

Mapping the mind: On the status of functional explanation in the cognitive sciences



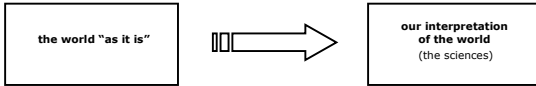
Rimas Čuplinskas
Department of Philosophy
University of Bonn, Germany

Hans-Christian Schmitz
Inst. of Comm. Research and Phonetics
University of Bonn, Germany



1. Scientific Theories

It seems plausible to maintain a basic distinction between **the world "as it is"** (or objective reality) and **our interpretations of the world**. Another way of depicting this distinction would be to call the first the "non-metaphysical realm", thereby contrasting it with the "metaphysical realm", in which the sciences, the theories they deploy, as well as the entities posited in such theories are to be found.



The relation between the two realms can be generally described such that the truth or utility of our interpretations depends on them somehow corresponding to happenings in the real world. Precisely how the relation is to be adequately characterized, distinguishes different positions in philosophy of science:

- **classical view of theories** (Hempel, Carnap): Theories are to be construed as formal or syntactic structures in which theoretical principles are linked to empirical content by means of correspondence rules, thus involving a sharp distinction between theoretical and observational sentences and terms. Theories make assertions about the way the world is. Scientific change is cumulative, so any new theory will be consistent with (though more inclusive than) previously accepted theories. Since it must be possible to confirm or disconfirm theories (they either do or do not account for empirical data), they are *objective* accounts of reality.
- **historicist view of theories** (Feyerabend, Kuhn): Scientific change is not cumulative, it is occasionally characterized by radical paradigm shifts. What counts as empirical data is not theory-independent. Scientific theories are *subjectively* significant, i.e. their meaning cannot be fully understood independently of the scientific community advancing them.
- **semantic view of theories** (Suppes, van Fraassen): Scientific theories are models in which certain relational properties of physical systems are represented. The model is structurally isomorphic to the system. If the representation is quantitative, a model is a mathematical model and its adequacy is measured by comparing its behaviour to that of the physical system. Each theoretical model posits a set of objects whose properties and behaviour are characterized by certain laws. The truth of theoretical assertions lies in the extent to which these comply with an accurate model.

2. Explanation, Understanding, Informativity

We look for explanations *in order* to understand. The extent to which an explanation leads to understanding can be called the *informativity* of that explanation. We can distinguish three main **theories of scientific explanation**:

- **the classical theory** (Hempel, Oppenheim): To explain an event scientifically is to give a deductive or inductive argument for that event. At least one of the premisses of that argument must be a scientific law.
- **the causal-statistical theory** (Reichenbach, Salmon): Scientific explanations must provide both the set of factors statistically relevant to an event and the causal network underlying the statistical regularities. *B* is said to be statistically relevant to *A* if the probability of *A* given *B* is different from the probability of *A*.
- **the pragmatic theory** (van Fraassen): An event is explained scientifically if it is the subject of a telling scientific answer to a why-question. Why-questions are identifiable by their *topics of concern* (what the questioner supposes to be the case), *contrast classes* (a set of alternatives including the topic of concern) and *explanatory relevance conditions*.

The development from the classical to the pragmatic model of scientific explanation reflects a move from a restrictive notion of explanation to a more inclusive one. Thus, the pragmatic theory incorporates the criterion of informativity more than its predecessors.

A **system** is a set of entities and relations between these entities. The entities can themselves be systems. The **internal structure** of a system is the set of relations between its entities. The **behaviour** of a system is defined as its external relations (input-output-transformations). The set of systems affecting or being affected by a given system is the **environment** of that system.

A **model** is a description of an object or state-of-affairs in which the latter is reduced to its relevant relational properties. What is deemed relevant depends on the scientist's interests. In this way, a model idealizes its subject. A **model of a system** is a **performance model** if it adequately models its subject's behaviour. It is a **functional model** if (i) it is a performance model, (ii) it consists of subsystems which are performance models of the subject's subsystems, and (iii) the subsystems are related in the same way as the subject's subsystems. A functional model is isomorphic to its subject. [The definition of a "functional model" is rather strict, by requiring the empirical adequacy of the model instead of it simply specifying its structure.]

3. Reduction and the Unity of Science

Classic programme of a reductivist account of nature

Oppenheim and Putnam (1958) developed a reductivist programme in which different levels of organization in the world (sociology, psychology, biology, chemistry, physics) are connected such that one level is explainable by (and reducible to) the next. The relation between different levels in the hierarchy is one of a whole to its parts, whereby the parts of entities posited on one level serve as entities on a lower level. The Oppenheim/Putnam programme involves a number of presuppositions:

- **ontological claim**: There is a structural unity in nature such that each branch of science is associated with a particular kind of object, the branches being arranged in a hierarchical order (part/whole-relation).
- **epistemological-methodological claim**: The unity of the world is reflected in the unity of science, the basis of which is a physical description of the world. Low-level (ultimately physical) explanations are more informative than high-level explanations.
- **logical core**: The concept of inter-theoretical reduction.

Three concepts of reduction

The classical reductivist strategy is ultimately an eliminativist model. According to this view, the special sciences are nothing but convenient (and relatively inaccurate) short-cuts waiting to be reduced to a more complete theory of physics.

A non-eliminativist, yet physicalist view is that the different branches of science legitimately deal with levels of reality involving laws and entities which in some sense are irreducible to lower levels. Ayala (1974) distinguishes three modes of reduction:

1. **constitutive (ontological) reductionism**: There are no elementary substances or forces other than those of elementary physics.
2. **theoretical (epistemological) reductionism**: All concepts and laws of special sciences can in principle be redefined in terms of physical concepts and laws.
3. **explanatory (methodological) reductionism**: Legitimate scientific explanations can only be given at a molecular level.

The modes of reduction are listed in order of increasing strength and dependency: one cannot claim (3.) without claiming (1.) and (2.), or, accordingly, claim (2.) without claiming (1.). The first position, constitutive reductionism, is one all physicalists must hold.

4. Functional Explanation and the Disunity of Science

The relation between physics and the special sciences can be described using the **concept of supervenience**. One set of properties *P* supervenes on another set of properties *Q* if the *Q* properties of an object determine what its *P* properties are. Supervenience is an *asymmetric covariance relation*. To claim, for example, that psychological systems are physical systems is to claim that there can be no difference in their psychological properties without a difference in their physical properties.

Functionalism is the view that types of biological or psychological states are individuated solely by the functional role they play within the system (Putnam 1960; Fodor 1968). They are not identical with *types* of physical states or processes, since a given function can be realized in many different ways (**multiple realizability**).

In **functional analysis** a system is described as a complex function which can be decomposed into simpler functions (sub-systems). Cummins (1983) proposes a three-stage methodology. First, the complex function is defined. Second, the function is decomposed into an organized set of simpler functions (stage of analysis). This analysis can proceed recursively by decomposing some (or all) of the subfunctions into sub-subfunctions. Thirdly, the operation of the bottom level of functions is explained by appealing to natural laws (e.g. mechanical or biological principles).

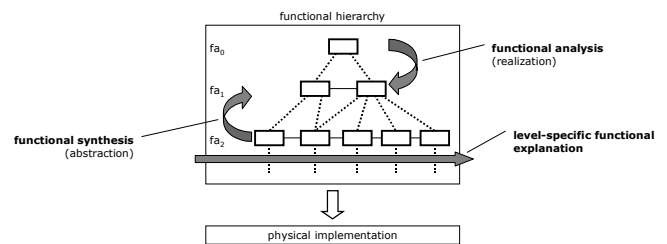
Functional synthesis can be described as the reverse process in which we explain which complex function is realized by a set of subfunctions (or which system a set of subsystems comprises).

The results of functional analysis and synthesis are functional models. A functional model represents the **functional architecture** of a system on a certain level of abstraction. We call the ordered set of all relevant functional architectures of a system its **functional hierarchy**. It is important to note that each element in the functional hierarchy can in principle be multiply realized. The operation of the lowest level in the functional hierarchy is explained by appealing to natural laws.

Explanation within a functional hierarchy is possible in three different ways:

1. **Realization**: Moving from a higher level of abstraction to a lower one (functional analysis) explains which subfunctions realize a complex function. In doing so we learn more about how a certain function is executed.
2. **Abstraction**: Moving from a lower level of abstraction to a higher one (functional synthesis) explains which complex functions are realized by the given functions. In doing so we learn more about what purpose the lower-level functions serve. This is especially informative when comparing two different systems, since we can learn e.g. that they are both executing the same high-level function.
3. **Level-specific description**: Each level of the functional hierarchy in isolation explains how the system works (relative to the level of description).

Idealizations and **approximations** are possible within the functional hierarchy.



5. Functional Explanation of Cognition

Let us construct a functional model of an ant. On a high level of description, the model consists of the ant's sensory, navigation, motor, reproductive systems and so on. These systems are described as performance models. On a lower level of description, these systems are decomposed and described as functional models. In doing so we learn how the navigation system etc. is realized.

Alternatively, we could imagine constructing a functional model of an ant on the cellular level (low level of description). We move to a higher level of description by abstracting from individual cell functions. We explain which functions are realized by certain cell groups.

Upon investigating the ant's behaviour, it seems plausible to decompose its navigation system and into the two subsystems GO_TO_FOOD and AVOID_DANGER. However, our model obviously describes the ant in an idealized way, since its navigation behaviour does not consistently comply with these two functions (e.g. it runs into ant traps).

We may draw the conclusion that the process of abstracting from a lower level to a higher level may involve *idealizing* the behaviour of the system. This, in turn, means that moving in the other direction (high-level to low-level) may involve the superfunction being only *approximately* realized by its subfunctions.

What permits us to make such idealizations in the functional hierarchy of a system? This depends on the system in question. If we are dealing with a **technical system** (e.g. a burglar alarm), then such functional idealizations are based on what we believe to be the interests of those who build or use such systems. If, on the other hand, we are dealing with a **biological system**, such idealizations can be made on the basis of its evolutionary history.

The theory of evolution allows us (i) to make sense of the idea that only some of the effects of a system are functions of the system, and shows (ii) how assigning a function to a system does not constitute projecting beliefs and desires into it. The central concept is that of adaptation, and the central claim is that a system is an adaptation for a certain function if this function amounted to a fitness advantage and was thus selected due to this advantage. (Sober 1993)

The cognitive sciences are in the business of describing, simulating and developing functional architectures of cognitive systems. When describing natural cognitive systems it is important to keep in mind possible idealizations and approximations involved within a functional hierarchy. When taking an *analytic* approach (top-down), one should not expect to find an inferred function to be perfectly realized by the subfunctions subserving it. By the same token, when taking a *synthetic* approach (bottom-up), one will not find a superfunction in the subfunctions realizing it, since in most cases we are merely dealing with approximate realizations.

- F. Ayala (1974), "Introduction", in: F. Ayala / T. Dobzhansky (eds.), *Studies in the Philosophy of Biology. Reduction and Related Problems*, Berkeley.
- A. Bartels (1996), *Grundprobleme der modernen Naturphilosophie*, Paderborn.
- R. Cummins (1983), *The Nature of Psychological Explanation*, Cambridge.
- J. Fodor (1968), *Psychological Explanation*, New York.
- J. Fodor (1975), "Special sciences (or: The disunity of science as a working hypothesis)", *Synthese* 28, 97-115.
- J.G. Kemeny / P. Oppenheim (1956) "On reduction", *Philosophical Studies* VII, 6-19
- K. Lambert / G.G. Brittan (1987), *An Introduction to the Philosophy of Science*, Ridgeview.
- P. Oppenheim / H. Putnam (1958), "Unity of science as a working hypothesis", in: H. Feigl et al. (eds.) *Minnesota Studies in the Philosophy of Science*, vol. 2.
- R. Pfeifer / C. Scheier (1999), *Understanding Intelligence*, Cambridge.
- H. Putnam (1960), "Minds and machines", in: S. Hook (ed.), *Dimensions of the Mind*, New York.
- Z.W. Pylyshyn (1984), *Computation and Cognition*, Cambridge.
- E. Sober (1993), *Philosophy of Biology*, Oxford.
- W. Stegmüller (1983), *Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie*. Bd. 1: Erklärung, Begründung, Kausalität, Berlin.
- G.H. von Wright (1971), *Explanation and Understanding*, London.

This is a small-scale copy of a poster presented at the 4th annual conference of the Association for the Scientific Study of Consciousness in Brussels (Belgium) on 30 June 2000. If you are interested in receiving a copy of a future publication on this subject, please contact us: Rimas Čuplinskas <cuplinskas@uni-bonn.de> or Hans-Christian Schmitz <hcs@ikp.uni-bonn.de>