Theoricity and Testing

Ariel J. Roffé[1]
Federico N. Bernabé[2]
Santiago Ginnobili[3]

1. Introduction

In the classical conception of scientific theory held mainly within the framework of logical empiricism, the distinction between theoretical and observational concepts played a central role. In the 1970s, several authors began to support the idea that such an absolute distinction should be replaced by a distinction relative to a theory, which would leave aside observability and deal with theoricity by focusing on the different role that concepts play in a specific theory. One of the most influential proposals was made by Joseph Sneed (1971), who can be considered the founder of metatheoretical structuralism (Balzer, Moulines, and Sneed 1987), who proposed to replace the theoretical/observational distinction by the T-theoricity distinction. This second distinction leaves aside observability to pay attention only to the way in which the concepts of a theory are determined. T-theoretical concepts would be those that can only be determined by appealing to the theory T, whereas T-non-theoretic concepts can be determined independently of T.

This distinction was extremely fruitful within the framework of metatheoretical structuralism in reconstructing scientific theories. Specifically, it allows us to clearly show how, when testing scientific theories, it is indispensable to appeal to other underlying theories that allow us to determine the concepts of the "empirical basis" of T independently of T itself. Even with the outstanding conceptual sophistication that philosophers of science have achieved in the move from the naïve theoretical/observational classical distinction to the theoretical/non-theoretical relatively-to-a-theory distinction, there remains a metatheoretical knot which must be disentangled:

i. The identification of theoretical and explanatory terms;

ii. The identification of non-theoretical and theory-testing terms.

[1]National University of Quilmes - National Scientific and Technical Research Council; University of Buenos Aires. To contact the author, write to: arielroffe@filo.uba.ar.

[2]National University of Quilmes - National Scientific and Technical Research Council; National University Arturo Jauretche; National University of Hurlingham. To contact the author, write to: federico.bernabe@unahur.edu.ar.

[3]National University of Quilmes - National Scientific and Technical Research Council; University of Buenos Aires. To contact the author, write to: santi75@gmail.com.

Abreu, C. (ed.) *Philosophy of Science in the 21st Century. Contributions of Metatheoretical Structuralism*.
Florianópolis: NEL/UFSC, 2023, pp. 89–111.

Santiago Ginnobili and Christián Carman (2016) were concerned with showing that the T-theoreticity/T-non-theoricity distinction proposed within metatheoretical structuralism does not coincide with the explanatory/non-explanatory role that concepts have in a theory (i.e., they argued against the first of the above-mentioned identifications). Following this path, in this paper, we will deal with the second identification. We will do three things: (i) Introduce an account of theory testing inspired by the structuralist metatheory, and with it, a new distinction between the concepts that can be used to test a theory (which we call T-testing concepts) and those that cannot (T-non-testing); (ii) argue that the T-testing vocabulary of a theory T does not necessarily match its T-non-theoretical vocabulary; and (iii) provide a few examples to show this, and how the different distinctions interact in explanation and testing.

To do all this, we will proceed as follows. In the second section, we present a brief history of the conceptual identification between theoricity/explainability, on the one hand, and non-theoricity/testing-capability, on the other. We also review how the structuralist metatheoretical program has served to sophisticate and specify the notion of theoricity, emphasizing the most recent contributions that allow us to disaggregate this notion from that of explanatoriness. In the third section we introduce the distinction of T-testability and argue that it coincides neither intensionally nor extensionally with the distinction of T-theoricity nor with that of T-explainability. In addition, we will show how this new distinction allows us to better explain how the testing of a theory works. To illustrate this, in the fourth section, we present two cases of theories that contain terms that are T-non-theoretical and T-non-testing at the same time: natural selection theory and cladistics. The fifth section will introduce some nuance to the considerations introduced before. We will distinguish between strongly T-testing concepts (what we focus on in this article) and weakly T-testing concepts. Section 6 adds some general philosophical considerations and consequences of our approach. Finally, we draw some conclusions.

2. Structuralism and the Refinement of the Theoricity Criterion

The received, syntactic or classical view in the philosophy of science established a very influential distinction within scientific language, between the observational vocabulary, whose sentences are composed exclusively of observational concepts, and the theoretical vocabulary, whose sentences contain at least one theoretical concept.[4] This distinction allows, in the classical conception, to account for the way in which theories are tested (through the hypothetical deductive method) and for how explanation works (through the nomological deductive model of explanation). The idea, specifically, is that the introduction of theoretical concepts makes it possible to more adequately *explain* the behavior of certain phenomena described through observational concepts, and that observational concepts make it possible to *test* the theoretical statements made within the framework of a theory.

It is straightforward to understand why the classical philosophers of science, being empiricists, considered that the theoretical/observational distinction performs this dual work. In Carl Hempel's

[4]If they contain only theoretical concepts, those sentences are said to be "pure", or "theoretical" proper. In the classical metatheory, laws contain only these kinds of sentences. If a sentence contains both theoretical and observational terms it is called "mixed". Mixed sentences act as a bridge between the theoretical laws and the observational terms describing observed "facts", i.e., they give theoretical concepts an empirical meaning. See below for more on this.

words: "[s]cientific systematization is ultimately aimed at establishing explanatory and predictive order among the bewilderingly complex 'data' of our experience, the phenomena that can be directly 'observed' by us" (Hempel 1958, p. 41). In his view, those "empirical facts", "data of our experience" or "phenomena that can be directly observed by us" must take the form of *observational statements* (i.e., statements that refer only to observable entities). Laws, on the other hand, may contain concepts that refer to non-observable (i.e., theoretical) entities. Conversely, these laws and theories must be tested by comparing their observational consequences with the (observational) statements we accept, based on our observations. Thus, the empiricist account of both explanation and theory testing depends crucially on the distinction between these two supposedly independent vocabularies, which would have distinct functions in science: observational vocabulary would serve as an independent *judge* for our theories, while theoretical vocabulary would be used to *explain* the data we collect by observing the world that surrounds us.[5]

The theoretical/observational distinction has been widely criticized, primarily by "historicist" philosophers of science, e.g., Feyerabend (1962), Hanson (1958) and Kuhn (1962). These authors formulated their criticisms in different ways, but the common thread among them is the position that all observation (and observational language) is *theory-laden*. Even our most immediate experience is already "conceptualized". So, according to them, there is no theory-independent firm basis of observational statements from which we can judge the adequacy of our theories. Furthermore, the "conceptual scheme" we use to "organize" the sense data is not fixed/universal. We can learn to see things in a certain way. Scientific education, these critics would say, consists in large part of educating our perception (Hanson 1958). But, if theories can teach us to see what we use to test them, then there is a risk that theory testing becomes circular. In other words, if the concepts of the empirical basis of a theory T are loaded by T, then every test of T will have to presuppose this theory, in order to conceptualize the "empirical" phenomena used to build the sentences with which we test it.

These objections pushed philosophers to abandon the observational/theoretical distinction, but the problem of circularity remained. In the paper in which he abdicates the standard conception of scientific theory, Hempel (1970) proposes an alternative solution, by modifying the theoricity criterion in two aspects. Firstly, he made the distinction relative to particular theories (i.e., a term may be theoretical in one theory and non-theoretical in another). And secondly, by replacing the idea that the empirical basis of theories must be described in *observational* terms with the condition that it be described using only *previously understood* (and therefore interpreted) terms. The (now called) *antecedent* vocabulary of a theory T may thus be loaded with theory and not be observational, but the theories with which it is loaded must be *temporally* prior to the formulation of T—and hence, applicable independently of T.

Similarly (although independently, since he had not read Hempel's paper), Sneed (1971) proposed the T-theoretical/T-non-theoretical distinction. Like Hempel's, the structuralist criterion is relative to particular theories. Also, like Hempel's, it postulates that T-non-theoretical terms have to be *prior* to T. The difference lies in what Sneed understands by "prior". According to him, what

[5]Strictly speaking, in the classical metatheory, observational and theoretical *statements* perform those functions. Concepts, by themselves, do not explain or act as judges of anything. We will continue to speak of explanatory, explained, testing, etc. concepts, in an indirect sense, by considering the roles that the statements (or theoretical structures) formulated with them play.

matters is not *temporal* precedence, but rather *operational* precedence.[6] That is, T-non-theoretical terms are those that can be operationalized or determined independently of T, while T-theoretical terms are those that *require* using the laws of T to be operationalized. As we shall see below, the possibility of independent operationalization is what guarantees that theories can be tested against something independent of themselves, avoiding the circularity issue.

More precisely, a term *t* is T-non-theoretical if and only if there exists a determination method for it that does not presuppose T; a term is T-theoretical if and only if *every* determination method for it presupposes T. Put simply, *determining* a term means finding out its denotation. A *particular* determination is said to be "T-dependent" if it uses the laws of T to find out the denotation of *t*. A determination *method* for a term *t* is a way of systematically determining the denotation of *t*. A determination method will presuppose a theory T just in case every one of its instances (every particular determination following this method) is T-dependent, since the whole procedure will thus presuppose that T is true/empirically adequate/justified.

The structuralists have also developed a second criterion of theoricity, sometimes referred to as the "formal" criterion (Balzer 1983; 1985; Balzer, Moulines and Sneed 1987; Gähde 1983). To avoid confusion, we will call the terms that are T-theoretical in this second sense "T-determinable" (the reason for this will be clear in a moment), reserving the term "T-theoretical" for the first criterion. A term *t* is T-determinable if and only if there exists a determination method for it that presupposes T (i.e., a T-theoretical determination method). That is, T-determinable concepts are those that *can* be operationalized by using the laws of T. T-non-determinable concepts are those that cannot—i.e. those that *have to* be operationalized by appealing to something outside the theory in question (for some application examples see Section 4, as well as Balzer, Moulines and Sneed (1987, pp. 73-78)).

The structuralist T-theoricity distinction is a notable sophistication over previous accounts and, more importantly, it has been applied in the reconstruction of countless scientific theories from different scientific disciplines. However, in the standard structuralist account, there remained a link between T-non-theoretical/T-explained/T-testing vocabularies, on the one hand, and T-theoretical/T-explanatory/T-non-testing vocabularies, on the other. This identification is similar to the one present in the classical observational-theoretical distinction: T-theoretical concepts are introduced to explain the behavior of phenomena described by means of T-non-theoretical terms, and T-non-theoretical concepts allow the theory to be tested. In more technical terms, the global empirical basis of a theory coincides with its global *explanandum*.[7]

One clear instance of this identification can be found in the following passage by Díez:

[6]In many cases, these two distinctions will in fact coincide since the vocabulary that is temporally previous to some theory T will also tend to be operationally independent from it. However, it is possible for a researcher to, *at the same time*, recognize a phenomenon to be explained and give an explanation for it. More importantly and frequently, as Hempel (1970) himself notes, there are cases where some *term(s)* are previously available, but the meaning of the concepts they denote change, making the application of the criterion very difficult in practice (how much conceptual change is necessary to consider that we are not dealing with the same concept?). With Sneed's criterion these problems do not arise.

[7]The notion of "global empirical basis"—proposed by Pablo Lorenzano (2012)—would be analogous to the notion of "empirical basis" used in the classical conception, in the sense that it is a concept relative to a theory in general and not to its particular applications. And, in the same sense, the "global *explanandum*" of a theory should be understood as a way to speak of the systems whose behavior theory intends to explain, and not to refer to particular explanations (Ginnobili and Carman 2016).

> We will call T-testing, or T-non-theoretical, or T-empirical vocabulary, that part of the characteristic vocabulary of the theory T used in the description of the "T-data", that is, of the "phenomena" of which the theory wants to account for (explain/predict) [...] We will call T-explanatory, or T-theoretical, the characteristic vocabulary of T that is not T-testing vocabulary, that is, the concepts used in the formulation of the laws of T, which cannot be determined/measured without presupposing the validity of any of these laws. (Díez 2012, pp. 68-69, our translation)

Let us begin with the identification of the first two respective vocabularies (theoretical and explanatory, non-theoretical and explained). Although structuralism *per se* does not provide an account of scientific explanation, there are some conceptions of explanation that take the structuralist metatheory into consideration (Bartelborth 1996; Díez 2013; Forge 2002), since knowing how theories are logically structured is extremely relevant for understanding how they explain phenomena.

Take for example Díez's account, known as *ampliative, specialized embedding*. Succinctly, in this view, *explananda* are conceptualized as data models of the form $DM = \langle D_1, \ldots, D_n, f_1, \ldots, f_i \rangle$ where the Ds are domains of objects and the fs are relations and functions over those domains. Explaining a phenomenon means embedding its DM representation into a theoretical model TM, such that: (i) DM is a substructure of TM (i.e., TM contains every domain, relation and function in DM, plus some others); and (ii) TM satisfies some theoretical laws, which restrict the possible interpretations of the concepts, and in that way make some of the *explananda* phenomena expected. The details of this proposal, which include some additional criteria to distinguish between adequate and inadequate explanations, do not matter here. What matters to us is that, according to Díez's initial view (Díez 2002), every new concept that TM introduces to account for DM must be T-theoretical.[8] Thus, as one can readily see, the T-theoretical concepts are the ones that play explanatory roles (and in that sense can be called "T-explanatory"), while the T-nontheoretical concepts figure exclusively in the *explananda* of theory T (and therefore can be called "T-explained").

This identification of the T-theoretical and T-explanatory vocabularies was put into question by Ginnobili and Carman (2016), also in structuralist terms. These authors argued, firstly, that the T-theoretical status of a concept could change over time without its explanatory role changing. For example, scientists could find some new ways of operationalizing "mass" and "force" independently of classical mechanics. If this happened, "mass" and "force" would become T-non-theoretical for classical mechanics but their explanatory role over the movements of particles would remain the same within that theory. The only reason to think otherwise would be an *a priori* identification of T-theoretical and T-explanatory vocabularies. Secondly, and perhaps more importantly, these authors showed that there are clear and convincing examples of theories that, to account for their global *explanandum*, expand conceptually by appealing to *both* theoretical and non-theoretical concepts, in some cases, and *only* to non-theoretical concepts in others. In both cases T-non-theoretical concepts can figure as part of the T-explanatory concepts of the theory.

In the following sections we go deeper into Ginnobili and Carman's path and show that, in the same way that the distinctions between T-theoretical/T-non-theoretical and T-explanatory/T-

[8] Later on, and partly due to discussions with Ginnobili and Carman (see below) he weakened this point, and only demanded that at least one new concept be T-theoretical (Díez 2013).

explained concepts are not the same, the T-testing/T-non-testing distinction is also independent of both of them.

3. Theory Testing and T-Testing Vocabulary

In this section, we provide a sketch of an account of theory testing inspired by the structuralist metatheory (and which, we believe, is implicit in a lot of structuralist practice), that will allow us to characterize more appropriately the distinction between T-testing and T-non-testing concepts, and to show how this distinction is not identifiable with the other two distinctions introduced above.

Put simply, a particular test of a theory consists in a pair of determinations of a term t, one of which is T-dependent and the other is T-independent. In other words, to evaluate a theory one must determine (at least) one term by using a determination method that presupposes T and another that does not presuppose T. The test is said to be successful if both determinations result in the same value for t (or in a value that is close enough given some standard of approximation), and unsuccessful otherwise.[9]

To illustrate this in a simple manner, consider the following example. In classical genetics (CG),[10] the non-theoretical level is comprised of (among other things) the traits of the organisms of the breeding population, and their proportions. That is, one can determine the frequencies of traits in a population without applying the laws of classical genetics. For instance, if the trait being investigated is the height of a particular species of plant, then there exists a determination method that does not involve using the laws of CG (e.g., using a ruler) for determining plant height.[11] On the other hand, determining gene and genotype frequencies *requires* applying the laws of CG. That is, the *only way* of knowing that an organism has a given genotype, or a population a given distribution of genotype frequencies, is to apply CG itself.[12] In a typical case of application, one begins with a set of trait frequencies for two subpopulations at t_1 (say, a distribution of plant heights); one then postulates a given genetic architecture (*loci*, allele-types, etc.) and establishes what the trait frequencies would look like at t_2 if that were the case—i.e., one performs a CG-theoretical determination of trait frequencies. Lastly, one measures plant heights at t_2—i.e., performs a CG-non-theoretical determination of trait frequencies—and sees if the value coincides with the CG-theoretically determined value. If it does, then the postulated values for genotype frequencies are taken to be adequate, if it does not, then they are discarded. It is also easy to see that, in the first case, this would not only confirm that a given genotypic architecture is present, but also represent a confirming case for CG. On the other hand, if the values did not coincide, then we would be in front of a Kuhnian unresolved puzzle, which, if failed to be subsumed after repeated attempts—e.g.,

[9]This would correspond to what, in section 5, we will call a *strong* test. There may be other types of tests. See below for more on this.

[10]Structuralist reconstructions of CG, and treatments of the issue of T-theoricity there, can be found in Balzer and Dawe (1997) and Lorenzano (1995).

[11]In this particular case, the trait looks “observational” in the sense that measuring something with a ruler does not seem to use any theory. But we could have chosen a trait that requires some more independent theorizing to be determined (e.g., blood type, which does satisfy mendelian inheritance).

[12]Or at the very least, this was the only way of knowing that at the time Mendel proposed his laws, and at the beginning of the twentieth century.

after postulating different possible genetic architectures—, could become an anomaly.

To put our account in enunciativist terms, testing a theory would involve making a derivation for *both* the theoretical and non-theoretical determination methods.[13] That is, instead of thinking of theory testing as an instance building an argument and then "observationally" checking whether the conclusion holds, the picture would look more like building two arguments:

$$\frac{\begin{array}{l} L_1, \dots, L_n \\ C_1, \dots, C_m \end{array}}{E_1} \qquad\qquad \frac{\begin{array}{l} L^*_1, \dots, L^*_i \\ C^*_1, \dots, C^*_j \end{array}}{E_2}$$

Where $L_1, \dots, L_n/L^*_1, \dots, L^*_i$ are (two different sets of) laws, one of which includes laws of T and the other of which does not, $C_1, \dots, C_m/C^*_1, \dots, C^*_j$ are (different sets of) initial conditions, and E_1/E_2 represent the occurrence or non-occurrence of the same phenomenon. This would further illustrate the holistic character of theory testing (i.e., theories are never tested against "raw" experience, but only against a background that includes other theories). Also, as in the classical account, our perspective assumes that the laws of the theory perform the function of connecting independent areas of our experience. In our case, the laws perform this function within the T-dependent determinations that use them.

With all this in mind, it is straightforward to see which part of the vocabulary of a theory is relevant for testing it. If testing a theory consists in determining a concept both theoretically and non-theoretically, then the concepts that can be used for that purpose are those that can be determined in both ways. In terms of the distinctions introduced above, a concept will be said to be T-testing if and only if it is T-determinable and T-non-theoretical. On the other hand, a concept will be T-non-testing if it is either T-theoretical or T-non-determinable. This distinction is interesting because, as we will show in the next section with concrete examples, some concepts are both T-non-theoretical and T-non-determinable (and thus T-non-testing). Put more simply, that there are cases where some part of the empirical vocabulary of the theory is not relevant for testing it. Consequently, even though the T-theoretical vocabulary is always T-non-testing, the theoricity distinction does not collapse with the testability distinction because the vocabulary that is used to test a theory may be a proper subset of the empirical (non-theoretical) vocabulary. To put it graphically, the relations between T-(non)-theoretical, T-(non)-explanatory and T-(non)-testing vocabularies would look as follows.[14]

[13] If we equated determining a concept with making an inference, which is doubtful. At least some accounts of determination methods would not agree here. For example, Roffé (2020c) recently proposed that determination methods should be thought of as algorithms, and following an algorithm and making an inference are two different things.

[14] It may surprise the reader that we include a space for T-explanatory and T-non-determinable concepts (which must then also be T-non-theoretical, since all T-theoretical concepts are T-determinable, see section 5). That is, we reserve a space for concepts that the theory adds to account for its *explananda*, and that cannot be determined by the theory itself. The next section will also give some examples of this.

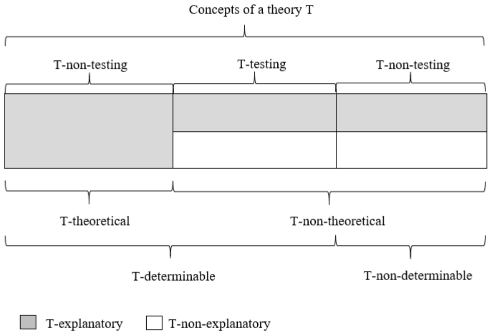

Figure 1: Relations between the three distinctions.

Before moving on to the examples just mentioned, it is worth making some technical precisions to our proposal. In the standard structuralist account (e.g., the one presented in Balzer, Moulines and Sneed 1987) the T-non-theoretical language of a theory T is represented as a class of models M_{pp} or "partial models", which results from eliminating the T-theoretical concepts from the class of "potential models" M_p (which represents the entire language of the theory). The intended applications (I) of the theory are presented as a subset of the M_{pp}. If Ginnobili and Carman are right, then I is not a subset of M_{pp} (since the representation of the global *explanandum* phenomena may not contain all the T-non-theoretical terms). Rather, one may introduce a class of models M_e as a substructure of M_{pp}, containing only the explained vocabulary, and I will be a subset of M_e. In that way the distinction between T-non-theoretical and T-explained vocabularies is captured formally. In our case, if we call M_d the class of models that contains the denotations of the T-determinable concepts (another substructure of M_p), then the class of T-testing models (call it M_t) can be obtained as the substructures of M_p that contain the concepts present in both M_{pp} and M_d. Visually:

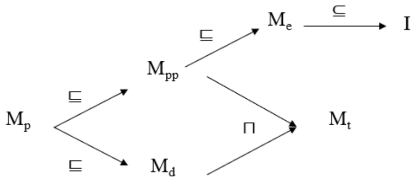

Figure 2: New proposed landscape of the structuralist classes of models.

4. Cases

In this section we present two theories as cases of application of our distinction, which contain terms that are at the same time T-non-theoretical and T-non-testing, namely, natural selection theory (NST from hereafter) and cladistics (CLAD from hereafter). Before going into that, a brief consideration of the strategy we will employ may be useful going forward. The key for upholding that those terms are T-non-testing will be to defend that they are T-non-determinable. The *prima facie* obvious way of arguing that a term t is T-non-determinable would be to show that, for any arbitrary model of the theory, even if the denotations of all the other terms are known, the laws of the theory would still not allow us to infer a univocal denotation for t, and hence that there would always be at least two possible values for it that render the laws true. However, things are more complicated than that, because a concept (e.g., a domain D) might figure as part of the signature of another concept (e.g., a function $f : D \to X$). In such cases, it would be logically impossible to know the extension of f but not of D. Hence, the way to argue that a concept c is T-non-determinable will be to show that, for any arbitrary model of the theory, even if the denotations of all other terms *that do not depend logically on c* are known, the laws of the theory would still not allow us to infer a univocal denotation for t.[15]

In what follows, we provide informal (or semi-formal) reconstructions of NST and CLAD and argue mostly informally about the T-non-determinability of the terms in question. This will be enough for, at least, giving some plausibility to our theses. A fuller, more formal and more rigorous examination of the cases below would be desirable, but falls outside the scope of this writing (the respective formal reconstructions of both theories can be found in Ginnobili (2012; 2018) and Roffé (2020b; 2020d)).

4.1. Natural Selection Theory

Ginnobili and Carman (2016) appeal to NST to show that explanatory conceptual ampliation does not always appeal exclusively to theoretical concepts. Here we will appeal to the same theory to show that not all NST-non-theoretical concepts are NST-determinable, and therefore, that there are NST-non-theoretical terms that are NST-non-testing.

The nature and structure of the theory of natural selection have been an important topic of discussion in the philosophy of biology. For the point we want to discuss we will rely on Ginnobili's structuralist reconstruction of it (Ginnobili 2010; 2011; 2016; 2018).

Darwin proposed the theory of natural selection to explain how populations of organisms evolve adaptively. That is, how they acquire traits that allow them to succeed in the environments in which they live. For example, the explanation of how certain birds have acquired a coloration pattern that camouflages them in their environment is as follows: In the past, there was a population of birds with different colors of plumage. Those that were more similar to the place where they fed were longer-lived, because predators visually confused them with the background, and left more offspring that in turn carried that trait. Generation after generation, new variations emerged that increased the resemblance of the pelage with the environment and these were spread by having a greater reproductive success, until we reached the current population (Darwin 1859, pp. 84-85). This historical explanation (in the sense that it appeals to changes over many, many generations) appeals to iterations of what Ginnobili considers the fundamental law of the theory, which schematically states that in a given generation, those organisms that carry certain traits will have greater reproductive success.

[15] A concept c_1 is logically dependent of another c_2 if and only if it is *logically* impossible to know the extension of c_1 without knowing the extension of c_2 (for example, because c_1 is a function that has c_2 as a domain).

Here we will not include the debatable details of the reconstruction offered. Suffice it to point out that, according to Ginnobili (2016, pp. 18-19), the fundamental law of the theory would be the following (we modified slightly it to fit our terminological uses):

[(The trait t_1 is more effective than trait t_2 in performing function f in environment $e \rightarrow$ organisms that possess t_1 are fitter than those who possess t_2 in e) and t_1 and t_2 are inheritable] $\rightarrow$ the organisms that possess t_1 will be more successful in differential reproduction than those that possess t_2 in e.

Where t_1 and t_2 are two variants of a trait-type T—e.g., 1 m and 1.10 m for the trait-type of the length of the neck of a giraffe—, f is a particular biological function—e.g., reaching the higher branches of trees—and e is a particular environment—e.g., the African savanna in a period of scarceness.

Although we are not going to present the full formal reconstruction of NST (for that see Ginnobili (2012; 2018)), it is useful for the purposes of our work to list of some of its fundamental concepts.

- O is a set of organisms of a given population to which NST is applied.
- T $(=\{t_1, t_2, \dots t_n\})$ is a set of trait-variants of the same trait-type.
- E is the set of different environments, and e is a distinguished individual (a particular environment) within it.
- *DESC* is a function that assigns a trait to a particular organism.
- H (heritable traits) is a subset of T. Ginnobili (2012; 2018) introduces it from a definition and not as a primitive concept.
- F is a set of (biological) functions, and f is a distinguished individual within it.
- *EFEC* is a 4-ary relation that establishes an order in the effectiveness with which a pair of traits perform a function. $EFEC(t_1, t_2, f, e)$ symbolizes that t_1 is more effective than t_2 in performing f in environment e. It is a comparative concept, not a metric one.
- *FIT* (fitness) establishes an order among of different kinds of organisms of a generation in the particular environment in which they are found. It is also a comparative, non-metric concept.
- *RS* (reproductive success) establishes an order among of different kinds of organisms of a generation in the particular environment in which they are found. Ginnobili (2012; 2018) does not introduce it as a primitive concept, but rather as being defined from the mathematical language presupposed by the theory.

What this theory explains is why certain types of organisms in a population have greater reproductive success than others. The intended applications of the theory, therefore, are organisms in a population that differ in the possession of a trait and that differ in their reproductive success in a given environment. Thus, *e*, *E*, *O*, *T*, *RS* and *DESC* allow us to describe the global *explanandum* of the theory (M_e in the previous section). The explanation consists in pointing out that organisms that possess a certain trait, which performs a function more effectively, improve their fitness in that environment, thus improving, if the trait is heritable, their reproductive success. The concepts with which the intended applications are conceptually enriched, and are thus T-explanatory, are: *f*, *F*, *H*, *EFEC*, *FIT*.

Ginnobili and Carman (2016, following Ginnobili 2011) argue that this is at least a case of mixed conceptual extension, because the functional attribution, the effectiveness with which a function is fulfilled, and the heritability of the trait can all be determined outside of NST. They even argue that this could be a case of purely non-theoretical explanatory conceptual extension, arguing that the concept of fitness could be considered as NST-non-theoretical, since its different specifications can be determined independently of the theory. For our purposes it is not necessary to discuss this point. What we must ask ourselves is which of the non-theoretical concepts appearing in the fundamental law of NST are NST-determinable, and consequently, serve to test NST, i.e., are NST-testing.

The most obvious prediction that can be made with the theory has to do precisely with the determination of its *explanandum*. That is, with predicting or retrodicting which kinds of organisms will have greater reproductive success. Reproductive success (as a comparative concept) can be ascertained in an NST-theoretical way and can also be determined in a non-theoretical way (since one can simply count the number of organisms of each kind that are present in each generation). Therefore, it is a T-testing concept. Even when it is not a traditional way of testing the theory, it is possible to think of NST-theoretical determinations of heritability. If we had all the other concepts of the law determined we could find out whether a trait is heritable or not by determining whether the trait affects the reproductive success of its possessors. The same can be said of the effectiveness with which the function is performed. Having all the other concepts of the law determined, we can establish—perhaps in a non-deductive way—which variant of the trait best fulfills its function by determining the reproductive success of its possessors.

What about the rest of the concepts? There remain *e*, *E*, *f*, *F*, *O*, *T*, *DESC*, and *FIT* to consider. Let us leave aside the question of the determinability of the concept of fitness, which would involve a discussion beyond the scope of this article.

The functional concepts *f* and *F* are also somewhat controversial, and the answer depends on the account of functions one adopts. For instance, some people have argued that the function(s) of a trait is just the effect(s) that it has been selected for in the past (Millikan 1989; Wright 1973). Ginnobili's reconstruction is incompatible with this approach since the theory contains functional notions, and consequently that definition of function would be circular. Here, we will assume this reconstruction to be adequate. If that is the case, then the notion of function (more specifically, the concept that states that a certain trait fulfills a certain function) would be NST-non-determinable, and consequently, NST-non-testing. There is no way to find out from NST what the function of a *feature* is. This is knowledge that comes from physiological and behavioral studies. It is important to note that this does not depend on reconstructive decisions made. As we presented the conceptual framework, what we must say is that the distinguished element *f* cannot be determined from NST (i.e., we cannot know which, out of all the possible functions of the trait, is the relevant one). However, we could have introduced a specific concept for functional attribution. In that case we would say that it is impossible to determine that concept. The point is that, as we said, it is not possible to perform the functional attribution only by considering NST.

Moving on to other concepts that can be found in the law, the terms that represent the environment, particularly *e* (the actual environment), is clearly T-non-determinable. Firstly, as the vocabulary is presented above, it is a domain in *EFEC*, *FIT* and *RS*, so the question would be if knowing what the organisms and their (heritable) traits are, as well as what the function at stake is, would be sufficient for establishing what the selective environment consists of. And the answer is obviously no. For instance, knowing that in a particular application of NST there is a population of giraffes, some with longer and some with shorter necks (both heritable) and that the relevant function is feeding, does not permit us to infer that the environment is one in which the leaves are high and the food is scarce. It might have consisted in any number of other scenarios. For instance, it might have been the case that giraffes with longer necks could submerge their heads deeper in the water and catch algae or fish better, or that they could see farther away to find food sources, etc.[16]

Another clear case is the function that assigns traits to organisms (*DESC*), which allows us to partition the population into varieties and consequently, to predict differences in reproductive success among these varieties. One could argue that knowing what the relevant traits are and which is more effective, and seeing which particular organisms are having greater reproductive success could allow us to assign traits to organisms. But this has at least two problems. Firstly, the law only states that one *kind* of organism has greater reproductive success than the other, not that every individual organism of one kind has greater reproductive success than every individual organism of the other kind. However, *DESC* assigns traits to individual organisms. Secondly, and more importantly, *RS* implicitly contains *DESC* in its definition, since the kinds of organisms that have differential reproductive success are defined as groups of organisms that share the possession of a common trait (which presupposes *DESC*). Therefore, *DESC* can in no way be determined from NST for a particular organism.

Finally, *O* and *T* are also not very clear cases of T-non-theoretical and T-non-testing concepts. Not because they could be T-determinable. They clearly are not, since almost every other concept in the theory depends on them (has them as a domain), and thus without having both of them determined we cannot determine almost anything else. This is a common characteristic of the most basic domains of theories (think, for example, about the set of particles in classical mechanics). In that way, they are clearly T-non-testing. What is doubtful is whether they should be considered T-non-theoretical. We will expand on this in the next section.

If this analysis is correct, NST would be a case where none of the above distinctions is coextensive. There are explanatory NST-non-theoretical concepts (which are not part of the global *explanandum*) and NST-non-theoretical concepts which are not determinable and therefore are not part of the testing basis of the theory.

4.2. Cladistics

Even though Darwin's most well-known development is NST, it is not the only theory he proposed, nor the one he considered to be his most important contribution. While the iteration over many generations of NST explains the presence of adaptations, it does not account for the possession of certain *structurally* similar (and sometimes functionally very dissimilar) traits that biologists call *homologies*. A famous example is the tetrapod limb, which has approximately the same bones, arranged in approximately the same pattern, in a wide variety of species (for example, in humans, bats, whales and moles), even though they serve for widely different purposes in each (grabbing, flying, padding and digging).

Prior to Darwin, one popular explanation was that organisms were structurally similar in this sense because God created every species (or at least every vertebrate) parting from a common plan or *archetype* (see for example Owen 1847; 1849). Darwin realized that the archetype was not an idea in the mind of a God, as Owen had thought, but an actual ancestor (Darwin 1872, p. 384).

[16] These considerations would also hold even if *EFEC*, *FIT* and *RS* did not have *e* as a domain (i.e., if we had chosen to somehow present the vocabulary and the laws in a different way).

Thus, the fact that organisms share homologies is indicative of, and indeed *explained by*, the fact that they share a common ancestor. At some point in time, some subpopulations of this ancestor became reproductively isolated and their traits evolved independently by adapting to the local environments, but preserving, however, the general structure of the ancestor's trait. In this way, the evolution of life on Earth can be depicted by a tree that parts from a single root species and subsequently divides into the rich diversity of species found today.

Moreover, even at that time, it was obvious to biologists that some organisms share some homologies among themselves that they do not share with other organisms, and that this pattern is nested. For instance, all spider species share some homologies that scorpions do not; and in turn, spiders and scorpions share many homologies that shrimps do not. The darwinian explanation for this is that the most recent common ancestor of spiders is not an ancestor of scorpions (i.e., spider species diversified among each other later than they did with the ancestor of all scorpions), and the same goes with spiders + scorpions and shrimps. In other words, a particular tree (a subtree of the entire tree of life) explains the particular (nested) *distribution* of homologies among these species.

However, not everything is as easy as it may seem from the paragraphs above. Many times, homologies (in the sense of structurally similar traits) do evolve independently in separate lineages, i.e., *convergence* does occur in nature.[17] And this, in turn, can obscure what the phylogenetic relations between a set of species are. Consider for example the following very simplified example:

Species/Character	C_1	C_2	C_3	C_4
S_1	1	1	1	0
S_2	1	1	0	1
S_3	0	0	1	1
S_4	0	0	0	1

Table 1. Example data matrix (homology distribution scheme) for four species S_1-S_4 and three homologous characters C_1-C_4 , each with two alternative states codified by 0 and 1.

Here, according to characters C_1 and C_2, S_1 and S_2 share a common ancestor that is not an ancestor of S_3 and S_4 (and vice-versa, since the latter two share state 0). However, according to C_3, S_1 and S_3 have an ancestor that is not an ancestor of S_2 and S_4 (and vice-versa). C_4, on the other hand, seems to suggest that S_2, S_3 and S_4 have a common ancestor that is not one of S_1. In other words, we have (among others) the following three possible hypotheses of relatedness:

[17] This is sometimes expressed by saying that *primary* homologies—i.e., structurally similar traits—are not always *secondary* homologies—traits inherited from a common ancestor—(see for example Blanco, Roffé and Ginnobili 2020; Pearson 2010; 2018; de Pinna 1991; Roffé, Ginnobili and Blanco 2018). Here we stick to the terminological choice of using the term "homology" for primary homologies, though we intend nothing of weight with this.

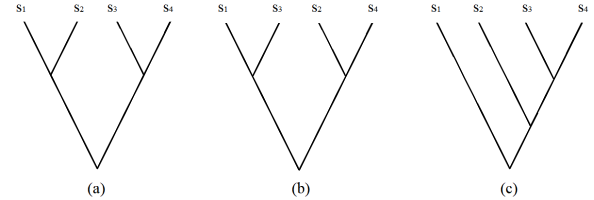

Figure 3: Three of the fifteen possible trees for 4 taxa.

The question is, how do we choose among them? How do we decide which character state(s) are the convergences and which are not? In Darwin's time there was no systematic procedure for doing this. This is what can be achieved with the methods of cladistics.

In the cladistic methodology, each character is mapped into each tree to see how many evolutionary changes one would need to postulate to account for the currently "observed" distribution. The score or *length* of a tree is simply the sum of the changes needed to account for every character under consideration. In the example from Figure 3 above, the length of tree (a) is 5, while the lengths of (b) and (c) are both 6, making (a) preferable to them—one out of all the possible optimal character mappings is shown in Figure 4 below.

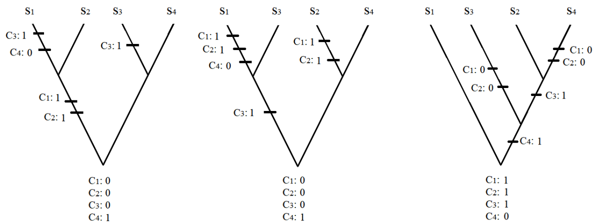

Figure 4: Character optimizations for the three trees shown in the figure above.

Notice that, despite first appearances, character C_4 is actually uninformative for phylogenetic purposes, since it accommodates equally well in every possible tree. In trees (a) and (b) it suffices to postulate that the root ancestor had state 1 and that its state changed in the branch leading to S_1 (and we could have even done the same in tree (c)).

The tree that is inferred as the actual one, among all possible trees, is simply the minimum length tree (or one of them, if there is more than one). There are many complications to the general sketch just presented (for more complete presentations of the theory the reader may see Kitching *et al*. (1998) and Wiley and Lieberman (2011)) but the above will be enough for our purposes.

Moving on to how to reconstruct this theory, its concepts can be formally represented in the following way.

- T, a set of taxa (the set of species whose phylogenetic relations will be inferred).

- A, a set of trees, such that each tree is a graph containing the members of T as leaves, and satisfying some formal properties (each tree is directed, acyclic, rooted, etc.). Note that A is univocally defined for each T.
- $a_R \in A$, a distinguished individual (a particular tree) that represents the actual evolutionary history (the one we wish to infer as the correct one).
- C $(= \{C_1, ..., C_j\})$, a set of characters, such that each character might be further thought of as a set of states (i.e., $C_i = \{s_1, ..., s_i\}$).
- $DESC$, a function that assigns a state for each character C_i to each taxon in T. Note that T, C and $DESC$ together comprise the data matrix shown above in table 1.
- LEN, a function that takes an input tree and a character assignment and computes the length of the tree.

Given all this, the fundamental law of CLAD would simply state that, given a set of taxa T, of characters C and an assignment $DESC$, the actual evolutionary tree a_R is among the minimum length (LEN) trees (once again, this is a semi-formal and very simplified version see Roffé (2020d)).

As said above, what cladistics explains is the "observed" distribution of homologies (i.e., shared character states) among a set of taxa. Thus, the CLAD-explained vocabulary would consist of T, C and $DESC$. To account for this distribution, the theory extends that vocabulary with a set of trees, a function to compute the length of such trees and the actual tree. This is the CLAD-explanatory vocabulary.

Of course, the explained concepts T, C and $DESC$ are CLAD-non-theoretical; the data matrix is typically built previously and independently to the beginning of the phylogenetic cladistic analysis (in fact, we will argue that it is always built that way).[18] The actual evolutionary tree a_R is CLAD-non-theoretical as well. Even though in the usual cases of application it cannot be determined independently of CLAD because the relevant evolutionary events are in the deep past, this is only an empirical limitation not a conceptual one. There are, in fact, applications to experimental phylogenies where the actual tree is known independently (for more on this see Hillis *et al*. (1992) and Roffé (2020b)).[19]

A and LEN have a more confusing T-theoricity status. At first glance, one might think that these are the T-theoretical concepts of the theory. However, they are both *defined* functions. For instance, given a set of taxa, the set of trees is automatically determined as the set of all possible graphs with certain mathematical characteristics. Thus, the determination of trees and lengths does not seem to presuppose the fundamental law of CLAD, and the extensions of both can be determined for any possible set of taxa and assignments, even if the law were false (the resulting minimum-length tree did not coincide with the actual one).[20] In that sense, even if their definitions

[18] See Roffé (2020a) for a fuller discussion of this point in the context of dynamic homology.

[19] This characteristic is common of many other theories. Some particle trajectories may also be in the past, for example, and thus only be determinable CM-theoretically. That does not mean that trajectories are theoretical for classical mechanics, since there are other applications in which trajectories can be determined non-theoretically. That a concept is T-non-theoretical does not mean that it should be determinable independently if T *in every application* of the theory. This point, however, tends to confuse philosophers of systematics (and systematists themselves) more than physicists, since actual trees are *almost always* in the past.

[20] If, however, one included the definitions as part of the fundamental law, this would not hold. This does not seem like a very reasonable option though.

are part of the theory, they seem to be CLAD-non-theoretical, and we would again be in presence of purely T-non-theoretical conceptual extension.

Going back to the main subject of this paper, to answer which of these concepts are T-non-testing, we need to examine which are T-determinable. The most obvious T-testing concept is a_R. As said above, it can be determined independently of the theory, and it can be determined with it by finding the minimum length tree and applying the fundamental law.

A and *LEN* can be determined simply by applying their definitions. If this counts as a CLAD-theoretical determination is, once again, doubtful, and we will not discuss it here in greater length.

One case of a clearly T-non-theoretical and T-non-testing concept is the function *DESC*, the assignment of states to the terminal taxa. That it is T-non-testing stems from the fact that it is T-non-determinable. One could think that there are some applications in which having the data matrix and a_R determined does induce a unique assignment of states to the terminal taxa. For instance, given a character C_1 comprised of two states, 0 and 1, and the following tree:

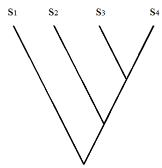

Figure 5: Actual tree which could be thought to (but does not) result in a univocal assignment of states to taxa.

One might think that the only possible interpretation is the one that assigns 0 to S_1 and S_2 and 1 to S_3 and S_4. However, the reverse is also possible, also with a minimum length of 1. And it is easy to see that the same procedure (inverting the 0s and the 1s) can be done for each possible assignment, obtaining another one that yields the same length for every possible tree. Since every change (from 0 to 1 and from 1 to 0) counts the same, the corresponding trees will have the same length.[21] Furthermore, this is not just a change of codification/scale. Even if one fixes in advance what a 0 and a 1 mean (for example, the absence and presence of a given morphological structure, respectively) one cannot know solely by applying CLAD which taxa have the feature and which do not.

Finally, *T* and *C* are as organisms and traits in NST, they are the most basic domains of CLAD and one cannot do anything with the theory without them (such as determining assignments, lengths, trees, etc.). Therefore, they are CLAD-non-testing as well.

In the next two sections, we consider other ways of testing theories, and consider some additional general points and suggestions, and finally, in Section 7, we draw some conclusions.

[21] If there are more than two states, the procedure is analogous. Things start getting more complicated when one allows for non-uniform cost transformation schemes, but this is almost never done in phylogenetic practice, so we keep the uniformity assumption for simplicity.

5. Weak T-determination

In Section 3, we characterized a test of a theory as a pair of determinations of a term t, one of which is T-dependent and the other is T-independent. This would actually correspond to the strongest way in which a theory can be tested, which is when the theory makes a single theoretical prediction (i.e., a univocal T-theoretical determination) for the denotation of a term t. In that case, as we said above, we simply compare the predicted value (the T-theoretical determination) with the "observed" one (the T-non-theoretical determination). However, theories sometimes make predictions that are weaker than that, in the two following ways.[22]

Firstly, theories sometimes make "disjunctive" predictions, in the sense that they establish that the denotation of a given term t must be either d_1 or d_2 or d_3 or... So long as the set of values in that disjunction is a subset of all the possible interpretations of t (alongside the rest of the concepts) the theory is putting a restriction on what can happen and is thus making a prediction in some sense. Hence, we can characterize a term t as *weakly* T-determinable if, given the denotation of every other term that does not depend logically on t, the laws of the T restrict the possible interpretations of t. More precisely, if we take a concept c and the set C of all its possible interpretations in the $\mathrm{M_p}$, and the set ∂ consisting of the class of actual models M "cutting of" (the denotation of) the concept c, then c will be weakly T-determinable if and only if there is an $m = \langle D_1, ..., D_i, R_1, ... R_j, f_1, ..., f_k, c_x \rangle$ such that (*i*) $\langle D_1, ..., D_i, R_1, ... R_j, f_1, ..., f_k \rangle \in \partial$, (*ii*) $c_x \in C$, (*iii*) $m \in \mathrm{M_p}$, and (*iv*) $m \notin \mathrm{M}$.[23]

An example would be the concept *DESC* in cladistics, presented in the previous section. What our argument about inverting the 0's and 1's showed is that, given the denotations of all the other terms, there will always be at least two ways of assigning states to the terminal taxa which will make the actual tree be an optimal tree. However, this does not imply that *every* possible assignment will make the actual tree optimal. The set of assignments that do this will, in many cases, be a subset of that. In other words, there will be some possible assignments that the laws of CLAD rule out. Thus, *DESC* can be said to be strongly CLAD-non-determinable but weakly CLAD-determinable.

With this in mind, one could also think of introducing the notion of *weakly* T-testing concepts, as those that are T-non-theoretical and weakly T-determinable. A test using a disjunctive T-theoretical determination for the relevant term t could be called a weak test. Note that a weak test would still consist of a non-theoretical and a (weak) theoretical determination for some concept, so our conception of testing is still adequate. Also note that strong tests (those in which the theoretical prediction is univocal) are usually preferred to weak tests, since they allow stricter testing of theories (in Popperian terms, a theory that makes a single non-disjunctive prediction it has more potential falsifiers).[24]

There is a second relevant phenomenon, which we can also characterize as a weaker version of testing. Some theory could only put (disjunctive) restrictions on the denotations that a set of concepts can take, but not univocally determining the denotation that each particular concept must have. For a very simplified example consider a theory with the following law:

[22] We thank José Díez for bringing these two phenomena to our attention.

[23] Again, we thank José Díez for his help in formulating this condition.

[24] In the same way, a test involving a finite disjunctive prediction would be preferrable to one in which the relevant concept can have an infinite number of values (while still being a subset of the values in $\mathrm{M_p}$)

(L) If the patient has fever and a sore throat, then she has the flu.

From the fact that Alice has a fever and a sore throat, one can determine that she has the flu. Thus, "having the flu" would be a strongly T-determinable theoretical concept. However, from the fact that she has the flu and has fever, one cannot establish that she must have a sore throat, because L only gives a sufficient condition for having the flu, not a necessary one. Thus, having a sore throat would not be strongly T-determinable in this fictitious theory. However, note that from the fact that Alice does not have the flu one can establish that she does not have both fever *and* a sore throat. This is a disjunctive restriction on the possible interpretation of both concepts that does not univocally determine the denotation of either of them.

What is interesting about our analysis from above is that making the concept of strong testing explicit allows us to subsequently characterize these (and possibly other) weaker senses of testing as well.

6. General Considerations

Before moving on to our conclusions, we can draw three additional considerations, to illustrate how these developments of metatheoretical structuralism can also be relevant to certain general discussions within the philosophy of science.

The first concerns theory testing. Since its inception, the central theme of structuralism has been the explication of the structure of scientific theories, and the way in which it changes over time. Undoubtedly, the reconstruction of theories and their links is relevant to an understanding of how the testing of theories is carried out. Structuralists usually state (informally) that the standard conception of theory testing, the HD method, must be sophisticated and improved, but by no means discarded. Some of the things that are typically said to need revision are the dependence of the HD method on the observational-theoretical distinction, the fact that it confuses theory testing with hypothesis testing, and the fact that what is tested in a deductive hypothetical way is not the fundamental law of a theory, nor any special law, but the empirical assertion.[25] However, a better account of testing has not yet been worked out in detail. To undertake this task, it will be fundamental to keep in mind the way in which the T-theoricity and T-determinability distinctions interact in testing, the proposed distinction between T-testing and T-non-testing concepts and the correct presentation of the global empirical testing basis.

The second general point concerns the criterion of demarcation. Since the beginnings of professionalized philosophy of science in the early twentieth century, the refutability or testability of scientific theories has been discussed as a criterion of adequate factual knowledge. In some cases, to distinguish it from pseudoscientific or metaphysical theories (Popper), and in other cases as a criterion of cognitive significance (logical empiricism). Since then, we have learned that the discussion in its beginnings was somewhat naïve, and that the criteria provided by the classical philosophers of science turned out to be too restrictive (almost all interesting science turned out to be pseudoscientific/metaphysical/cognitively meaningless). However, the questions these philosophers posed may still be relevant today. In particular, one may wonder if there are (or could exist) theories that, by their very conceptual constitution, are impossible to test. This question is

[25] Which is a factual statement that states that a certain empirical system can be subsumed under one of the lines of specialization of a specific theory-net.

still interesting and could be relevant for those who want to continue to discuss the demarcation criterion in a more sophisticated way. The discussion carried out in this paper could help identify what a theory without a global testing basis (i.e., one without T-testing concepts) would consist of.

Take for example the case that is usually presented as an example of spurious unification: *What God wants to be the case is the case* (Kitcher 1881, p. 528). Kitcher's intuition is that the reason why this statement fails as a law has to do with the fact that it does not provide a genuine unification, and that this has to do with the stringency of the explanations provided. Díez (2002; 2014), continuing previous attempts to deal with explanation from a structuralist point of view (Bartelborth 1996; 2002; Forge 1999; 2002), deals with the question of spurious unification by pointing out the extreme malleability of abstract principles that lack special laws through which to increase their empirical content. For example, if we were to take the second principle of classical mechanics by itself, without the additional restrictions imposed by special laws, something similar to what happens with the principle above would occur. It would be possible to apply it trivially to any case we could imagine (the second principle, considered in isolation, is empirically unrestricted, as Moulines (1982), points out).

The discussion we carried out above allows us to collaborate with the characterization of what is wrong with the spurious statement provided. Let us take a more specific version of the given spurious law: *Organism x has trait y because God created it that way*. This statement has as its concepts *x possesses trait y* and *God created x with trait y*. We think it is quite intuitive to hold that the first concept would be T-theoretical (it would be impossible to determine what God wants independently of this law) and that the second would be T-non-theoretical, since it is possible to determine the possession of traits independently of the law. Note, however, that (in the absence of criteria restricting the divine will) it would not be possible to determine the possession of the trait from the law. This leads to the concept of trait possession being T-non-determinable (both strongly and weakly). The testing basis of the theory would be empty. And in this particular sense, the theory would not be testable (it would not be possible to theoretically ascertain the value of the non-theoretical concept), even when it has non-theoretical terms. This, which seems quite intuitive, had not been pointed out in the discussions regarding the spurious character of this type of statements. And it could collaborate with the elucidation of the sense in which the statement is irrefutable.

Additionally, and independently of the discussion about demarcation, this can be useful to collaborate with the structuralist approach to explanation. The standard characterization of special law given, for example, *Architectonic* (Balzer, Moulines and Sneed 1987, p. 170), is extremely weak (i.e., a law is a special law of itself). This discussion allows us to give an extra requisite that makes genuine special laws increase the empirical content of the theory, so that genuine testing becomes possible (the point made by Díez). The requirement (or one of the requirements) would be that what special laws must achieve is to provide procedures that allow theoretical determination for the non-theoretical concepts that appeared in the fundamental law. In other words, they must make some T-non-theoretical concept(s) T-determinable, thus allowing the theory to have at least one T-testing concept.

Finally, we can return to the issue of the status of the most basic domains of theories (such as organisms, particles, etc.). Although this is a more specific structuralist discussion, it does have wider implications for more general philosophical discussions (e.g., a lot of effort has been, and continues to be, devoted to understanding what an organisms is). As shown in the examples above, in many (if not most) cases, these concepts will turn out to be T-non-testing for their respective theories. Remember that a concept *c* will be T-determinable when, knowing the denotations of all other concepts that do not depend logically on *c*, the laws of T allow us to infer (either a univocal or a disjunctive) denotation for T. However, since almost every other concept will contain them as domains, then no other denotations will be known, and nothing will be inferable. This, however, is merely a conjecture that needs fuller examination in more applied work. What is doubtful, however, is if these concepts fit into our new category of T-non-theoretical and T-non-testing, because the former can be uncertain.[26]

7. Conclusions

Having gone a long way, we can now be more explicit about what was said in the introduction. The classical philosophers of science intended to account for the role that concepts had in explanation and in testing with a single distinction, by appealing to what, because of their empiricist attitude, they considered key: observability. Within the framework of metatheoretical structuralism, the question of the role of observation in science was separated (not because it is unimportant or because it has no role) from that of better understanding the functioning of the independent testability of scientific theories. A more sophisticated and fruitful distinction than the classical observational-theoretic one was then proposed, paying attention to theory-dependent and theory-independent criteria of determination.

However, the idea that independent testability was the key to understanding the role that concepts have in explanation persisted, as Ginnobili and Carman (2016) have argued. Even within structuralism itself other proposals emerged, which provided additional distinctions to Sneed's original one. This allowed these authors to have a more sophisticated account of both explanation and testing, and of the different roles that concepts can play in these. This did not imply a criticism of the T-theoricity distinction, but rather, an establishment of its proper scope and role, a condition of possibility to establish its usefulness.

In that vein, Ginnobili and Carman (2016) proposed the T-explainability distinction and argued that it does not coincide neither extensionally nor intensionally with that of T-theoricity. To hold this, they showed there are theories that conceptually extend their intended applications

[26] It is debatable whether the T-theoreticity distinction applies to all concepts of a theory, since some concepts of some theories do not seem to be determinable through theoretical determination methods at all (neither T-theoretical nor T-non-theoretical). The most debated case is the concept of *particle* in classical mechanics, for there is no explicit theory that allows the application of this concept. Moulines (1991, p. 224) has argued that *particle* could come from a very elementary implicit theory that provides such application criteria. Falguera (2006, p. 76) has doubted that such a theory exists, suggesting that it is a concept that, being non-theoretical in classical mechanics, depends in some sense on its laws, since we consider particles to be those entities that follow the laws of classical mechanics. He characterizes such concepts as T-non-theoretical but with "theoretical charge" of T. Ginnobili, Carman and Lastiri have suggested instead that the distinction does not apply to such concepts since their application does not appeal to the laws of any theory as is usually the case (Ginnobili 2018, pp. 147-50; Ginnobili, Carman and Lastiri 2008). Such concepts seem rather to be used to refer to the domain of intended applications of the theory. This could be the case of concepts such as *O* in NST (which does not always apply to organisms in the strict sense). If this were the case, the T-theoreticity distinction would not establish a partition between the concepts of a theory. Some concepts might be neither T-theoretical nor T-non-theoretical in T. We leave the question aside for the sake of simplicity.

with non-theoretical concepts. However, we can make the conjecture (partly reasonable and partly based on the fact that we know of no case that refutes it) that every T-theoretical concept is T-explanatory.

In this paper we have tried to show how a third distinction, that of T-determinability (which could already be found in the standard literature of structuralism), interacts with T-theoricity in the testing of a theory. We have also attempted to show that not every T-non-theoretical concept is T-determinable. This allowed us to introduce one additional distinction, T-testability, which permits us to talk more precisely about the global testing basis of a theory.

Additionally (although this was not our main goal), we have shown that not every T-explanatory concept is T-determinable. The picture of the situation, then, is quite complex (see Figure 1 in Section 3). As is often the case, the theoretical or metatheoretical frameworks that we propose to account for certain phenomena are usually too simple in their origin (because it is rational to begin by thinking that the phenomena we want to account for are simple) and tend to become more complicated and sophisticated as time goes by, in the development of the program.

Acknowledgments

This work has been financed by the research projects PUNQ 1401/15 (Universidad Nacional de Quilmes, Argentina), UNTREF 32/19 80120190100217TF (Universidad Nacional Tres de Febrero, Argentina), PICT-2018-3454 y PICT-2020-SERIEA-01653 (ANPCyT, Argentina), UBACyT 20020190200360BA (Universidad de Buenos Aires, Argentina) and PIUNAHUR-2021 *La incorporación de la perspectiva de género en la enseñanza de la biología en el marco del Profesorado Universitario en Biología de la UNAHUR*. We would like to thank José Díez for his valuable comments on previous versions of this paper.

References

Balzer, W. 1983. Theory and Measurement. *Erkenntnis* 19(1-3): 3–25.

Balzer, W. 1985. On a New Definition of Theoreticity. *Dialectica* 39(2): 127–145.

Balzer, W. and Dawe C M. 1997. *Models for Genetics*. Berlin: Peter Lang.

Balzer. W.; Moulines, C. U. and Sneed J. D. 1987. *An Architectonic for Science. The Structuralist Program*. Dordrecht: Reidel.

Bartelborth, T. 1996. Scientific Explanation. *In*: W. Balzer and C. U. Moulines (eds.), *Structuralist Theory of Science. Focal Issues, New Results*, pp. 23–43. Berlin/New York: Walter de Gruyter.

Blanco, D.; Roffé, A. J. and Ginnobili, S. 2020. The Key Role of Underlying Theories for Scientific Explanations. A Darwinian Case Study. *Principia: An International Journal of Epistemology* 24(3): 617–32.

Darwin, C. R. 1859. *On the Origin of Species by Means of Natural Selection*. London: John Murray.

Darwin, C. R. 1872. *On the Origin of Species by Means of Natural Selection*. 6th ed. London: John Murray.

Díez, J. A. 2002. Explicación, unificación y subsunción. *In*: W. González (ed.), *Pluralidad de la explicación científica*, pp. 73–93. Barcelona: Airel.

Díez, J. A. 2012. Inconmensurabilidad, comparabilidad empírica y escenas observacionales. *In*: P. Lorenzano and O. Nudler (eds.), *El camino desde Kuhn. La inconmensurabilidad hoy*, pp. 67–118. Madrid: Biblioteca Nueva.

Díez, J. A. 2013. Scientific w-Explanation as Ampliative, Specialized Embedding: A Neo-Hempelian Account. *Erkenntnis* 79(8): 1413–1443.

Falguera, J. L. 2006. Foundherentist Philosophy of Science. In: G. Ernst and K.-G. Niebergall (eds.), *Philosophie der Wissenschaft -Wissenschaft der Philosophie. Festschrift für C. Ulises Moulines Zum 60. Geburstag*, pp. 67–86. Mentis: Paderborn.

Feyerabend, P. 1962. Explanation, Reduction and Empiricism. In: H. Feigl and G. Maxwell (eds.), *Scientific Explanation, Space, and Time* (*Minnesota Studies in the Philosophy of Science*), pp. 28–97. Minneapolis: University of Minneapolis Press.

Forge, J. 2002. Reflections on Structuralism and Scientific Explanation. *Synthese* 130(1): 109–121.

Gähde, U. 1983. *T-Theoretizität und Holismus*. Frankfurt am Main/Bern: Peter Lang.

Ginnobili, S. 2010. La teoría de la selección natural darwiniana. *Theoria* 25(1): 37–58.

Ginnobili, S. 2011. El estatus fenomenológico de la teoría de la selección natural. *Ideas y Valores* 60(145): 69–86.

Ginnobili, S. 2012. Reconstrucción estructuralista de la teoría de la selección natural. *Ágora. Papeles de filosofía* 31(2): 143–69.

Ginnobili, S. 2016. Missing Concepts in Natural Selection Theory Reconstructions. *History and Philosophy of the Life Sciences* 38(3): 8.

Ginnobili, S. 2018. *La teoría de la selección natural. Una exploración metacientífica*. Bernal: Universidad Nacional de Quilmes.

Ginnobili, S. and Carman, C. C. 2016. Explicar y contrastar. *Critica* 48(142): 57–86.

Ginnobili, S.; Carman, C. C. and Lastiri, M. 2008. Una reflexión acerca de la distinción T-teórico-/T-no-teórico de la filosofía de la ciencia estructuralista. *IX Coloquio Internacional Bariloche de Filosofía*.

Hanson, N. R. 1958. *Patterns of Discovery : An Inquiry into the Conceptual Foundations of Science*. Cambridge: Cambridge University Press.

Hempel, C. G. 1958. The Theoretician's Dilemma. *In*: H. Feigl; M. Scriven and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*. Volume 2, pp. 37–98. Minneapolis: University of Minnesota Press.

Hempel, C. G. 1970. On the 'Standard Conception' of Scientific Theories. *In*: M. Radner and S. Winokur (eds.), *Minnesota Studies in the Philosophy of Science*. Volume 4, pp. 142–163. Minneapolis: University of Minnesota Press.

Hillis, D. M.; Bull, J. J.; White, M. E.; Baggett, M. R. and Molineux, I. J. 1992. Experimental Phylogenetics: Generation of a Known Phylogeny. *Science* 255(5044): 589–592.

Kitching, I. J.; Forey, P. L.; Humphries, C. J. and Williams, D. M. 1998. *Cladistics: The Theory and Practice of Parsimony Analysis*. New York: Oxford University Press.

Kuhn, T. S. 1962. *The Structure of Scientific Revolutions*. Chicago. London: University of Chicago Press.

Lorenzano, P. 1995. *Geschichte und Struktur der klassischen Genetik*. Frankfurt am Main: Peter Lang.

Lorenzano, P. 2012. Base empírica global de contrastación, base empírica local de contrastación y aserción empírica de una teoría. *Ágora* 31(2): 71–107.

Millikan, R. G. 1989. In Defense of Proper Functions. *Philosophy of Science* 56(2): 288–302.
Moulines, C. U. 1982. *Exploraciones metacientíficas*. Madrid: Alianza Editorial.
Moulines, C. U. 1991. *Pluralidad y recursión*. Madrid: Alianza Universidad.
Owen, R. 1847. *On the Archetype and Homologies of the Vertebrate Skeleton*. London: John E. Taylor.
Owen, R. 1849. *On the Nature of Limbs*. London: John Van Voorst.
Pearson, C. H. 2010. Pattern Cladism, Homology, and Theory-Neutrality. *History and Philosophy of the Life Sciences* 32(4): 475–92.
Pearson, C. H. 2018. Theoricity and Homology: A Reply to Roffe, Ginnobili, and Blanco. *History and Philosophy of the Life Sciences* 40(4): 62.
de Pinna, M. C. C. 1991. Concepts and Tests of Homology in the Cladistic Paradigm. *Cladistics* 7(4): 367–94.
Roffé, A. J. 2020a. Dynamic Homology and Circularity in Cladistic Analysis. *Biology & Philosophy* 35(1): 21.
Roffé, A. J. 2020b. El estatus fáctico de la cladística: aportes desde una reconstrucción estructuralista. *Metatheoria – Revista de Filosofía e Historia de la Ciencia* 11(1): 53–72.
Roffé, A. J. 2020c. Reconstructor: A Computer Program That Uses Three-Valued Logics to Represent Lack of Information in Empirical Scientific Contexts. *Journal of Applied Non-Classical Logics* 30(1): 68–91.
Roffé, A. J. 2020d. *Contrastando reconstrucciones con herramientas computacionales: Una aplicación a la Cladística*. Dissertation: Ph. D. Dept. of Philosophy. University of Buenos Aires.
Roffé, A. J.; Ginnobili, S. and Blanco, D. 2018. Theoricity, Observation and Homology: A Response to Pearson. *History and Philosophy of the Life Sciences* 40(3): 42.
Sneed, J. D. 1971. *The Logical Structure of Mathematical Physics*. Dordrecht: Reidel.
Wiley, E. O. and Lieberman, B. S. 2011. *Phylogenetics: Theory and Practice of Phylogenetic Systematics*. New Jersey: John Wiley & Sons.
Wright, L. 1973. Functions. *Philosophical Review* 82(2): 139–68.