Using Conceptual Spaces to Model the Dynamics of Empirical Theories[1]

Peter Gärdenfors and Frank Zenker

University of Lund

Abstract: In *Conceptual Spaces* (Gärdenfors 2000), dimensions and their relations provide a topological representation of a concept's constituents and their mode of combination. Consequently, as concepts are *n*-dimensional geometrical structures, conceptual change denotes the dynamic development of these structures. Following this basic assumption, we apply conceptual spaces to the dynamics of empirical theories and show, in particular, that the terms of the *structuralist view of empirical theories* can be largely recovered. Based on the logically possible change operations which a concept's dimensions can undergo (singularly or in combination), we identify four types of (increasingly radical) change to an empirical theory. The incommensurability issue as well as the importance of measurement procedures for the identification of a radical theory change are briefly discussed.

Key-words: theory dynamics, structuralism, conceptual spaces, theory dislodgement, normal and revolutionary science, incommensurability, integral and separable dimensions

1. Introduction

The aim of this paper is to apply *conceptual spaces* as developed by Gärdenfors (2000) to give a new account of the dynamics of scientific theories. We shall compare this account to the *structuralist view* of empirical theories (Sneed 1971, Stegmüller 1973, Balzer et al. 1987, Moulines 2002). Using a reconstruction of *Newtonian Particle Mechanics* (NPM) as a paradigmatic example, we explain how the structuralists' terms are applied, and show that, by using conceptual spaces, we can recover *most* of the key concepts of the structuralist view without the set-theoretical overhead. There is also some loss in comparison with structuralism, because we are not after a *mathematically general* model. Moreover, we do not situate our approach with respect to the statement vs. non-statement dichotomy, as this appears (to us) as an overstressed distinction.

[1] Forthcoming in: Olsson, E.; S. Rahman, T. & Tulenheimo, T. *Science in Flux: Philosophy of Science Meets Belief Revision Theory* (Logic, Epistemology, and the Unity of Science-Series). Berlin: Springer.

Our aim is to provide philosophy of science with a new tool by which to comprehend in general terms the dynamics of empirical theories. We shall argue that our approach, which is based on geometrical notions, is more suited as a general framework for representing theories and their dynamics than structuralism is. We believe it also fits better with the intuitions of practicing scientists. In particular, we will show by means of examples that conceptual spaces provide a clearer account of different kinds of changes of scientific theories.

We start with a summary of the structuralist view of empirical theories (section 2), followed by an outline of our modeling tool, *conceptual spaces* (section 3). In section 4, we show how the central notions of structuralism can be expressed in terms of conceptual spaces. We then summarize the structuralist reconstruction of theory changes and point out its limitations in the application to *radical* or *revolutionary* changes (section 5). In section 6, we let conceptual spaces prove their mettle by presenting four types of increasingly more severe changes to empirical theories.

2. A brief summary of the structuralist view

For a structuralist, an empirical theory is a set of *mathematical (set-theoretical) structures* – hence the name. These structures happen to satisfy some *axioms* that can be expressed as set-theoretical predicates (cf. Suppes 1957: ch. XII, Sneed 1971). Since the structures and not the axioms are central, structuralists characterize their program as a *non-statement* or *semantic view* of empirical theories. Ultimately, the aim of the program is to provide a framework for the detailed representation of the logical structure of an empirical theory in order to achieve a rigorous reconstruction of *changes* either to the theory or the conditions of its application to empirical phenomena, for example, a reconstruction of how NPM changed over time.[2]

We now turn to a brief presentation of the basic concepts of structuralism. According to Sneed's (1971) account (further developed by Stegmüller (1973)), an empirical theory is represented as a pair <*K*, *I*>, which consists of a *formal core*, *K*, and a set of *intended applications*, *I*. The intended applications are identified pragmatically,

[2] The construction and application of an empirical theory can naturally be seen as the paradigmatic example of rational human belief and its use. In this sense, structuralism also becomes a framework, albeit limited, for doxastic dynamics.

while *K* is specified via the set-theoretic structures that systematize its parts, most notably its *models* (see below). Although structuralism has developed over time,[3] we choose to focus on Sneed's original account since it contains the most essential components.

The mathematical structure of a theory core is described, firstly, by a set of *measures* (variables) for different magnitudes of the objects that are studied. The measures are functions in the set theoretic sense of sets of ordered pairs. In NPM the relevant variables of an object are *position* (location in space), *time*, *mass* and *force*. Secondly, there are *constraints* for these measures. For example, in every model of NPM, mass is supposed to be a *conservative* (any object has the same mass in all applications) and *additive* magnitude (the mass of a complex object is the sum of the masses of its components).[4]

The magnitudes which cannot be measured without the theory itself being applied are called *T-theoretical* and those which do not presuppose the theory for their measurement are called *T-non-theoretical* (Sneed 1971). In NPM, force and mass are theoretical while position and time are non-theoretical, because we are not required to use NPM to determine values of the former. Instead we rely on an "antecedently accepted" theory for measuring space and time.[5]

The distinction between theoretical and non-theoretical magnitudes motivates a corresponding one among a theory's models: An empirical structure (a.k.a. *data-structure*) of values for the measured variables (space and time) is called a *partial potential model* and the set of theses structures denoted *Mpp – partial*, because the model *lacks* theoretical functions. If values for non-theoretical terms (*time* and *location* in NPM) are specified, but the values for the theoretical terms are unconstrained, one obtains the set of *potential models Mp – potential*, because all possible values for theoretical functions appear in some member of *Mp*. Thus, in *Mp* kinematical descriptions (co-

---

[3] Cf. Moulines (2002) for a brief outline of the current state of the structuralist program and Balzer et al. (1987) for a full account.

[4] Balzer et al. (1987:105) call these the constraints of *equality* and *extensivity*. Cf. their ensuing discussion with respect to the simplifying assumption that gives rise to a third constraint: In any subsystem considered, e.g., moon and earth, masses are assumed to be impressed upon by the *same* forces, if these masses *were* related to the system as a whole, i.e., the cosmos. "Although this is not so if we look at things quite accurately, physical calculations work with such and similar assumptions" (1987:106).

[5] NPM does, of course, presuppose its own theory of space-time, namely that of *absolute* (Euclidian) *space* and *absolute simultaneity*, cf. DiSalle (2006:17-35, 98-130).

ordinations of indexes for time over Euclidean space) are set in relation to a system's masses and the forces impressed upon them, that is, the dynamical factors. Of the potential models, only some will satisfy the central axioms of the theory, for example Newton's second law $F = ma$ in NPM. A potential model that satisfies the axioms of the theory is called a (full) *model* of the theory. The set of models is denoted M. Hence: $M \subseteq Mp$, while the relation between Mp and Mpp is that formally expressed by a "forgetful functor" (alternatively a restriction function (projection) from Mp onto Mpp).

In order to address changes to empirical theories, it is useful to focus on Sneed's (1971) notion of the *core of a theory*. It consists of the following five parts: (i) a set of variables of the theory, (ii) a set of constraints for the theoretical variables, (iii) a set of models, determined by the central axioms for the theory, (iv) a set of potential models and (v) a set of partial potential models.[6] The core captures what remains *constant* in a theory during the time of its use.

Some changes of a core are so-called *core expansions* (Stegmüller 1976:107). They are reached from the core through a process of *specialization*, thus adding special laws of a theory, e.g., the Newtonian law of gravitation (see below). Thus, in the case of NPM, its core does not yet establish any *quantitative* relation between given masses, forces and accelerations. Rather, a core specifies so-called *basic laws* of a theory.[7]

In this sense, basic laws underlie – thus come to be valid in, thus unify – all of the theory's models. Of course, the core does not yet exhaust a theory. Rather, it spells out presuppositions with which *more specific* characterizations – formulated as core-expansions and by means of *special* or *substantial laws* – must be consistent. Only the latter are (sought to be) applied to real world situations.[8]

Through a *specialization* of the core, that is, by adding restricting characterizations, the structuralist can specify a theory's *internal structure* hierarchically

[6] This is not exactly Sneed's definition, which may be found in his (1971:171). Our above characterization of a core, however, captures what is essential about this concept.

[7] This is Sneed's term, cf. Sneed (1979:300), Stegmüller (1976:107f.) uses *fundamental laws*. In addition, there are assumed characterizations of the single components of the model, so-called *frame conditions*. E.g., for NPM, that the set of particles is finite or that mass is a positive real function. These conditions "do not say anything about the world (or are not expected to do so) but just settle the formal properties of the scientific concepts we want to use" (Moulines 2002:5).

[8] Hence, both basic laws and frame conditions are not *open* to refutation in the same sense that special laws are. Certainly, they can be *revised*, but not (without using stronger assumptions) *falsified*.

as a partial order over so-called *theory-elements*.[9] For example, in the case of NPM, from the basic law $F = ma$ (expressed in theory element T_0) one may "carve out" (Balzer et al. 1987:169) first, T_1, the *actio-reactio* principle (Newton's third law), followed by specifying *conservative forces* in T_2, then *central forces*, T_3, followed by *forces inversely proportional to distance*, T_4, and, finally, *forces for which the gravitational constant*, G, *holds*, T_5.[10]

Thus, one reaches the law of gravitation, $F = G \cdot Mm/r^2$, as a five-fold specialization from the core of NPM. In this way, one has construed a series of core expansions and thereby specified the core. Note, again, that with only $F = ma$ in the core, we do not have any information about the interaction of particles. Models that satisfy $F = ma$ might just as well be about single particles. It is only with such special laws as the law of gravitation that connections between objects are introduced.

The decision on what to reconstruct as the core of a theory is relevant, because the potential models are characterized *only* by the core structure. Therefore, every more specialized theory T_{n+1} must have the *same* set of potential models as the less specialized T_n, i.e., $Mp(T_{n+1}) = Mp(T_n)$.[11] Differences arise, amongst others, with respect to a theory-element's full models, *M*, and its intended applications *I*, that is, $M(T_{n+1}) \subseteq M(T_n)$ and $I(T_{n+1}) \subseteq I(T_n)$.

Since there is, in principle, more than one way in which a partial potential model can be completed to a full model, the *empirical claim* of a theory is rendered as the contention that every partial potential model (representing data structures arrived at by measurement) *can* be successfully enriched to a full model. This claim *may* very well turn out to be false, in which case one would say: A particular theoretical enrichment of a partial potential model constitutes an *anomaly* for the theory. Among the partial potential models are, most notably, the *paradigmatic applications* (i.e., systems to which the theory has already been successfully applied and which are considered central to the

[9] The term 'theory-element' denotes the set of the sets *M, Mp, Mpp, L* (links to theory elements specialized form different cores), *C* (constraints), *I* (intended applications) and blurs, *B*, used for approximation (cf. Balzer et al. 1987).

[10] Cf. Gähde (1997, 2002).

[11] *A fortiori* for the so-called *partial potential models, Mpp*, i.e., $Mpp(T_{n+1}) = Mpp(T_n)$.

theory). These form a subset of the *intended applications I*. The latter are (in a non-formalized sense) *similar* to paradigmatic applications.

With all the parts of the core and the intended applications above in place, the structuralist disposes of a seemingly powerful terminology to describe changes within one theory and connections across several empirical theories.

3. Conceptual spaces

With this brief summary of structuralism as a background, we next turn to a presentation of our modeling tool. Conceptual spaces represent information by *geometric* structures rather than by set theory. Information is represented by points in the space (standing for objects or individuals), and regions (standing for properties and relations) in dimensional spaces. A great deal of the structure of a theory can be modeled in a natural way by exploiting *distances* in the space. These distances represent degrees of similarity between objects.

A conceptual space consists of a number of *quality dimensions*. Psychological examples of such dimensions connected to sensory impression are color, pitch, temperature, weight, and the three ordinary spatial dimensions. However, in scientific theories the dimensions are determined by the variables presumed by the theory. We have already noted that within NPM the relevant dimensions are three dimensions of space, time, mass and three dimensions of force.

The primary role of the dimensions is to represent various "qualities" of objects in different *domains*. The notion of a domain can be given a more precise meaning by using the notions of *separable* and *integral* dimensions. These concepts are adapted from cognitive psychology (see e.g. Garner (1974), Maddox (1992), Melara (1992)). In that context, certain quality dimensions are said to be integral if, to describe an object fully, one cannot assign it a value on one dimension without giving it a value on the other.

For example, an object cannot be given a hue without also giving it a brightness value. Or the pitch of a sound always goes along with its loudness. Dimensions that are not integral are said to be *separable*, as for example the size and hue dimensions. Within the context of scientific theories, the distinction should rather be defined in terms of *measurement procedures*. If two dimensions (or sets of dimensions) can be measured by independent methods, then they are separable, otherwise they are integral.

In NPM, space and time are separable. In contrast, a relativity theory construes space-time as an integral set of dimensions. In the kinematics (descriptions of moving bodies) presupposed by Newtonian mechanics, the measurement procedures for location in space (measuring rods or optical signals) and time (pendulum motions, i.e., clocks) were considered to be independent of each other. By defining velocity and acceleration as the first and second derivates of position with respect to time, Newton proposed a theory that *coordinated* spatial and temporal measurement. In this sense, Newton presupposes a theory of space-time. Still, there is no interaction in the measurement of *distance* and that of *duration*, i.e., between trigonometry and chronometry.

Using this distinction, the notion of a *domain* can now be defined as a set of integral dimensions that are separable from all other dimensions. In NPM, the domains are four: *space*, *time*, *mass* and *force*. The domains form the framework used to assign properties to objects and to specify relations between them (see below). The dimensions are taken to be independent of symbolic representations in the sense that we can represent the qualities of objects, for example by vectors, without presuming an explicit language in which these qualities are expressed.

The notion of a dimension should be understood literally. It is assumed that each of the domains (integral set of dimensions) is endowed with certain *topological* or *metric* structures.[12] It is part of the meaning of "integral" dimensions that they share a metric. Considering NPM, space is a three-dimensional Euclidean space, time is a one-dimensional structure that is isomorphic to the line of real numbers, mass is a one-dimensional structure that is isomorphic to the positive half-line of real numbers, and force is isomorphic to vector a three-dimensional Euclidean (vector) space.[13]

A consequence is that the topological structure of different quality dimensions entail that certain statements will become *analytically true*. For example it follows from the linear structure of the length dimension that comparative relations like "longer than"

[12] For examples of different topological and metric assumptions within psychological domains, see Gärdenfors (2000).

[13] We say "isomorphic to" rather than "homomorphically embedded in". Infinitesimally small differences, e.g., in the lengths of objects, will fall below the threshold of our cognitive capacity, and are thus not measurable. Still, such a difference clearly can be part of a conceptual space (cf. Batitsky 2000:96).

are *transitive*. This is thus an analytic feature of such a relation (*analytic-in-S,* that is). Similarly, it is analytic that everything that is green is colored (since "green" refers to a region of the color space) and that nothing is both green and blue. Analytic-in-S is thus defined on the basis of the topological and geometrical structure of the conceptual space S.[14] However, different conceptual spaces will yield different notions of analyticity, which leads to a form of *relativism* that would be foreign to a classical notion of analyticity.

4. Correspondence between structuralism and conceptual spaces

After the brief accounts of structuralism and conceptual spaces, this section will show how most of the structuralist notions can be expressed in terms of conceptual spaces. Since our account does without the set-theoretic paraphernalia of the structuralists, we believe that it is more palatable for practicing scientists and fits better with their intuitions. Furthermore, conceptual spaces will allow us to highlight new aspects of the structure and dynamics of theories. In particular, we will show in section 6 that our approach generates a new way of classifying theory changes.

For each of the five components of a theory core (see above), we will show what its correspondence in terms of conceptual spaces is. Let us begin with the measurements (variables) that form the building blocks of a theory core. In general, they correspond to the domains of a conceptual space. For example, in the structuralist account, *space* and *time* were the two T-non-theoretical terms of NPM. The distinction between T-theoretical and T-non-theoretical terms is basically the same in conceptual spaces, except that we put emphasis on the separability of measurement procedures. The importance of this will show up in section 6.3.

Secondly, let us consider the constraints on the theoretical variables. In general, they are determined by the assumptions concerning the metric (or scale) that is connected with the domain. NPM introduces as theoretical terms the variables *mass* and *force*. In the conceptual space of NPM, mass is a separable dimension isomorphic to the non-negative real numbers. The *extensivity* or *additivity* of mass, i.e., $m_1 + m_2 = m_{1+2}$ (where 1+2

[14] The "phenomenological" assumptions that Carnap (1971:78f.) formulates are validated by the very structure of the space and need not be added as meta-linguistic constraints.

denotes the object composed of objects 1 and 2), is accounted for by the assumption that mass is measured on a quotient scale (Stevens 1946, Ellis 1968).

Another constraint on the mass dimension is that it be *conservative* which means that mass is a property of an object that stays constant over different applications. *Force* is represented as three integral dimensions isomorphic to Euclidian 3-D space. A constraint on this variable is that the component forces of a body must be *independent* of the system to which the body belongs.

In NPM, any object (particle) is represented as a point in the eight-dimensional space spanned by the *space*, *time*, *mass* and *force* dimensions. Once an object has been assigned a value for all of these eight dimensions, it is fully described as far as the conceptual apparatus of Newtonian particle mechanics is concerned.

We then turn to the three kinds of models. Whenever the structuralist speaks of various models, the conceptual space framework generally speaks of sets of *vectors* or *points* in dimensional space. Each point represents the properties of an object. Thus, a *partial potential model* of NPM is a set of points (partial vector) in 4-dimensional space (3-D for space, 1-D for time). A *potential model* is a set of points in the entire 8-dimensional space of NPM. Finally, a full model is a set of points (full vectors), the values of which satisfy the core axioms.

In NPM, the partial potential models are construed from partial vectors with values for space and time, while a full model involves vector values for all eight dimensions such that they always satisfy Newton's force law and may also satisfy a special law. In particular, Newton's second law $F = ma$ determines a hyper-surface in the eight-dimensional space. However, it should be noted that the law in itself only concerns single objects. Only when specialized laws are added will NPM introduce forces that connect several objects in an application.

The upshot is that once the conceptual space is specified and the core axioms formulated, the three kinds of models fall out very naturally as:

(i) sets of points in the subspace with values for T-non theoretical dimensions (*partial potential model*)

(ii) sets of points in the space with values also for T-theoretical dimensions (*potential model*)

(iii) sets of points in the space with values for all dimensions such as to satisfy the core structure (and perhaps some special law) (*full model*).

As can be seen from this reconstruction, the three kinds of models, which play such a central role for structuralism, are not really required as separately specified entities when the conceptual space plus the theoreticity distinction are given.

How are the values for the various vectors determined? Just like the structuralists, we assume that the values for non-theoretical dimensions are determined by observations and that this is done by careful measurement according to established procedures. The values for the theoretical dimensions are obtained by presuming that the applied theory yields an empirically correct prediction. This parallel is not changed by taking a different view upon the description of the conceptual apparatus of a theory.

As regards the intended applications, the framework of conceptual spaces has little new to offer.[15] Naturally, theories lend themselves to make predictions which are based on special laws. For example, the law of gravitation is applied to a given partial potential model of NPM, i.e., data on the time- and location-function of a number *n* of objects. In our way of speaking, this comes out as *n* trajectories in the 8-D space (described above) as constrained by the law of gravitation. It should be noted that, unlike Newton's second law, the law of gravitation introduces forces that *connect* several objects in an application.

Finally, the empirical claim of a theory is rendered as follows: *Any* partial vector (partial potential model) can be completed to a *set of points* (one for each object in the application) in the eight dimensional space (full points) that satisfy the constraints and axioms of the core. In particular, in NPM the points are predicted to lie on the hyper-surface spanned by $F = ma$. Certain applications will also be expected to satisfy further special laws.

[15] However, it should be investigated whether the similarity naturally offered by distances in conceptual spaces can be exploited to give an account of the similarity of different applications.

Of course, at a given moment, one will start by considering only the finite number of special laws that have already been established. Yet, strictly speaking, *any* special law consistent with the frame-conditions will qualify, *including those not yet forwarded* (cf. Stegmüller 1976:105). Thus, our rendering of the empirical claim is just as *weak* as that offered by structuralism. This, we think, is as it should be. We can thus show that the components of an empirical theory as identified in structuralism can also be identified in terms of conceptual spaces.

5. Structuralist change operations and their limitations

Having shown the correspondences between structuralism and conceptual spaces, we shall next discuss the change operations that structuralism can identify and point to a limitation in the reconstruction of so-called *radical theory change*.

On the pragmatic side, structuralism can reconstruct the following changes by respecting the set of *intended applications I* as a part of a theory element: By correcting earlier measurements, the numerical values of a partial potential model for, say, Mercury's orbit within the application of NPM to the solar planet system – which had, so far, been *successfully completable* to a full model – may no longer be completable without offsetting a neighboring application, say, Venus' orbit (cf. Roseveare 1982, Gähde 1997, 2002). In this case, a particular intended application, unless it is simply *retracted* from the set of intended applications, becomes an *anomaly*.[16]

Furthermore, Mercury's orbit (which had been calculated with the aid of Newton's law of gravitation) can "move" to a *new* theory element, by means of which one hopes to calculate the application, after all. This had been the case when a correction term to the exponent of Newton's gravitational law was proposed or when Clairaut or Hall proposed alternative gravitation laws that feature different correction terms (cf. Roseveare 1982).

On the formal side, the basic change operations that structuralism offers are the *addition* or *deletion* of any of the parts that constitute a theory element, i.e., any change to

[16] Naturally, a data structure that can no longer be successfully completed to a full model may itself be *hypothetically* enriched, such as to include so far unobserved additional objects. E.g., postulation of Vulcan is such a case. This, however, does not seem to be a relevant change, as the new hypothetical data structure – *qua* also obeying $F = ma$ – already is among the set of partial potential models, but had merely not been proposed as one that is fit for *theoretical* enrichment, in the very first place.

the elements of the sets *M*, *Mp*, *Mpp*, *C*, *L* (see above) – possibly including so-called *blurs* (cf. Moulines 2002). Thus, the addition/deletion of an "entire" theory element to/ from the structure of a theory – e.g., the proposal of a new law that can be specialized from an already present element – is merely a special case of the changes to the parts of a theory.

Note that, through the identification of the set of potential models via satisfaction of basic law plus constraints, there is – despite appearances to the contrary – not exactly a whole lot of room for the structuralist to trace changes of a more *radical* bend. While the internal (logical) structure of a theory (as revealed through a structuralistic reconstruction) may change in the most minute ways, every system to which *the same* theory is applied will still be "forced" to co-operate within the conditions that are spelled out in the most basic theory element, such as $F = ma$ in NPM.[17] Thus, one defines the potential models of NPM into which we can "squeeze" a given kinematical system. At the same time, one does thereby *exclude* any alternative which does not obey $F = ma$.

Clearly, structuralism provides a very fine-grained view on changes to *one* basic structure. However, in the case of a so-called *revolutionary theory-dislodgement* – e.g., the transition from *Newtonian Particle Mechanics* (NPM) to *Einstein's General Relativity* (GR) – the structuralist must resort to speaking of a *core-replacement*, because $F = ma$ simply is a basic law that is *not* valid in all of GR's potential models.[18] Speaking of ´core-replacement or *theory dislodgement* (Stegmüller) strongly suggests that there are "jumps" in theory evolution – an assumption we deny.

On our view, symbolic formulations of a theory, i.e., *equations*, specify *quantitative relations* between the ranges of values that this theory's terms can take. If two theories that quantitatively relate the *same* terms in a *different* way or – which is the more likely case – quantitatively relate a subset of these terms to *terms not included in the former theory*, one is well advised to take a step back from the equations. Instead, it is

[17] This is the sense in which Sneed can explicate 'having a theory' as "being committed to use a certain mathematical structure, together with certain constraints on theoretical functions to account for the behavior of a, not too precisely specified, range of phenomena" (1971:157).

[18] Cf. Diederich (1996:80) for the claim that "the classical problems of incommensurability have [thereby] been circumvented", rather than resolved. Cf. also Balzer et al. (1987:306-19) for attempts at tackling the incommensurability issue by relating two theories, *T* and *T**, through their sets *Mp* and *Mp**. For Kuhn's largely negative reaction to the structuralist's endeavor, cf. Kuhn (1976).

worthwhile to consider the conceptual spaces that the theories span. In this way – or so we submit – one may better understand how the old and the new space are connected.

This way of viewing the matter provides – we think – an interesting approach to the incommensurability issue that, however, we will not take up here.[19] For our present purposes, it is sufficient to have shown that we can fruitfully address theory-dislodgement. This very issue appears not to be satisfactorily answered by the structuralists endeavor. *Nolens volens,* by speaking of core rejections, the structuralist will have to admit that she cannot trace the *continuities* between, e.g., NPM and GR by means of her reconstructive apparatus. Thus, she cannot reconstruct the transition, but only the theory's "initial and final sets of generalizations" (Kuhn 1987:19).

6. Four types of theory change within the framework of conceptual spaces

Compared with structuralism, we know of no comparable attempts in the literature at a similar formal representation of concrete cases of theory evolutions. However, this account of theory dynamics appears weak when we look at its ability of completely model radical theory change. The rather heavy set-theoretical apparatus that has been developed for the purpose gives little substantial insight into the processes of such changes of theories in return. In this section our aim is to show that the framework of conceptual spaces fares better.

Given the notion of an empirical theory *as* a conceptual space (presented in section 4), changes to a theory core (including expansions) can naturally be divided into four types:

(i) addition and deletion of special laws;
(ii) change of scale or metric as well as the salience of the dimensions;
(iii) change in the separability of dimensions;
(iv) addition and deletion of dimensions that make up the space.

[19] Following Kuhn, incommensurability has been predominantly identified as a problem that occurs in relating the symbolic forms of two theories or frameworks. With our shift away from the symbolic and towards the conceptual level, we may end up not finding incommensurability at all. This sounds odd, but it is how it should be. After all, the practicing scientists that we know do not admit to any problems whatsoever in making a transition from one set of generalizations to the other.

We shall argue that this ordering of the changes represents increasing *degrees of severity*. To show that these are generally applicable distinctions, we consider examples from the history of science. In the following, we discuss which change operation a case exemplifies. To be clear, the hypothetical completion of an application to one that features additional data, e.g., postulation of the planet Vulcan for the solar planet system, does not constitute a change to the theory in question (see above), but is rather a change in one application.

6.1 Addition of special laws

In general, the addition of a special law to a theory core only further specifies the class of models of the theory and thus increases the empirical content of the theory. Historically, it is not unusual that special laws, e.g., Hooke's law of the spring or the law of the pendulum, are formulated as specializations of the frame conditions *after* the latter had been specialized – in this case – to the law of gravitation. Such a process only extends the range of applications of a theory core, without causing any significant change in the hitherto available theory structure.

Generally, the addition of special laws seems to be characteristic of what Kuhn (1962/1969) has called *normal science*. It should be regarded as the mildest form of change to an empirical theory since it does not involve any change in the theory core. In line with the choice of the expression 'expansion' from Sneed/Stegmüller, it could also be seen as an *expansion* of the theory core in the sense of belief revision (Gärdenfors 1988). In the present terminology, the dimensions presupposed in an empirical theory are simply used in new quantitative ways within the same qualitative space.

Special laws could also be *deleted*, but this is often not quite what happens when a theory core encounters anomalies. Rather, if an application of a special law results in predictions that do not fit with data, the application itself may be *retracted* from the set of intended applications and, possibly, be moved to a new theory in which it may be more successful. For example, Newtonian mechanics was once presumed to apply to the phenomenon of light. It was later realized that this would not work, whence light phenomena ceased to be intended applications of NPM. However, any "problematic" special law may persist as part of the theory without any application being assigned to it, as one may hope to find a new application for it in the future. Hence, special laws are never really deleted. Rather their intended applications may be temporarily suspended.[20]

6.2 Change of metric/scale

It is part of the description of a conceptual space to assign every domain (set of integral dimensions) its own metric. In theories that use classifications based on features that depend on several domains, the metrics for the domains must be weighed together. In other words, their relative salience must be determined (see Gärdenfors 2000, section 4.7.2). For example, pre-Linnaean botany was based on holistic features of the flowers, such as *size* and *color*, while Linnaeus' classification made the *numbers of pistils* and *stamens* the most salient features. What is involved in this kind of theory change is the principle for *combining* different domains of a conceptual space.

However, even *within* a single domain there may occur changes of metric. It is trivial that temperature can be measured on both the Celsius and the Fahrenheit scales. These scales are equivalent since they both involve an interval scale (invariant under all positive linear transformations (Stevens 1946, Ellis 1968)). However, temperature can also be measured on the Kelvin scale, which is stronger since it has an absolute zero point (and thus has less invariance). The change from Celsius to Kelvin leads to different predictions concerning temperature. It is part of the theory associated with the Kelvin scale that no object can have a temperature below absolute zero, while no such prediction could be made only from assuming that temperature is measured on an interval scale. Thus the change leads to a change in the empirical content of the theories of temperature associated with the scales. [21]

Another example of a more severe form of changing metric is obtained when, in reaction to experimental findings in early chemical theory, it was argued that the "fire substance" (*phlogiston*) would need to have *negative* mass in order for the theory to square with experience. Here, the negative range of the mass scale needs to be introduced – a rather radical, but possible move to bring the theory in line with the empirical results.

---

[20] Of course, such changes are traceable and, therefore, not without repercussions in the theory. If T_n is the respective theory element, its set of applications $I(T_n)$ will "drop to zero."

[21] A science-historical account of the process leading to a current concept of temperature as *mean kinetic energy* is provided in Chang (2004).

A change of metric involves changes of the predictions of the theory and is thus a form of *revision* in the sense of Gärdenfors (1988). It is therefore a more drastic change than adding new special laws. However, the change is still relatively mild, since the basic framework of the conceptual space and the core axioms are maintained.

6.3 Change in integrality or separability of dimensions

In NPM time and space are separable domains. A remarkable change occurred as a reaction to the Michelson Morley null result on ether drift. It was hypothesized that the rods by which one measured length are *shortened* in the direction of the ether drift, resulting in the so-called Lorentz-Fitzgerald contraction. Effectively, one thereby "squeezed" the length-scale to account for the null result within an ether theory.

A most basic assumption today is that light signals propagate at a finite velocity. An ether theorist, on the other hand, expected a difference in the speed of light signals as a function of their direction of travel relative to the ether. To uphold the ether hypothesis after the null result, it was proposed that, the rod – along which the light beam traveled and then returned by reflecting from a mirror at its end – was *shortened* in the direction of the "ether-wind". Moreover, shortened just enough to let the (predicted, yet unobserved) drag of the ether-wind onto the light beam cancel out.

This is an effective way of interpreting an experiment in favor of one's theory. However, it can hardly be called a plausible hypothesis, if – as Einstein did – one doubts that there be (any necessity for) an ether to begin with and is committed to the constancy of lengths on observational grounds. In fact, Einstein's solution does not presuppose a *mechanism* by which the length-scale is squeezed. From suitable assumptions, he could rather deduce the contraction factor such that there will be only an *apparent* contraction.[22]

The Lorentz-Fitzgerald contraction seems to be an exceptional case of integrating domains within the framework of a theory core (i.e. NPM). In general, dimensions are not

[22] If L_0 is the length of an object in a rest frame, and L_1 the length measured by an observer, then the contraction is given by $L_1 = L_0/\gamma$, where γ is defined as follows (with *u* for relative velocity between observer and object, and with *c* for the speed of light): $\gamma = \frac{1}{\sqrt{1-\frac{u^2}{c^2}}}$

separated, nor are unconnected dimensions integrated, unless some more severe change also takes place in the form of adding or deleting dimensions.

6.4 Addition and deletion of dimensions

The most fundamental change of a conceptual space occurs when dimensions are added or deleted. Most, perhaps all of the historical changes Kuhn (1962/1969) calls revolutions can be analyzed as changes in the fundamental dimensions of a scientific area.

A paradigmatic case for the addition of a dimension is Newton's introduction of the *mass* dimension as distinct from the *weight* dimension of Galilean physics and adding the dimension of *force* in his mechanics. Given the distinction between weight and mass, an object of a given weight is now analyzed as an object of a given mass under the influence of a given gravitational force.[23] Effectively, the weight dimension is deleted and replaced by the separate dimensions of mass and force, which function as theoretical variables in Newton's theory.

For a second example, in order to save electro-magnetic phenomena, Einstein introduced the energy dimension as a theoretical variable in his relativity theory and eliminated force as a fundamental dimension. Energy, in the form of kinetic energy, was a derived variable of NPM, but in GR it becomes a fundamental variable.

As a third example, following Chen (2003), one can characterize the particle theory of light by assuming that it postulates *at least* two integral dimensions for *velocity* and *size* (both taking continuous values) and one dimension for *side* (taking a binary value), separable from velocity and size. In a wave theory of light, the *velocity* dimension remains, but it now becomes integral with the dimensions *amplitude* and *wavelength*, while the dimension of *size* is deleted and replaced by *phase difference* (also taking a binary value).

6.5 Discussion

We will leave a full discussion of the incommensurability issue for future work, because we are first required to have a good *definition* of incommensurability. The extent to

[23] In 1901, the *Bureau International des Poids et Mesures* conventionally defined as much "to put an end to the ambiguity which in current practice still exists on the meaning of the word *weight*, used sometimes for *mass*, sometimes for *mechanical force*" (BIMP 1901:70).

which our approach provides an interesting answer to this issue will largely be a function of the definition we chose.

However, the main point may already have been anticipated: If one accepts as plausible the idea of literally *developing* a conceptual space into a new one by the change operations identified above, then its description as one of the above four types of change *is* the answer to the incommensurability issue. In other words, the traditional *problem* of incommensurability – finding ruptures between the symbolic forms of predecessor and successor theory – is a consequence of treating the symbolic level of representation as primary. On the conceptual level, the four kinds of change operations establish the continuities between an old and a new theory.

It appears to us that a set-theoretical apparatus (such as structuralism) is prone to conceal insights among precise yet cumbersome formulations. This is especially so in cases of so-called large scale changes, i.e., theory dislodgement, for which, in our opinion, structuralism only offers a pseudo-reconstruction by offering the concept of a core-rejection. This ultimately seems to miss out on tracing continuity.

We have provided several examples of theory dynamics that can be analyzed in terms of changes of conceptual spaces; other examples need to be studied. The reader might have noticed that we do not lay much stress on the mathematical symbolism that usually accompanies accounts of theory change. In fact, we suggest that the inclusion of such symbolism in the presentation of the history and philosophy of science is a myopic outgrowth of established standards fostered by a philosophical training which is almost exclusively devoted to the symbolic level of representing theories and other forms of knowledge.

This, we hold, should not persist as the *only* framework in which to discuss nor understand changes to empirical theories. The formulas themselves do not reveal the underlying assumptions concerning the variables involved: their geometrical structures, their integrality and their determining measurement procedures. Of course, studying formulae is indispensable in the analysis of empirical theories, but understanding conceptual change should not proceed exclusively on this level.

It is important to realize that the conceptual structure of an empirical theory as well as its dynamics can be approached by *abstracting* from the quantitative relations (formulae) and by focusing on the qualitative relations between the terms postulated by the theory, in our terminology: on the *dimensions* that constitute an empirical theory. We do not thereby deny that exact science should be expressed in the language of mathematics, but we deny that insights into the conceptual development of theories are generated by staring at formulae. Rather, insights arise from having a clear and simple geometrical conception of what it means to be a (separate or integral) dimension and from having defined change operations which, when applied to the space, can transform it into a new one. Furthermore, in our opinion, the role of the measurement procedures associated with the central variables has been underestimated.

Taking stock of what is gained by our approach when compared with structuralism, then – a full answer on the incommensurability issue pending – we would claim that a reconstruction of empirical theories in terms of conceptual spaces will generate the insights of structuralism (without the set-theoretical apparatus) and provides a fruitful way of describing different kinds of theory changes which goes beyond what is possible within structuralism. Already, one can say that our approach bears evident benefits for educating a wider audience in theory change. Future work should treat additional cases, the relation between measurement and the separability of dimensions, and the similarity relations between applications that can be formulated in terms of conceptual spaces.

7. Conclusions

Conceptual spaces allow us to present a more unified view than structuralism of empirical theory cores and their expansions and, in particular, theory dynamical aspects in a way that fits better with actual scientific practice. The geometrical notions developed in Gärdenfors (2000) provide insights into how a theory develops by classifying changes according to the operations that have been identified in the previous section. On our account, the fact that the symbolic formulations of a theory change over time is but an effect of the dynamics of the underlying conceptual space. While we believe that our approach will allow a more fruitful approach to the problem of incommensurability than structuralism has been able to offer, we must leave this to future work.

We have shown how the structuralist's set theoretical constructs can find their correspondences in the theory of conceptual spaces. In particular, it poses no major difficulty to recover the T-theoreticity distinction and account for the distinction into partial potential, potential and full models of a core (and its expansion). Basically, these distinctions are reached by separating among a theory's dimensions those that are grounded in an antecedently available measurement processes from those that are not.

In general, the significance of measurement procedures for our account is the following: When it comes to Kuhn's revolutionary change, we hypothesize that any introduction of a new or any deletion of an old dimension will also reveal a change in the measurement procedure. In this sense, the measurement procedures also come out as the pragmatic links between concepts and the empirical world that they represent.

Acknowledgements

We would like to thank the organizer and the audience of the December 2006 workshop at Lund University and especially with respect to our exposition of structuralism, J. D. Sneed and C. U. Moulines for helpful comments and criticism. Frank Zenker acknowledges funding from the *Swedish Institute* (SI) and Peter Gärdenfors from the Swedish Research Council.

References

Balzer, W., C. U. Moulines & J. D. Sneed (1987). *An Architectonic for Science. The Structuralist Program*. (Synthese Library Vol. 186). Dordrecht: Reidel.

Balzer, W., D. A. Pearce & H.-J. Schmidt (1984). *Reduction in Science. Structure, Examples, Philosophical Problems* (Synthese Library Vl. 175). Dordrecht: Reidel.

Balzer, W., C. U. Moulines & J. D. Sneed (eds) (2000). *Structuralist Knowledge Representation. Paradigmatic Examples*. Amsterdam: Rodopi.

Batitsky, V. (2000). Measurement in Carnap's late Philosophy of Science. *Dialectica* **54**, 87-108.

BIMP (1901). Comptes Rendus de la 3^{e} Conférence Générale des Poids et Mesures, Paris 1901. http://www1.bipm.org/en/CGPM/db/3/2/

Carnap, R. (1971). A basic system of inductive logic, part 1. In: Carnap, R., and Jeffrey, R. C. (eds). *Studies in Inductive Logics and Probability. Vol. 1*. Berkeley, CA: UCP, 35-165.

Chang. H. (2004). *Inventing Temperature. Measurement and Scientific Progress*. Oxford: OUP.

Chen, X. (2003). Why Did John Herschel Fail to Understand Polarization?: The Differences between Object and Event Concepts. *Studies in the History and Philosophy of Science* **34**, 491-513.

Diederich, W. (1996). Pragmatic and diachronic aspects of structuralism. In: Balzer, W. & C. U. Moulines (eds). *Structuralist Theory of Science. Focal Issues, New Results*. Berlin/New York: De Gruyter, 75-82.

DiSalle, R. (2006). *Understanding Space Time. The Philosophical Development from Newton to Einstein*. Cambridge: CUP.

Ellis, B. (1968). *Basic Concepts of Measurement*. Cambridge: CUP.

Gähde, U. (1997). Anomalies and the revision of theory-elements: Motes on the advance of Mercury's perihelion. In: Dalla Chiara, M. L. et al. (eds). *Structures and Norms in Science. Vol. 2* (Synthese Library Vol. 260). Dordrecht: Kluwer Academic Publishers, 89-104.

Gähde, U. (2002). Holism, Underdetermination, and the Dynamics of Empirical Theories. *Synthese* **130**, 69-90.

Gärdenfors, P. (1988). *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge, MA: MIT Press.

Gärdenfors, P. (2000). *Conceptual Spaces*. Cambridge, MA: MIT Press.

Garner, W. R. (1974). *The Processing of Information and Structure*. Potomac, Wiley, New York: Erlbaum.

Kuhn, T. (1962/1969). *The Structure of Scientific Revolution*. Chicago: CUP.

Kuhn, T. (1976). Theory Change as Structure Change. Comments on the Sneed Formalism. *Erkenntnis* **10**, 179-99.

Kuhn, T. (1987). What are Scientific Revolutions? In: Krüger, L. et al. (eds.). *The Probabilistic Revolution. Vol. 1*. Cambridge: MIT Press, 7-22.

Maddox, W. T. (1992). Perceptual and decisional separability. In: Ashby, G. F. (ed.). *Multidimensional Models of Perception and Cognition*. Hillsdale, NJ: Lawrence Erlbaum, 147-180.

Melara, R. D. (1992). The concept of perceptual similarity: from psychophysics to cognitive psychology. In: Algom, D. (ed.). *Psychophysical Approaches to Cognition*. Amsterdam: Elsevier, 303-388.

Moulines, C. U. (2002). Introduction: Structuralism as a Program for Modeling Theoretical Science (Special Issue on Structuralism). *Synthese* **130**, 1-11.

Roseveare, N.T. (1982). *Mercury's perihelion from LeVerrier to Einstein*. Oxford: Clarendon Press.

Sneed, J. D. (1971). *The logical Structure of Mathematical Physics*. Dordrecht: Reidel.

Stegmüller, W. (1976). *The Structuralist View of Theories*. Berlin, Heidelberg, New York: Springer.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science* **103**, 677-680.

Suppes, P. (1957). *Introduction to Logic*. New York: Van Nostrand.

University of Lund
Departments of Cognitive Science and Philosophy
Kungshuset, Lundagard
22 222 Lund
Sweden

peter.gardenfors@lucs.lu.se
mail@frankzenker.de