine proposition.

For we do not add $\alpha(0, 3)$ to our stock whenever we think we could truthfully do so, i.e. whenever we suppose $(\exists\alpha, \beta, \gamma) :$ stock . $\alpha(0, 3)$ to be true. $(\exists\alpha, \beta, \gamma) :$ stock . $\bar{\alpha}(0, 3)$ might also be true. We have to think what else we might be going to add to our stock, or hoping to add, and consider whether $\alpha(0, 3)$ would be certain to suit any further additions better than $\bar{\alpha}(0, 3)$. E.g. in our little theory either $\beta(n, 3)$ or $\bar{\beta}(n, 3)$ could always be added to any stock which includes $\bar{\alpha}(n, 3) . \vee . \bar{A}(n) . \bar{B}(n)$. But we do not add either, because we hope from the observed instances to find a law and then to fill in the unobserved ones according to that law, not at random beforehand.

So far, however, as *reasoning* is concerned, that the values of these functions are not complete propositions makes no difference, provided we interpret all logical combination as taking place within the scope of a single prefix $(\exists\alpha, \beta, \gamma)$; e.g.

$$\overline{\beta(n, 3) . \bar{\beta}(n, 3)} \text{ must be } (\exists\beta) : \overline{\beta(n, 3) . \bar{\beta}(n, 3)},$$

$$\text{not } \overline{(\exists\beta)\, \beta(n, 3) . (\exists\beta)\, \bar{\beta}(n, 3)}.$$

For we can reason about the characters in a story just as well as if they were really identified, provided we don't take part of what we say as about one story, part about another.

We can say, therefore, that the incompleteness of the 'propositions' of the secondary system affects our *disputes* but not our *reasoning*.

5. This mention of 'disputes' leads us to the important question of the relations between theories. What do we mean by speaking of equivalent or contradictory theories? or by saying that one theory is contained in another, etc.?

In a theory we must distinguish two elements:

(1) What it asserts: its meaning or content.

(2) Its symbolic form.

Two theories are called *equivalent* if they have the same content, *contradictory* if they have contradictory contents, *compatible* if their contents are compatible, and theory A is said to be *contained* in theory B if A's content is contained in B's content.

If two theories are equivalent, there may be more or less resemblance between their symbolic forms. This kind of resemblance is difficult if not impossible to define precisely. It might be thought possible to define a definite degree of resemblance by the possibility of defining the functions of B in terms of those of A, or conversely; but this is of no value without some restriction on the complexity of the definitions. If we allow definitions of any degree of complexity, then, at least in the finite case, this relation becomes simply equivalence. For each set of functions *can* be defined in terms of the primary system and therefore of those of the other secondary system *via* the dictionary.

Two theories may be compatible without being equivalent, i.e. a set of facts might be found which agreed with both, and another set too which agreed with one but not with the other. The adherents of two such theories could quite well dispute, although neither affirmed anything the other denied. For a dispute it is not necessary that one disputant should assert p, the other $\bar{p}$. It is enough that one should assert something which the other refrains from asserting. E.g. one says 'If it rains, Cambridge will win', the other 'Even if it rains, they will lose'. Now, taken as material implications (as we must on this view of science), these are not incompatible, since if it does not rain both are true. Yet each can show grounds for his own belief and absence of grounds for his rival's.

People sometimes ask whether a 'proposition' of the secondary system has any meaning. We can interpret this as the question whether a theory in which this proposition was denied would be equivalent to one in which it was affirmed.

This depends of course on what else the theory is supposed to contain; for instance, in our example $\beta(n, 3)$ is meaningless coupled with $\bar{\alpha}(n, 3) \vee \bar{\gamma}(n)$. But not so coupled it is not meaningless, since it would then exclude my seeing red under certain circumstances, whereas $\bar{\beta}(n, 3)$ would exclude my seeing blue under these circumstances. It is possible that these circumstances should arise, and therefore that the theories are not equivalent. In realistic language we say it could be observed, or rather might be observed (since 'could' implies a dependence on our will, which is frequently the case but *irrelevant*), but not that it will be observed.

Even coupled with $\bar{\alpha}(n, 3) \vee \bar{\gamma}(n)$, $\beta(n, 3)$ might receive a meaning later if we added to our theory some law about the colour of 3. [Though then again $\beta(n, 3)$ would probably be a consequence of or a contradiction to the rest: we should then, I think, say it had meaning since e.g. $\beta(n, 3)$ would give a theory, $\bar{\beta}(n, 3)$ a contradiction.]

It is highly relevant to this question of whether propositions have meaning, not merely what general axioms we include in our theory, but also what particular propositions. Has it meaning to say that the back of the moon has a surface of green cheese? If our theory allows as a possibility that we might go there or find out in any other way, then it has meaning. If not, not; i.e. our theory of the *moon* is very relevant, not merely our theory of things in general.

6. We could ask: in what sort of theories does every 'proposition' of the secondary system have meaning in this sense?

I cannot answer this properly, but only very vaguely and uncertainly, nor do I think it is very important. If the theory is to correspond to an actual state of knowledge it must contain the translations through the dictionary of many particular propositions of the primary system. These will, almost certainly, prevent many 'propositions' of the secondary system from having any *direct* meaning. E.g. if it is stated in the theory that at time n I am at place 1, then for place 2 to be blue at that time n can have no direct meaning, nor for any very distant place at time $n + 1$. If then such 'propositions' are to have meaning at all, it must either be because they or their contradictories are included in the theory itself (they then mean 'nothing' or 'contradiction') or in virtue of causal axioms connecting them with other possible primary facts, where 'possible' means not declared in the theory to be false.

This causation is, of course, in the second system, and must be laid down in the theory.

Besides causal axioms in the strict sense governing succession in time, there may be others governing arrangement in space requiring, for instance, continuity and simplicity. But these can only be laid down if we are sure that they will not come into conflict with future experience combined with the causal axioms. In a field in which our theory ensures this we can add such axioms of continuity. To assign to nature the simplest course except when experience proves the contrary is a good maxim of theory making, but it cannot be put *into* the theory in the form 'Natura non facit saltum' except when we see her do so.

Take, for instance, the problem "Is there a planet of the size and shape of a tea-pot?" This question has meaning so long as we do not know that an experiment could not decide the matter. Once we know this it loses meaning, unless we restore it by new axioms, e.g. an axiom as to the orbits possible to planets.

But someone will say "Is it not a clear question with the *onus probandi* by definition on one side?" Clearly it means "Will experience reveal to us such a tea-pot?" I think not; for there are three cases:

(1) Experience will show there is such a tea-pot.

(2) Experience will show there is not such a tea-pot.

(3) Experience will not show anything.

And we can quite well distinguish (2) from (3) though the objector confounds them.

This tea-pot is not in principle different from a tea-pot in the kitchen cupboard.