Modelling Experiments in Scientific Discovery

Peter C-H. Cheng
Department of Psychology
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
U.S.A.
(cheng@b.psy.cmu.edu)

Abstract

Investigating the character of scientific discovery using computational models is a growing area in Artificial Intelligence and Cognitive Science. Scientific discovery involves both theory and experiments, but existing discovery systems have mainly considered the formation and modification of theories. This paper focuses on the modelling of experiments. A general characterization of the nature of experiments is given and more specifically Galileo's motion experiments are examined. The STERN scientific discovery system has been used to model Galileo's investigations of free fall, and is introduced here. The system has an extensive representation for experiments and uses experiments to: (i) confirm existing hypotheses; (ii) find new hypotheses; (ii) enhance its own performance; and, (iv) make intractable hypotheses tractable.

1 Introduction

As a way to investigate the nature of scientific thinking and discovery, modelling episodes of scientific discovery in computer programs is now well established in Artificial Intelligence and Cognitive Science. For example, the BACON program (Langley *et.al.* 1987) has shown how laws, such as Galileo's law of free fall, can be discovered from empirical data. ECHO (Thagard, 1989) has been used to assess the acceptability of competing mature scientific research programmes, such as the oxygen and phlogiston theories in the history of chemistry.

STERN is a scientific discovery system that continues this line of research. The program is an instantiation of a framework that attempts to characterize the nature of scientific research programmes in a general way. STERN has extensively modelled Galileo's investigations of naturally accelerated motion.

The role of experiments in scientific discovery has not been the major concern in previous work in the area. Typically, the principal manifestation of experiments in existing discovery systems is in the form of correct empirical data. Just a few researchers have considered experiments in more depth (e.g. Kulkarni & Simon, 1988). Hence, I will concentrate on the experimental component of STERN in this paper. First, the extent to which experiments have been considered in previous work will be discussed. Second, we will consider the structure of experiments as posited by the framework and then look at some of the experiments Galileo used in his investigations. Third, the instantiation of the experimental component of the framework in STERN is described. Finally, four of STERN's processes that involve experiments will be considered in detail.

2 Experiments in Previous Work

There are now a great variety of scientific discovery programs. Here, we will just consider the extent to which experiments have been modelled.

In most existing systems the only manifestation of experiments is in the form of "observations" or correct empirical data. There are many programs that have empirical data as input. For example, BACON (Langley et.al., 1987) is given sets of numerical values from which it finds laws. For instance, given data relating to the radius and period of revolution of the planets BACON finds Kepler's third law. Systems that have followed BACON have combined both quantitative and qualitative methods in the formation of laws from data (e.g. Nordhausen & Langley, 1987; Falkenhainer & Michalski, 1986). Empirical data may also be used in other tasks; such as the modification of existing models (e.g. Buchanan & Feigenbaum, 1978; Kokar, 1986; Rose, 1988), or for assessing the acceptability of competing hypotheses in mature research programmes (Thagard, 1989). Although an important part of scientific discovery, the use of empirical data is only a one of many aspects of experimentation.

However, three systems have considered experiments in more detail. The first is Rajamoney *et.al.'s* (1985) program that considers processes involving liquids in containers. The system investigates an unknown processes by designing experiments to differentiate between different processes. For example, by maximizing a liquid's free surface area, whilst minimizing the contact area with the vessel, the system can distinguish between evaporation and absorption. The two other systems are HDD (Reimann, 1990) and KEKEDA (Kulkarni & Simon, 1988). Both have representations of experiments that: (i) specify the independent and dependent experimental variables; (ii) give the values of these variables; and, (iii) supply some details about the particular nature of a given experiment. However, HDD has no heuristics for instantiating its own experiments; all its experiments are given as user inputs. Of KEKEDA's many heuristics, there are several that propose different experimental tests depending on the nature of the hypothesis being investigated.

To summarize, the modelling of experiments has played only a modest part in previous work. Clearly, there is plenty of scope for further investigations. So, let us now consider the nature of experiments in more detail.

3 Experiments

The framework for computational models of scientific discovery was introduced by Cheng (1990). It proposes a minimum set of components as a guide to the construction of acceptable models of scientific discovery. The focal concept is the *research programme;* a body of research that investigates a delimited set of phenomena using a *theoretical component* and an *experimental component.*

Here, our main concern is with the nature of the experiments, but for completeness the theoretical component will be briefly outlined. The framework views theory as the abstract formal characterization of the phenomena within a research programme. Theoretical knowledge is in the form of *state transformation functions* that interrelated the *states* of the phenomenon - sets of values for characteristic attributes of the phenomenon. Three types of theoretical knowledge are distinguished: (i) *hypotheses* that characterize the phenomenon in all its different manifestations; (ii) *models* to characterize the phenomenon in just one situation; and (iii) *instances* that are series of states. The theoretical component also identifies different classes of theoretical inferences and considers criteria for assessing acceptability of theories.

We now consider the structure of experiments in detail.

3.1 The Structure of Experiments

The framework gives a general abstract conception in which experiments are mechanisms that treat phenomena as "black boxes". The scientist investigates a phenomenon via a set of specified inputs and outputs. The inputs attempt to control some aspect of the phenomenon and manipulate others, while the outputs reveal values that result from these particular inputs. In an experiment, a phenomenon is instantiated in a manner that allows input parameters (Inputs-M) to be manipulated and output parameters (Outputs) to be measured or observed. Some input parameters are fixed (Inputs-C), they are held constant to tightly control the experimental environment. The form of experiments can thus be represented by the equation:

E(Inputs-E(Inputs-C) = Outputs . . .(0)

where the phenomena in the black boxes determine the hidden functional relation, £, between the Inputs-M/Inputs-C and Outputs. In this scheme experimental apparatus is required to instantiate and manipulate the phenomena, and instruments are needed for measurement and observation. Ideally, just one Input-M should be manipulated at a time when performing an experiment to prevent ambiguity over the extent to which a parameter affects the phenomena.

More specifically this characterization of experiments is treated at three levels of generality in the framework: as *experimental paradigms, experimental setups* and *experimental tests.*

At the most general level, within most sciences there are distinguishable classes of experimental situations, which are quite different ways a phenomenon can be investigated within a research programme. These classes of experiments are called experimental paradigms.

At a more specific level there are experimental setups. These are instantiated experimental paradigms; manufactured experimental apparatus and instruments for manipulating and measuring Input and Output parameters, respectively. Different experimental setups provide variations on the way a phenomenon is instantiated, manipulated and observed a under a particular experimental paradigm.

Finally, at the most detailed level one has specific experimental tests. An experimental test refers to a particular experimental trial in an experimental setup. In an experimental test particular variables are chosen to be the Input-M, Input-Cs and Output parameters. The experiment is then performed with a series of Input-M values for which Output values are recorded, with fixed Input-Cs values.

In addition to the structure of experiments, the framework also considers: the processes required for the performance of experiments; the genesis of experiments; and the assessment of the reliability of experimental results. To make the characterization of the structure of experiments more concrete let us consider the experiments used by Galileo.

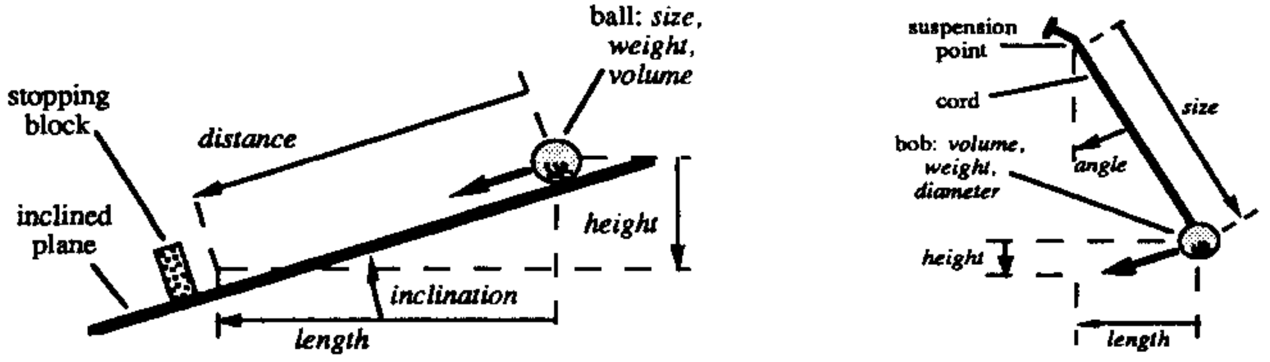
3.2 Galileo's experiments

Galileo is often considered to be the first scientist in the modern sense of the term because he not only theorized about phenomena but also performed experimental investigations. In his studies of naturally accelerated motion, the most important experimental paradigms used by Galileo include (Galileo, 1838): (i) swinging pendulums consisting of small weights attached to the end of long suspended cords (MacLachlan, 1976); and (ii) inclined planes, or ramps, made from long straight wooden batons along which spherical metallic balls are rolled (Settle, 1961). Figures la and lb show these two experiments schematically and indicate some of their parameters. An inclined plane, for example, provides an experimental setup in which several different types of experimental tests can be performed. For example, the setup may be used in different tests to investigate how the *distance* (Input-M) down the plane varies with *time* (Output), or how the *height* (Input-M) affects the *time* (Output) with the distance held constant (Input-C).

Pragmatic knowledge plays an important part in the use of all kinds of experiments. The relative ease of manufacture of experimental setups from particular paradigms plays a role in the selection of the setups. A pendulum is simpler to construct than an inclined plane. In an inclined plane experiment, *distance* can be determined from markings made on the side of the plane but *time* is measured with a water clock. Obviously *distance* is simpler to manipulate and measure so it is chosen as the Input-M parameter, forcing *time* to become the Output parameter.

Further, background knowledge also has an important role to play. When choosing which parameters in a given experimental setup to make the Input-M and Output it is essential to ensure that they are not trivially related, that is, tautologically or by definition. For example, when the *inclination* of an inclined plane is fixed, the *distance, height* and *length* will vary in proportion to each other just because of the physical geometry of the setup. However, simple geometrical knowledge can be used to infer that such combinations of parameters are completely independent of motion phenomena.

Galileo's skill as an experimental scientist is shown by



(a) Inclined Plane (b) Pendulum

Figure 1 Galileo's Inclined Plane and Pendulum Experiments (Experimental parameter names in *Italics)*

his invention of new experimental paradigms. The basic technique he employed was to combine known experiments using the Output of one to feed into another. For example, Figure 2 shows the combined projectile and inclined plane experiment (Drake & MacLachlan, 1975; Drake, 1975). In this experiment a ball descends an inclined plane, *PQ.* It is launched into the air with an imposed initial horizontal motion by the lip at Q, and describes a free path as a projectile until it lands at R. The first half (inclined plane) of such combined experiments will be called the *initial part* of the experiment, and the second half (projectile) called the *terminal part.* There are two ways, or *modes,* in which combined experimental setups can be used in tests. In the *initial mode* the Input-M is chosen from the initial part of the experiment, and the Output from the terminal part; for example *height* as Input-M in the inclined plane and horizontal projectile *length* as the Output. The *terminal mode* focuses just on the terminal part, with both Input-M and Output being parameters chosen from that part; for example, the projectile's *height* and *length* could be the Input-M and Output, respectively. The initial part's Output acts as a Input-C parameter to the terminal part. Galileo carried out investigations on combined experiments using the initial and terminal modes.

Experiments are not simple a matter. Different levels and types of experiments have an important part to play in making scientific discoveries. In the following two sections we consider how STERN models most of the aspects of experiments just described.

4 Representing Experiments

STERN considers all three levels of experiments posited in the framework (see §2.1). STERN has schemas for experimental paradigms, setups and tests; and also for experimental parameters. All are instantiated as frames.

Experimental paradigm frames have slots for information associated with each paradigm. The information includes: the name of the paradigms (e.g. 'incplane' for the inclined plane); lists of the relevant experimental parameters; what experimental setups available under the paradigm; the ease of setup manufacture (a number in the range [0 1]); the mappings between the parameters and the variables used to express background knowledge; and, details to distinguish the initial and terminal parts of combined experiments. The experimental setup frames have slots for: the name of setups (e.g. 'down_incplane'); the parameters specific to a setup; experimental tests; and, the name of the initial part of the setup in the case of combined experiments. The experimental test frames has slots for the Input-M, Output and fixed Input-C parameters and their values, and a slot to indicate the mode in which combined experiments are used.

The experimental parameters employed by experimental paradigms and setups are themselves frames. Some of the parameter's slots are particularly relevant to our present concerns. There are slots that name the parameter and hold its current value when required. Two slots indicate the maximum and minimum permitted values; restrictions on the magnitudes of a parameter imposed by the physical dimensions of the experimental apparatus. Finally, there is a slot containing a measure of how easy it is to manipulate

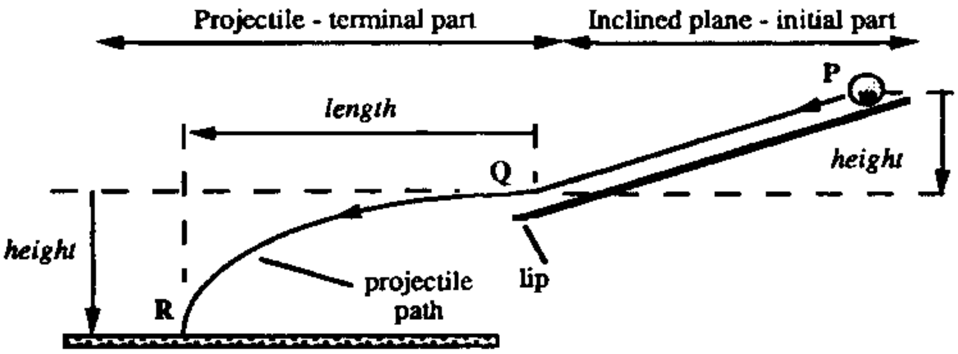


Figure 2 Galileo's Combined Inclined Plane And Projectile Experiment (Experimental parameter names in *italics)*

and observe the parameter.

Thus, STERN has four different frames that instantiate most of the aspects of experiments considered in the Section 3. We will now see how these various experimental frames are used by STERN.

5 STERN'S Experimental Abilities

In general terms, STERN has a production system architecture, but is not a typical example of a production system. A complex hierarchy of schemas is used for its memory. For example, on the experimental side of STERN, lists of experimental tests are stored under experimental setups, and lists of setups under experimental paradigms. A total of 64 rules are employed to carry out STERN's various processes. The rules are partitioned into a hierarchy of 16 groups that constitute tasks on various different levels. For example, a high level task is the confirmation of existing hypotheses, and a more specific task is the comparison of a single theoretical prediction and an experimental test result

Four main experimental abilities are present in STERN: (i) using experiments to (dis)confirm hypotheses; (ii) experiment-led generalization to hypotheses; (iii) controlling the availability of experiments as means to enhance performance efficiency; and (iv) constructing and using new experiments to overcome odierwise intractable hypotheses.

5.1 Experiments to Disconfirming Theories

One of STERN's high level tasks, or strategics, is the confirmation (or disconfirmation) of an existing hypothesis. Basically, this involves making a prediction, obtaining an experimental test result, and comparing the two. However, to be able to make a valid comparison the prediction must be relevant to the experiment; specifically, the theoretical terms mentioned in the prediction must correspond to measurable experimental parameters in the active experiment. Thus, STERN closely integrates its experimental and theoretical processes during this task.

STERN first selects an existing hypothesis that has not been previously tested. One such hypothesis is the Aristotelian *effective weight* law - the speed of a body is proportional to its "effective weight", or density. Leaving aside the proportionality constant, the hypotheses can be expressed thus,

$$V = DEN, \qquad \ldots (2)$$

where V and DEN are the speed and density of the body, respectively.

STERN then chooses an experimental paradigm that has not already been considered with the current hypothesis and that seems the most profitable to use. A record of the experimental paradigms used to test each hypothesis is kept by STERN. Of those not previously considered, STERN favours paradigms that combine ease of manufacture with the most experimental setups. The pendulum paradigm is typically preferred, but let us consider the inclined plane for our on going example.

As the current aim is to test the hypothesis using the experimental paradigm, STERN checks whether V and DEN correspond to experimental parameters that are directly measurable. However, in this case neither do, so STERN tries to find expressions in other terms to substitute for V and DEN using one of two different methods. First, STERN can replace a term by its own definition, when preconditions attached to the definition obtain. For V, the condition that V is constant is satisfied because the Aristotelian *instantaneous acceleration* law is assumed to hold at the time. Second, background knowledge can be used to replace a term, by searching through the various relations in each different set of background knowledge, for expressions equivalent to the term. This is the case with DEN. The equation that STERN finally infers from Equation 2 is,

$$D / T = W / VOL, \qquad \ldots (3)$$

where D and T correspond to the *distance* and *time* of travel, respectively, and W and VOL to the *weight* and *volume* of the ball, respectively.

STERN next selects a particular inclined plane experimental setup and then attempts to make specific predictions. STERN considers pairs of terms from Equation 3, in turn, as the basis for making predictions. For example, let us take D and T. By reference to the experimental setup, D is chosen as the independent term, because its corresponding *distance* experimental parameter is the most easily manipulated and measured. T becomes the dependent term. A series of values for T and D are calculated using Equation 3. The range over which D is permitted to vary is found from the magnitude limits of its corresponding experimental parameter. W and VOL are given values equal to the mid point in the range of their corresponding experimental parameters. Thus a quantitative prediction has been fully specified.

The next stage is to design an experimental test that matches the prediction. STERN chooses Input-M and Output parameters that correspond to D and T, respectively. The values of the Input-M and Input-Cs are set equal to the values of D, and W and VOL, respectively. A subprogram then simulates the performance of the experimental test. It returns Output parameter values with a realistic amount of noise. (In other systems the user typically supplies the results of experimental tests.)

Thus, STERN finally has a prediction and an experimental test result that can be validly compared with each other. A function combining two forms of correlation analysis is used to simultaneously measure how accurately the values of the Output parameter and dependent term match and to assess the amount of noise in the Output. This *predictive accuracy* value is in the range [0 1]. The whole processes is repeated with predictions based on other pairs of theoretical terms, each yielding an instance with a value of predictive accuracy. The acceptability of the model. Equation 3, is given by the quotient of the sum of the instance accuracy values and the number of instances. The acceptability of this model, and any others derived from the original hypothesis, are in turn used to assess the hypothesis, Equation 2. Hypothesis acceptability is given by the quotient of the sum of the model acceptability values and the number of models considered. Hence, the measure of acceptability of hypotheses is a function of the adequacy of its models (and instances) that also reflects any uncertainty due to the presence of noise in the experimental data.

The confirmation strategy is used many times throughout the modelling of Galileo's discoveries. Aristotelian laws like Equation 1 are found to be unacceptable using it. We will now consider how experiments also have an important part to play in the generation of new hypotheses.

5.2 Experiment-led Generation to Hypotheses

As we saw above, many previous scientific discovery systems have been successful in the generalization of empirical data into laws. STERN also infers new theoretical knowledge by generalization, but does so starting with experimental paradigms.

STERN first chooses an experimental paradigm, such as the pendulum paradigm. The pendulum's setups are considered in turn. For each, STERN designs a series of experimental tests based on pairs of parameters, which are to be the Input-M and Output of different tests. Initially, STERN considers all combinations of parameter pairs and then eliminates those that are trivially related or that have no bearing on the phenomenon being investigated. Consider for example, three pairs of pendulum parameters: *(size time), (angle weight)* and *(angle length)* (see Figure Ib). Trivially (tautologically) related pairs of parameters are found using background knowledge. For example, *(angle length)* is eliminated, because the *angle* of swing is related to the *length* merely by the physical geometry of pendulums. The pairs of parameters that are considered irrelevant are ones that have no bearing on motion phenomena: that is pairs that do not have a *distance, time* or *speed* parameter. The *(angle weight)* pair is eliminated for this reason. Of the 36 pairs that STERN originally considers for the pendulum setup, only 15 remain after the two methods are applied; *(size time)* is one of them. The space of experimental tests that STERN must consider has been significantly reduced.

For each of the remaining pairs of parameters STERN designs experimental tests. Given the *(size time)* pair, STERN makes *size* the Input-M and *time* the Output, because *size* is more easily manipulated. All other parameters are treated as Input-Cs. A series of Input-M values are calculated using the maximum and minimum permitted magnitudes of the *size* parameter. The Input-C parameters are set to their mid range values. The experimental test is then performed by the experiment simulator subprogram, which returns the values of *time* for each *size* value.

To allow the theoretical side of the process to generalize the results, STERN translates the lest results into theoretical terms: that is, *time* becomes T and *size* becomes S. The task of generalizing data into models is analogous to what BACON does (Langley et.al„ 1987). Later, when further models have been found using other experimental paradigms, those that are sufficiently general become hypotheses. Many qualitative hypotheses, and the several quantitative models, are found using this experimental-led generalization strategy. One such model is the law governing the relationship between the length *(size)* of a pendulum and its period of swing *(time).*

5.3 Controlling the Availability of Experiments

We have seen how STERN can use experiments to confirm hypotheses, and how new theoretical knowledge is gained in an experiment-led manner. Here we consider a quite different feature in STERN - controlling the availability of experiments to improve performance efficiency.

STERN is given six experimental paradigms as its initial experimental input and it is also able to construct new combined experiments (see below). Whilst engaged in the two strategies considered in the previous subsections, the system could consider every experimental paradigm in turn, but this would be quite inefficient. What STERN actually does is to limit the number of available experimental paradigms. For example, attempting to initially confirm a hypothesis using just two experimental paradigms saves considerable effort. Unacceptable hypotheses can be been found using just two paradigms and thus eliminated from further investigation. Any acceptable hypotheses can be further tested using other experimental paradigms, but redundant processing is avoided by not doing the same for hypotheses already shown to be unacceptable.

A mechanism in STERN limits the number of experimental paradigms that are available. A pragmatic measure of how worthwhile a paradigm is likely to be is calculated from the ease of manufacture its experimental setups and the number of setups. Hence, when controlling the number of available experimental paradigms, only those with a practicability above a certain limit are made active. Initially, the limit is chosen (by the user) so that two experimental paradigms will be available. When the first two experimental paradigms have been exhausted, the mechanism makes new paradigms available by lowering the value of the pragmatic limit.

This is an example of how the inclusion of experiments in a scientific discovery system not only makes it a more complete model, but also allows new heuristics to be devised for improving program performance.

5.4 Constructing and using novel experiments

During the modelling of the Galilean episode, STERN infers many new hypotheses from those already obtained using the generalization strategy. One such new hypothesis is the most general form of the law of free fall,

$$V = H^{1/2} , \qquad \ldots (4)$$

where H is the vertical distance. STERN must test the hypothesis using experiments to see whether it is really acceptable, using the confirmation strategy (see §5.1). However, when attempting to check the new hypothesis, STERN finds that is it intractable because V cannot be eliminated from Equation 4. Whereas, previously V was substituted using its own definition, this can no longer be done as the condition that V be constant is no longer satisfied (earlier Aristotle's *instantaneous acceleration* hypothesis fulfilled the condition, but by now it has been shown to be unacceptable).

STERN decides to construct a new combined experiment as a means to overcome this problem, just as Galileo did (see §3.2). For example, consider the combined inclined plane and projectile experiment (Figure 2). Separate equations that include V can be found for each of the experiment's two parts. The speed down the inclined plane is given by Equation 4. The horizontal speed of a projectile is constant, so an equation in terms of horizontal distance, speed and time can be used. Now, as the lip at the end of the inclined plane sends the ball horizontally into the air, this means that the speed terms in both equations are equal, and can thus be eliminated by substituting one equation onto the other. Hie rest of the confirmation strategy may then be applied as usual.

To be able *invent* combined experiments it is necessary to use real world knowledge; for example, that a lip can be attached to the end of the inclined plane. Such abilities are beyond the current version of STERN, which is given as inputs the legal combinations of combined experiments. However, STERN is able to construct new paradigms given this information.

To manufacture a new combined experimental paradigm, STERN finds experimental paradigms that can act as the terminal part of combined experiments. A paradigm, say the projectile paradigm, is made active. For this terminal paradigm, the other paradigms with setups that can be initial parts are also made active. For example, the inclined plane paradigm is a suitable initial part for the projectile as the terminal part. The actual construction of the combined experimental paradigm involves instantiating a new experimental paradigm frame and filling its slots using the information available for the active initial and terminal paradigms. For example, the ease of manufacture of the new combined experimental paradigm is calculated thus:

$$\text{manufacture ease} = (i \cdot t) / (i + t), \qquad \ldots (5)$$

where i and t are the manufacturing ease of the chosen initial and terminal paradigms, respectively. Equation 5 satisfies the conditions that the ease of manufacture is: (i) between 0 and 1; and (ii) less than the magnitude of either i or t alone.

STERN uses the new combined inclined plane and projectile experimental in the confirmation strategy to assess the acceptability of the law of free fall. The combined experiment is used in both the terminal and the initial modes. In the terminal mode, which focused only on the projectile part of the experiment, STERN discovers the correct equation describing the parabolic shape of the projectile path. All this models what Galileo also did with this combined experiment

That ends our consideration of the experimental abilities in the STERN discovery system. (STERN also has a purely theoretical strategy for generating new quantitative hypotheses from existing acceptable and unacceptable qualitative and quantitative hypotheses; Cheng, 1990).

6 Conclusions

We have seen how STERN has modelled several aspects of the role of experiments in Galileo's investigation of the naturally accelerated motion. The experimental procedures are closely integrated with STERN'S various theoretical inference processes and are crucial to the system's overall discovery abilities. This research has begun to examine the important part that experiments have in scientific discovery, and demonstrates that there is much interesting research yet to be done. The current work on STERN is concerned with the modelling of other Galilean research programmes, such as the strength of materials. Future work on experiments may concentrate on methods for assessing the reliability of experimental results, or model the physical structure of experimental setups in sufficient detail to enable STERN to *invent* experiments. Both issues will involve the development of more sophisticated control strategies, that will take into account previous theories and experiments, and model the scientist's aims and expectations.

Acknowledgements

This research was carried out under a studentship and a postdoctoral fellowship from the Science and Engineering Research Council. Thanks should go to: Mark Keane, Marc Eisenstadt and everyone in HCRL at The Open University for all their support during my PhD research; and, Herbert Simon for his comments on this paper.

References

Buchanan, B.G. & Feigenbaum, E.A. (1978). Dendral and MetaDendral: Their applications dimension. *Artificial Intelligence,* 11, 5-24.

Cheng, P.C-H. (1990). Modelling Scientific Discovery. Technical Report No.65 (PhD Thesis), Human Cognition Research Laboratory, The Open University, Milton Keynes, England.

Drake, S. (1975).Galileo's new science of motion. In M.L.R. Bonelli & W.R. Shea. *Reason, Experiment, and Mysticism.* London: MacMillan Press.

Drake, S. & MacLachlan, J. (1975). Galileo's Discovery of the parabolic trajectory. *Scientific American,* 232(3),102-110.

Falkenhainer, B.C. & Michalski, R.S. (1986). Integrating quantitative and qualitative discovery: the ABACUS system. *Machine Learning,* 1(4),367-401.

Galileo (1838, 1954). *The two new sciences.* New York:Dover.

Kokar, M.M. (1986). Determining arguments of invariant functional descriptions. *Machine Learning,* l(4),403-422.

Kulkarni, D. & Simon, H.A. (1988). The processes of scientific discovery: the strategy of experimentation. *Cognitive Science,* 12, 139-75.

Langley, P., Simon, H.A., Bradshaw, G.L. & Zytkow, J.M. (1987). *Scientific Discovery: Computational Explorations of the Creative Process.* Cambridge, Mass.: MIT Press.

MacLachlan, J. (1976). Galileo's experiments with pendulums: real and imaginary. *Annals of Science,* 33,173-85.

Nordhausen, B. & Langley, P. (1987). Towards an integrated discovery system. *Proceedings of the Tenth International Joint Conference on Artificial Intelligence.*

Rajamoney, S., Dejong, G., & Faltings, B. (1985). Towards a model of conceptual knowledge acquisition through directed experimentation. In *Proceeding of the Ninth International Joint Conference on Artificial Intelligence.*

Reimann, P. (1990). *Problem Solving Models of Scientific Discovery Learning Processes.* Frankfurt am Main: Peter Lang.

Rose, D. (1988). Discovery and Theory Revision via Incremental Hill Climbing. *Proceedings of the International Workshop on Machine Learning and Meta Reasoning.*

Settle, T.B. (1961). An experiment in the history of science. *Science,* 133 (19-23).

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences,* 12 (3), 435-67.