In Praise of The Representation Theorem

Nancy Cartwright[*]

1. Introduction

This paper will take up three of Patrick Suppes's favourite topics: representation, invariance and causality. I begin not immediately with Suppes's own work but with that of his Stanford colleague, Michael Friedman. Friedman argues that various high level claims of physics theories are not empirical laws at all but rather *constitutive principles*, principles without which the concepts of the theory would lack empirical content. I do not disagree about the need for constitutive principles. Rather I think Friedman has mislocated them, and entirely at the wrong end of the scale of abstraction. It is representation theorems, as Suppes pictures them, that are the true constitutive principles, and that is true for theories far beyond physics.

My disagreement with Friedman has two prongs. First, the high-level principles he calls 'constitutive' are not (at least in many cases). Second, representations theorems are. The first half of my claim depends on a quite different view of theory that I have from Friedman, a view I believe I share at least in part with Suppes. But it is not the central part of Suppes's work I want to connect with today. Here I shall focus on the second half, in favour of representation theorems. I begin by explaining Friedman's view and describing why I think representation theorems are better candidates than his own for constitutive principles. Then I shall illustrate by looking at a simple representation theorem in a subject in which Suppes has occasioned an important revolution, causality. The example I choose is a representation theorem that links causality with a second of Suppes's favourite subjects, invariance. I want to illustrate the role of representation theorems as constitutive principles by showing that the theorem on invariance is indeed a presupposition for (one kind of) empirical meaningfulness of the abstract concept of a causal law.

---

[*] LSE and UCSD

2. Constitutive principles

For the Kantian, part of the job of ensuring that our empirical knowledge fits the world of experience is done by the synthetic a priori. This provides the rational framework within which we experience the world. As Michael Friedman puts it, "synthetic a priori knowledge (typified by geometry and mechanics) …functions as the presupposition or condition of possibility of all properly empirical knowledge."[1] Friedman is keen to resurrect the role of the synthetic a priori, but not as a once-and-for-all framework necessary for empirical experience. Rather each proper theory in modern physics has its own framework that is held, relative to it, as a priori and that makes possible the genuinely empirical knowledge within that theory.

Friedman's principal examples are the three laws of Newtonian mechanics, which are a priori in his sense in the Newtonian scheme as currently understood. The law of universal gravitation – "that there is a force of attraction or approach, directly proportional to the two masses and inversely proportional to the square of the distance between them, between any two pieces of matter in the universe"[2] – is the one empirical law in the scheme. This, he points out, talks about *acceleration*. Newton defined acceleration relative to absolute space. Since we do not believe in absolute space, we cannot do this. We say rather that the law of universal gravitation holds in any *inertial frame*, "where an inertial frame of reference is simply one in which the Newtonian laws hold (the centre of mass frame of the solar system, for example, is a very close approximation to such a frame)."[3] This is why in our current rendering of Newtonian theory Newton's three laws must be taken as a priori:

> It follows that without the Newtonian laws of mechanics the law of universal gravitation would not even make empirical sense, let alone give a correct account of the empirical phenomena. For the concept of universal acceleration that figures essentially in this law would then have no empirical meaning or application: we would simply have no idea what the relevant frame of reference might be in relation to which such accelerations are defined.[4]

[1] Friedman (2001), 26
[2] Friedman (2001) 36
[3] Ibid.
[4] Ibid.

A priori here can have a very local meaning: prior to the empirical laws, though perhaps posterior to other theories, claims, operations and definitions. Friedman argues for something more general. He maintains that these principles cannot be warranted in the same way that genuine empirical laws can. Whether that is so for the claims he counts as constitutive, those I want to count under this label – representation theorems – certainly are a priori in a straightforward sense: they are meant to be provable.

According to Friedman constitutive principles, like those defining the frame of reference of the concept of acceleration, make the "empirical application of the theories in question first possible".[5] I think this is a misdescription. The principles that Friedman calls 'constitutive' make the concepts *intelligible*, not *empirically applicable*. They provide univocal and precise definitions that fit the concepts into the relevant high theory but they are not the principles that make possible the empirical application of these concepts. When it comes to the presuppositions for empirical knowledge, as opposed to presuppositions for fitting these concepts into a highly abstract theory, representation theorems are a far better candidate.

3. *Representation theorems*

We represent features of the empirical world with specific mathematical forms that have specific properties. These forms tend to be far more universal across applications than is any (univocal) interpretation or definition of the related concept. *Acceleration* for instance: no matter what frame of reference we define it relative to, we almost always represent it as d^2x/dt^2. So length itself must be represented as a quantity twice-differentiable with respect to time. But it also has a number of other built-in features as well. Probably the simplest is that length is represented by an additive measure. Can a mathematical representation with these characteristics adequately represent the phenomena to be associated with "length"?

---

[5] Ibid. 49 This of course, as he points out, does not guarantee that the empirical principles that we formulate using them will be true, just that they are candidates for truth or falsity.

The answer depends on the structure of the phenomena to be represented. In the case of length, the phenomena might include what happens to sets of measuring rods. For instance, as Suppes puts it, the collection A of rods is longer than B "if and only if the set A of rods, when laid end-to-end in a straight line, is judged longer than the set B of rods also so laid out."[6] Formalizing that fact, along with a couple of other obvious features we ascribe to the empirical concept of length (for instance, that any collection of rods is at least as long as the empty set), we can characterize the structure consisting of the set of rods and the longer-than relation as a *finite equally-spaced extensive structure*.

Now we are in a position to show that an additive measure is an appropriate representation for length by proving a *representation theorem.* In this case the theorem tells us that for any finite equally-spaced extensive structure, there is an additive measure μ such that for every pair of sets of rods, A and B, $\mu(A) \geq \mu(B)$ iff A is longer than B. That is only a start of course. In order to guarantee the empirical applicability of the concept of length as we represent it in mathematical physics, we need a representation theorem relating all the qualitative features we assign to length to its mathematical representation. And similarly for all the quantities of empirical reality and their features for which we provide mathematical representations.

It is, then, I urge, in the representation theorems we offer for the concepts in use in modern science that we find our best candidates for "constitutive principles". These are the preconditions for the application of our concepts to empirical reality. Our representations are consistent with the features we ascribe to empirical reality only if the appropriate representation theorems are true.

4. Causality and invariance: a representation theorem

I consider here only linear causal systems, where

$\{V, \leq, L, C\}$ is a *linear causal system* iff

---

[6] Suppes (2002), 64

V is a finite set of variables, $x_1,\ldots,x_n$

$\leq$ is a complete, antisymmetric, transitive ordering on V

L is a finite set of linear equations over V of form $x_i = \Sigma a_j x_j$, $j \neq i$

$C \subset L$

such that the following axioms hold:

1. *L-Consistency:* Members of L are mutually consistent.

2. *L-completeness:* All linear combinations of members of L are in L.

3. *Causal precedence and antireflexivity:* $x_i = \Sigma a_j x_j \ \varepsilon\ C \rightarrow x_j < x_i$[7].

4. *Causal antisymmetry:* $x_i = \Sigma a_j x_j\ (j \varepsilon N)\ \varepsilon\ C \rightarrow \forall j \varepsilon N, \forall a \neq 0,\ x_j = a x_i + \ldots \notin C$.

5. *Causal numerical transitivity:* $x_i = \Sigma a_j x_j\ (j \varepsilon N)\ \varepsilon\ C$ and $x_j = \Sigma b_k x_k\ (k \varepsilon M)\ \varepsilon\ C \rightarrow \Sigma\Sigma a_j b_k x_k\ (j \varepsilon N, k \varepsilon M)\ \varepsilon\ C$

6. *Causal independence:* Members of C are independent (i.e., not linear combinations) of each other except for numerical transitivity.

7. *Supervenience:* $x_i = \Sigma a_j x_j \ \varepsilon\ L$ iff it is a linear combination of equations in C.[8]

Define: a *C-intervention* on $x_i \ \varepsilon\ V$ creates a new linear causal system identical to the first but where C is replaced by C': C' = the closure under numerical transitivity of C – {all equations ε C of form $x_i = \Sigma\ldots$} ∪ {$x_i = X$}, for any value X in the range of x_i.

[7] $a<b =_{df} a \leq b$ and $\neg b \leq a$

[8] This should guarantee that whenever two laws for the same effect appear in C, the 'connecting' equation that gets from one to the other by numerical transitivity is itself in C since the connector must itself be in C or be a linear combination of equations in C. (This is needed for the proof.)

Invariance Theorem: For any linear causal system {V,≤,L,C}, an equation $x_i = \Sigma a_j x_j$ for $j \varepsilon N$ and $x_j < x_i \; \varepsilon$ C iff it is in C' under all C-interventions on x_j $\forall j \varepsilon N$, that is, iff it is *invariant* under C-interventions on right-hand-side variables.[9]

Notice that the set C of causal laws relative to any set V of variables and set L of functional laws over them is not unique. There are many subsets of L that have the right form to count as a causal basis for L, even once the causal order among the quantities represented by V is fixed. Which is the right set – the set of 'true' causal laws for a situation – is pinned down by the invariance theorem coupled with a physical interpretation of "intervention". If we suppose that we have a set of physical operations that count as interventions on the quantities represented by $x_1,\ldots,x_n$, *relative to the true causal laws* – then $x_{n+1} = \Sigma a_j x_j$ (for j = 1,…,n) is a true causal law iff it is invariant under all those interventions.

Consider now the formal characterization of a linear causal system on the one hand and the operational characteristic of invariance under intervention on the one other. The operational characterization does provide a concept of causal law that is immediately empirically meaningful. But it faces the same problems as do operational definitions in general. First, it is too narrow, picking out just one operation to characterize the concept, thus requiring a host of additional empirical laws to guarantee the connection of what is so characterized with the other operations we use to measure the concept and with the predictions we expect to make from it. Second, it is *ad hoc*. How do we know that this operation measures the concept we are concerned with? The representation theorem shows that it must: the invariance theorem must hold if the formal characteristics of 'causal laws' are to hold.

On the other hand, the formal characteristics by themselves are not enough. What it is to be a causal law is empirically underdetermined by these features, even though they include the central claim that causal laws are the 'ontological basis' for all functionally true laws plus the presupposition of some particular causal order. The tie to the invariance test provides empirical content. But again, it is *ad hoc* simply to assert the connection between the formal characteristics and an operation of choice for

[9] For a proof see Cartwright (2003)

measuring it. The representation theorem shows that the choice is not *ad hoc*. The representation theorem is a presupposition of the empirical meaningfulness of the concept of causal law. That is, as I said in my introduction, the representation theorem is the true site of constitutive principles.

References

Cartwright Nancy (2003) "Two Theorems on Invariance and Causality" *Philosophy of Science*, 70, 2003 (203-224)

Friedman Michael (2001) *Dynamics of Reason: The 1999 Kant Lectures at Stanford University*, Stanford, California: CSLI Publications

Suppes Patrick (2002) *Representation and Invariance of Scientific Structures*, Stanford, California: CSLI Publications