

# A FUNCTION MODELLING APPROACH TO EMPIRICAL SCIENCE

Dawn E. Gregory and Paul R. Cohen  
 Experimental Knowledge Systems Laboratory  
 Computer Science Department, LGRC  
 University of Massachusetts  
 Box 34610, Amherst, MA 01003-4610  
 Phone: (413) 545-3638 Fax: (413) 545-1249

We are interested in developing an intelligent software agent to assist scientists with empirical studies. Although a variety of software is available for collection and analysis of experiment data, the difficult problems of hypothesis formulation and experiment design have not been adequately addressed. We hope to support both of these tasks by designing an appropriate modelling language.

The Scientist's Empirical Assistant (SEA) will provide an intelligent, goal-driven approach to creating scientific experiments. From the perspective of Artificial Intelligence, this means that SEA will use a large repository of *knowledge* to create one or more experiment *plans*. This paper describes a representation that facilitates this process.

*Function modelling* has traditionally been used to document information flow within organizations. The notation is very powerful because it allows us to describe the *actions* associated with each *object* of interest. For example, in computer science we study both programs (actions) and the data they process (objects). It is the purpose of empirical science to continually improve our understanding of both actions and objects. In order to store descriptive information, we extend the models to include *attributes*. Attributes of an object can be *variables* in an empirical study. Attributes of an action capture *semantic* information which is derived from an experiment or provided by the user.

Function models will provide an interface between SEA and the user, so a scientist can easily describe the problem at hand. A function model can be drawn in several ways, but we prefer the boxes-and-arrows paradigm, where each box is a function and the connecting arrows are objects. These diagrams also represent the different types of relationships between objects and functions: as an input or output value, as a constraining parameter, or as a tool that the function uses. An example function diagram is shown in figure 1(A).

SEA's knowledge base will store the components of a function model (i.e. objects and actions) according to class and context. Components are placed into *classes* according to the values of their attributes. For example, we will distinguish objects that are numeric from those that contain strings. *Context* is a hierarchy of scientific topic areas, which is used to constrain the search for relevant models.

While class and context are orthogonal concepts, they are both useful in generating experiment hypotheses. Suppose we define *equivalence* between functions  $F$  and  $G$  to be a one-to-one mapping between the connections of  $F$  and  $G$  that preserves the class of each connection. Whenever two functions are similar, but not equivalent (i.e. the mapping is not class-preserving), it may be beneficial to explore the differences. Context is used to focus this "matching" heuristic on relevant areas of the knowledge base, reducing the difficulty of generating meaningful hypotheses.

Function models have an additional advantage, in that they map easily to other modelling techniques. One such mapping is to a *causal model*, as shown in figure 1(B). It is also possible to map directly to techniques used for process analysis and improvement (planning), abductive and deductive reasoning, and time-series analysis. The point is that we can define mappings between function models and these other schemes, incorporating many types of information and analysis in a single repository. It is primarily this diversity which will allow SEA to effectively design scientific experiments.

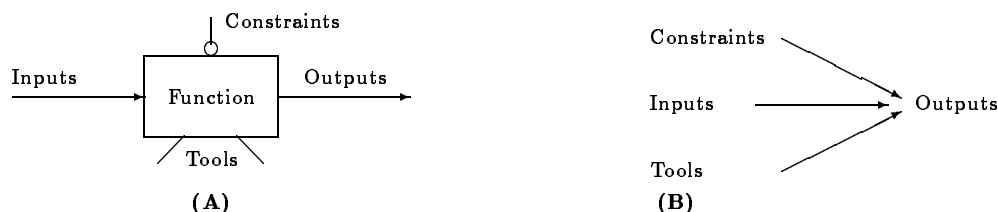


Figure 1: A function model (A) and its equivalent causal model (B).