Modelling Inductive And Deductive Discovery Strategies In Galilean Kinematics[1]

Peter C-H. Cheng
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213
pc2n+@andrew.cmu.edu

Abstract

This paper investigates how different strategies affect the success and efficiency of scientific discovery, by examining different approaches in Galilean kinematics. Computational models with biases for inductive or deductive approaches to discovery were constructed to simulate the processes involved in finding coherent and empirically correct sets of laws. The performance of the models shows that the best overall strategy is to begin with an inductive bias and then perform tight cycles of law generation and experimental testing. Comparison of the models with previous findings indicates that the best overall strategy for discovery depends on the relative ease of search in hypothesis and experiment spaces.

1 Introduction

Scientific discovery is an important and growing area of research in the cognitive sciences. A major issue in the area concerns what strategies scientists use to make discoveries and the effectiveness and efficiency of those strategies. Much of the empirical work has focussed on the biases that seem to exist in the process of seeking evidence to assess hypotheses, Gorman (1992). The work of Klahr and Dunbar (1988) and Klahr *et. al* (1990) addresses discovery strategies from a wider context. Their studies employed a simulated discovery environment consisting of a toy robot, BigTrak, controlled using a LOGO-like programming language. The task was to determine how a mystery programming key functioned by writing programs incorporating the key and observing the subsequent behavior of BigTrak. It was found that subjects could be classified as either experimenters or theorists according their preference to search the space of experiments or the space of hypothesis. Theorists were more successful and faster, because they generated new hypotheses using relevant prior knowledge. Experimenters were less effective and efficient, because they laboriously performed experiments and attempted to generalize the results into hypotheses. Klahr and Dunbar characterize scientific discovery as the dual search of the hypothesis and experiment spaces. The best discovery strategy for this task is initially to generate several hypotheses and then test them experimentally.

In contrast to the work of Klahr and colleagues, this paper describes computational work that models different strategies in Galileo's discovery of the laws of free fall. A large number of computational systems already exist that model many aspects of scientific discovery (see Cheng, 1992a, for a review). The conventional computational approach attempts to demonstrate the acceptability of a complete model by simulating one or more episodes of discovery. Two relevant examples are: Kulkarni and Simon's (1988) KEKEDA system, which demonstrates that prompt investigation of surprising experimental outcomes is a good strategy; and, Cheng's (1990, 1991) STERN system that has previously modelled Galilean kinematic discoveries. However, the approach adopted here is different. The modelling will focus on a particular factor that is important in discovery by constructing models that differ with respect to the factor but are otherwise as similar as possible. Thus the difference in the models' performance will directly demonstrate how the factor affects the success and efficiency in a particular discovery task. This approach was previously used to demonstrate the computational benefits of using diagrams in discovery, by comparing models using diagrammatic and conventional mathematical representations (Cheng, 1992b; Cheng & Simon, 1992). Here the models have been given biases for either inductive or deductive approaches to discovery, to examine how the approaches influence the overall success and efficiency of discovery in the Galilean domain.

We will begin by considering Galileo's kinematic discoveries.

[1]Thanks should go to Herbert Simon for the discussions that helped to shape this research. This work was supported by a UK Science and Engineering Research Council postdoctoral fellowship.

Figure 1

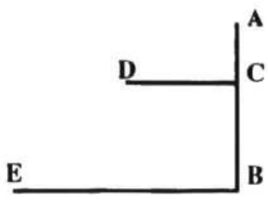

Figure 2

2 Galilean Kinematic Discovery

Although historians of science do not agree upon the precise manner in which Galileo found the laws governing the motion of bodies in free fall, there is reasonable agreement on the main stages of the episode (e.g., Drake, 1973, 1975, 1978; Drake & MacLachlan, 1975; Hill, 1988; Naylor, 1974). We take up the story from the point where Galileo had rejected the Aristotelian views of motion but incorrectly believed the speed of a naturally accelerated body to be in proportion to the distance travelled from rest, which may be expressed as an equation thus;

$$v_{ab}/v_{ac} = d_{ab}/d_{ac}\,, \qquad \ldots (1)$$

where v and d are speed and distance, respectively, and ab and ac indicate two different falls from rest. Galileo typically stated quantitative laws as sentences referring to ratios of similar variables, but for ease of comprehension equations will be used here, without affecting the claims being made. Galileo eventually found the correct laws of motion;

$$v_{ab}/v_{ac} = t_{ab}/t_{ac}\,, \qquad \ldots (2)$$

$$d_{ab}/d_{ac} = t_{ab}^2/t_{ac}^2\,, \qquad \ldots (3)$$

and, $$v_{ab}/v_{ac} = d_{ab}^{1/2}/d_{ac}^{1/2}\,, \qquad \ldots (4)$$

where t is time. The precise manner of their discovery is uncertain, the historians have conjectured many different paths, but it is clear that Galileo used a combination of deductive and inductive methods. They are considered in turn.

Galileo (1974) published his kinematic findings in the Third and Fourth Days (sections) of the *Two New Sciences*, TNS hereafter. The Third Day has two subsections. The first concerns constant speed motion, and presents laws relating speed, time and distance when acceleration is absent. For example, the fourth and sixth propositions are, respectively;

$$d_{ef}/d_{gh} = v_{ef}/v_{gh} \cdot t_{ef}/t_{gh}\,, \qquad \ldots (5)$$

and, $$v_{ef}/v_{gh} = d_{ef}/d_{gh} \,/\, t_{ef}/t_{gh}\,, \qquad \ldots (6)$$

where the variables have the same meanings as above, the values of v are constants, and the subscripts denote two different bodies or paths. The second subsection of the Third Day concerns the accelerated motion of bodies under many different circumstances. Beginning with a single postulate, a definition, and the laws of constant speed motion, 38 propositions are derived. The definition happens to be Equation 2, and the second proposition (TNS-III-2) and its corollaries are equivalent to Equation 3. TNS-III-2 is derived from TNS-III-1, which states that the mean speed of a body falling from rest is one half its maximum speed. As the speeds of two bodies are in proportion to their times, Equation 2, their mean speeds will thus be in proportion to the times. However, as the mean speeds are both constant by definition, Equation 5 holds, so times can be substituted for speeds to give Equation 3. Another proposition of TNS-III is the *double distance* law, TNS-III-9. Referring to Figure 1, if a body falls from rest through distance AB and is turned through a right angle so that it travels horizontally along BC, then the distance BC will be double that of AB when the time for AB equals the time for BC. This proposition is used in the derivation of Equation 4, considered next.

In TNS-IV-3 Galileo derives the relation for speed and distance, Equation 4. For vertical falls from rest over two distances AC and AB, Figure 2, the double distance law can be applied twice to distances CD and BE. As the speeds along CD and BE are constant their ratio is given by Equation 6,

$$v_{cd}/v_{be} = d_{cd}/d_{bfe} \,/\, t_{cd}/t_{be}. \qquad \ldots (7)$$

However, from the double distance rule we know: (i) the ratio of the times CD to BE is equal to the ratio of the times AC to AB; (ii) distance CD is double AC and distance BE double AB; and, (iii) the speed at end of the fall AC equals the speed along CD, and similarly for AB and BE. Therefore, the times, distances and speeds of AC and AB can be substituted into Equation 7;

$$v_{ab}/v_{ac} = d_{ab}/d_{ac} \,/\, t_{ab}/t_{ac}. \qquad \ldots (8)$$

Now, Equation 8 relates distances and times in free fall, so the times can be eliminated using Equation 3 giving Equation 4, so completing the derivation of the set of three laws[2].

The classes of motion experiments that historians are sure Galileo had at his disposal were pendulums, inclined planes or ramps, and projectiles. In the inclined plane experiments Galileo rolled a ball down the plane and measured how the distance along the plane varied with time. To perform experiments on projectiles, Galileo used an inclined plane with a lip at its end to launch the ball horizontally into the air. This combined inclined plane and projectile experiment allowed Galileo to determine speeds using Equation 6, because the horizontal speed of a projectile is constant and the time of fall is also constant when the vertical distance is fixed. Cheng (1991) describes the experiments in more detail. The

[2]Why Galileo did not just substitute for t in Equations 2 and 3, to derive Equation 4, is a mystery.

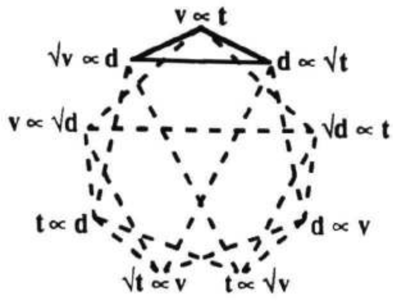

Figure 3

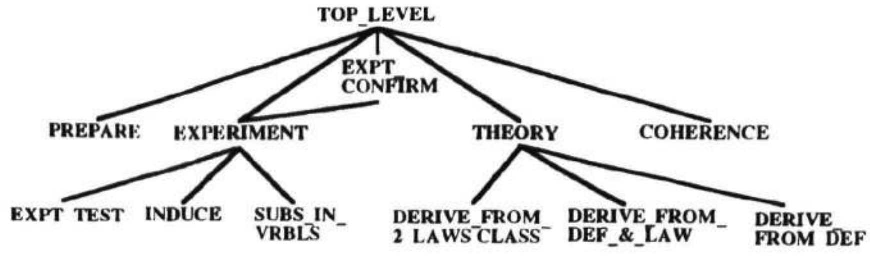

Figure 4

specific purposes behind the experiments are disputed because Galileo (1979) only left terse and cryptic records of them, but is possible that they could have been performed in either an inductive manner or a confirmatory mode. Galileo was a competent mathematician so was able to find expressions describing experimental data or to judge how well numerical predictions and data matched.

We will now define more precisely the discovery task to be modelled.

3 Simplified Discovery Task

A simplified version of the Galilean episode is considered for two reasons. First, it is necessary to reduce the complexity of the problem to make the construction of models and the performance analysis a practical proposition. Second, the simplification will focus on the aspects of the discovery with the greatest historical certainty.

The discovery task is to find an acceptable set of laws relating the speed, distance and time of a body in accelerated motion, given the laws of constant speed motion. The laws to be found will be power function of pairs of variables. A set of three laws is necessary to cover all the combinations of pairs of variables and a set is acceptable if it is coherent and empirically correct. A set is coherent if from any two laws the third can be validly inferred by elimination of a shared variable; which is the case for Equations 2, 3 and 4. Laws are empirically correct when they match experimental data with sufficient accuracy.

To clarify the nature of the discovery task, consider a subset of the space of laws in which the variables are linearly related or one variable is in proportion the square-root of the other. There are nine laws in the subset, so 84 ($=9!/(6!\cdot 3!)$) combinations of sets of three laws are possible. Figure 3 shows the nine laws (using proportionality signs rather than ratios of variables). The task is to find an acceptable set from amongst all the (84) possible sets. The triangles indicate the six sets of coherent laws. The uppermost triplet, with the solid triangle, is the empirically correct set. This characterization of the task suggests that enumeration of all the coherent sets may be an effective way to begin tackling the problem, because of the substantial reduction in the search space (from 84 to 6, in the example). However, the models will take a more historically constrained approach, employing methods that Galileo would have had at his disposal.

4 Program and Models

The basis for the models is a single production system that can be given a bias for either an inductive or a deductive approach to discovery. Care has been taken in the specification of the knowledge structures and productions to ensure a close match to the inferences steps seen in the historical material. The declarative knowledge representations will be considered first followed by a description of the various inference processes of the model.

The knowledge representations are simple predicate-argument-like structures, that can be considered at three levels. Information on the highest level relates to the overall task of discovery and includes: a list of the relevant variables (e.g., speed distance time); the permitted indices of the power functions (e.g., 1, 1/2, 2); the type of phenomenon (i.e., free fall accelerated motion); and, the identifiers of sets of laws that have been considered (nil initially). These top level items are given as run-time inputs. The second level of knowledge has information required for inferences involving sets of laws, including: the identifier of the set under active consideration; the pairs of variables for each law; and, a statement about a set's coherence, when tested. The lowest level concerns individual laws and associated information, including: the law's power equation; and records of the status of the deductive and inductive inferences made.

Figure 4 shows the hierarchy of classes of rules used by the program. The TOP_LEVEL rules control the overall operation of program by invoking the one of five classes on the next level. These classes are briefly described before returning to TOP_LEVEL

The PREPARE rules place second and third level knowledge structures into working memory for use by other processes, so it is the first class to be invoked by TOP_LEVEL when a simulation begins. It is also invoked whenever a set of rules is rejected to initialize a new set of structures. PREPARE may generate a new law as a definition based on a pair of preferred variables and an index not previously considered with the variables.

The EXPERIMENT class attempts to find laws by performing experiments. Descriptions of experimental tests, given as inputs to the program, specify which properties can be employed as *manipulated input* and *measured output* variables (Cheng, 1991). For example, time is the manipulated input and distance the measured output in an inclined plane experiment. Given a pair of variables EXPT_TEST is called to find an experiment test with matching input and output variables. When there is no direct match an alternative variable may be used as the input or output, if there is a known relationship between it and the given variable. For example, horizontal distance in the combined inclined plane and projectile experiment can be used to measure speed. When a suitable test is found, a special rule places the experimental results in working memory, as if the experiment had been performed. The INDUCE class is then engaged to seek a simple power law that accounts for the numerical data. Finally, the SUBS_IN_VRBLS class is employed to substitute variables in to the power law when necessary; e.g., speed for distance in the inclined plane and projectile experiment.

The function of EXPT_CONFIRM is to test experimentally a law found deductively. EXPT_CONFIRM simply invokes EXPERIMENT, described above, to induce a law that is compared with the derived law to see if they are the same. Galileo would have made a quantitative prediction from the derived law and compared this with the experimental data, rather than generalizing the data and comparing laws. However, the difference is minor and does not have a significant bearing on the conclusions to be drawn; so for simplicity it is acceptable.

The THEORY class is the deductive module of the program and calls three other classes. DERIVE_FROM_DEF and DERIVE_FROM_DEF_&_LAW attempt to find a law from a definition, or from a definition and an existing law, respectively. The two classes include rules instantiating constant speed laws and generalized versions of the inferences that Galileo employed in the TNS. The rules are general in two ways: (i) they employ variables for predicates and arguments; and, (ii) they consider any power function rather than a particular relation. General rules mean that the correct accelerated motion laws are not implicitly built into the deductive mechanisms of the system. For example, TNS-III-1 states that the mean speed is one half the maximum speed of a uniformly accelerated body, as the speed increases linearly with time. The equivalent rule is (translated into pseudo-English):

```
If the set of laws G is being considered,
   and the law M is being considered,
   and M is a definitional law of the form
      SSxy^p/SSxz^p=TTxy^r/TTxz^r,
   and G's context is free fall motion,
   and SSxy and TTxy are known,
Then SSab is the mean of SSxy,
   and SSab is constant with respect to TTab,
   and TTxy=TTab,
   and SSab=n.SSxy where n=p/(p+r),
   and note SSab & TTab can be determined.
```

G and M are names; SS and TT stand for predicates; n, p, and r are numbers; and, a, b, x, y and z are points on paths. This rule determines the mean of SS as a function of its maximum, as it increases as some power of TT.

DERIVE_FROM_2_LAWS attempts to find a new law from two given laws by substitution. All three derivation classes invoke the MATHS rules (not shown in Figure 4) to make simple mathematical inferences.

The remaining class, COHERENCE, tests whether a set of three laws is coherent by examining the values of the indices in the equations. Galileo would probably have performed the test by inspection or possibly by substituting values into the equations.

TOP_LEVEL calls various rule classes under different circumstances. PREPARE is invoked whenever a new set of laws is to be considered. COHERENCE is invoked whenever three laws have been found in the active set. EXPT_CONFIRM follows the COHERENCE class when a coherent set of laws has been found. Either EXPERIMENT or THEORY can be invoked when a set of laws is incomplete depending on the relative priority of the two classes set by the user. When EXPERIMENT has the higher priority it is always invoked in preference to the THEORY class, so making the program a model with an inductive bias. A model with a deductive bias is obtained when THEORY has the higher priority. TOP_LEVEL includes rules to evaluate the acceptability of a set of laws. The program halts when an acceptable set is discovered but a new set is sought when an unacceptable set is found.

That completes the overview of the program that can be employed as models with inductive or deductive biases. The following section considers the performance of the two models.

5 Performance of the Models

The two models have been run with different input conditions. One inductive and two deductive simulations are described. In all three simulations the inclined plane, and the combined inclined plane and projectile experiments, were given as inputs.

The first simulation employed the inductive model and no initial law was given as a definition. Once PREPARE had set up the various knowledge structures, EXPERIMENT was invoked because of the inductive bias. Pairs of variables were considered in turn and suitable experimental tests were found for two of the three pairs. The inclined plane was used to find data relating distance and time; and the combined inclined plane and projectile experiment used to find data relating speed and distance. The data were generalized into two equations, Equations 3 and 4, by INDUCE. No experiments were available to find the relation between speed and time, so THEORY was invoked to find the third law by combining the induced laws, resulting in Equation 2. COHERENCE then determined the laws were coherent, and as two laws matched the experimental data, an acceptable set had been found.

Included amongst the input of the first of the two deductive simulations was one of the correct laws of motion, Equation 2. The system was not told that the law was acceptable and it was used as the basis for the inferences made by THEORY, following the steps in the TNS. First, the relation between distance and time was found using the mean speed relation. Then, the law relating speed and distance was inferred by applying: the definitional law; the law just derived; and, the double distance scenario twice. The set of laws was found to be coherent so EXPT_CONFIRM was called to find whether the laws were compatible with experimental data, which they were. Thus an acceptable set was found. Approximately two thirds more computation, in terms of numbers of productions fired, was required for this simulation than the first, even though one of the correct laws was given as an input.

The second deductive simulation had the same inputs as the first except the given definition was Galileo's earlier incorrect definition, Equation 1. Again two further laws were derived;

$$t_{ab}/t_{ac} = d_{ab}^2/d_{ac}^2 , \qquad \ldots (9)$$

and,

$$v_{ab}/v_{ac} = t_{ab}^{1/2}/t_{ac}^{1/2} . \qquad \ldots (10)$$

The two laws and the definition form a coherent set, but during experimental testing Equation 9 did not match the data. This demonstrates that the deductive approach can generate a coherent set of laws that are not empirically acceptable. The whole set was rejected and PREPARE invoked to begin the search for a new set of laws. PREPARE generated a new law that happened to be Equation 2, so the simulation then followed the course taken by the first deductive run. The effort to generate the first but incorrect set of laws increased the amount of computation over the previous deductive simulation by approximately 70%.

The implications of the performance of the models will be considered next.

6 Discussion

In the present discovery task both inductive and deductive approaches are necessary for successful discovery, irrespective of whether the model has an inductive or deductive bias. This is consistent with the historical picture. Further, and more interestingly, the performance of the models indicates that the most effective strategy is initially to adopt an inductive or experimental approach that covers as many of the pairs of variables as is possible. When deductive inferences have to be made, it is preferable to employ tight cycles of law generation and experimental testing. Three related aspects of the task explain why this is the best strategy. First, the experiments will have to be performed no matter which approach is taken, because empirical data will be required to determine whether laws in a coherent sets are empirically correct. It is a waste of effort to derive several laws under the deductive approach only then to perform experiments from which the laws could have been more easily inferred, by inductive generalization, in the first place. Second, laws induced from experiments rule out a larger part of the search space of laws than those found by deductive inferences. Two experimental laws will uniquely identify a coherent and empirically correct set, whereas two deductive laws only define a coherent set without any indication of empirical acceptability. Third, the amount of work required to derive each law is substantial, and as the derivation does not guarantee the acceptability of the law, it is best to assess the empirical acceptability of a law as soon as it is found to avoid making further inferences based on an unacceptable law. In summary, the strategy aims to minimize the amount of computation that might be wasted in generating laws that are coherent but not empirically correct.

The overall strategy for the kinematic domain contrasts with Klahr and Dunbar's best strategy for the BigTrak discovery context (see §1). They contend that the best strategy is to maintain a bias for hypothesis space search and initially to think of hypotheses before conducting any experiments. Comparison of the tasks and models reveals the reasons for the differences in the strategies. Many of Klahr and Dunbar's (1988) subjects found it relatively easy to think of novel hypotheses, and the correct hypothesis was often among the small number generated. When the correct hypothesis was not one proposed, the initial search of the hypothesis space made it easier for subjects to induce the correct hypothesis from experimental outcomes. In the Galilean task, however, deductively generating sets of laws is difficult, and existing sets do not help in the generalization of experimental data into laws. This suggests that the experimentally lead strategy could be the most effective in the BigTrak task under certain circumstances; specifically, when the correct hypothesis is not a member of the small set of seemingly reasonable hypotheses, but is highly unexpected[3]. Subjects will not be able to think of the hypothesis initially, and considering seemly reasonable but irrelevant hypotheses will be unlikely to help subjects find the correct hypothesis when the experiments are conducted.

To summarize, it seems best to adopt an inductive experimental approach when the search of the hypothesis space for a particular domain is likely to be relatively difficult and computationally expensive. Alternatively, when the relative amount of effort to perform experiments is likely to be great, it is best initially to generate hypotheses before conducting the experiments.

7 Conclusions

The best strategy for the discovery task considered here is initially to take an inductive approach followed by tight cycles of law generation and experimental testing. The comparison of this finding with Klahr and Dunbar's work indicates that the most effective and efficient strategy will depend on the relative ease with which hypothesis and experiment spaces can be searched. How, or even whether, scientists make such evaluations, to aid the selection of the most appropriate strategy for a particular discovery task, remains as an important issue to be addressed by future research. The kinematic domain will be a good starting point for this work as Galileo's knowledge of Euclidean geometry and his ability as an experimenter could have been the basis for such reasoning.

References

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

Cheng, P.C-H. 1991. Modelling experiments in scientific discovery. *Proceedings of the 12th International Joint Conference on Artificial Intelligence*. Mountain View, CA: Morgan Kaufmann.

Cheng, P.C-H. 1992a. Approaches, models and issues in computational scientific discovery. In M.T. Keane & K. Gilhooly. *Advances in the Psychology of Thinking* Vol. 1. Hemel Hempstead, Hertfordshire, UK: Harvester-Wheatsheaf. In press.

Cheng, P.C-H. 1992b. Diagrammatic reasoning in scientific discovery: Modelling Galileo's Kinematic Diagrams. In *Working notes of the AAAI Spring Symposium Series: Reasoning with diagrammatic representations*. Stanford, March 1992.

Cheng, P.C-H. & Simon, H.A. 1992. The right representation for discovery: Finding the conservation of momentum. In D. Sleeman and P. Edwards eds. *Machine Learning: Proceedings of the Ninth International Conference (ML92)*, San Mateo, CA; Morgan Kaufmann.

Drake, S. 1973. Galileo's discovery of the law of free fall. *Scientific American*, **228**:85-92.

Drake, S. 1975. The role of music in Galileo's experiments. *Scientific American*, **232**:98-104.

Drake, S. 1978. *Galileo at Work*. Chicago: University of Chicago Press.

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

Galileo Galilei. 1974. *Two new sciences*. S. Drake trans. Madison, Wisconsin: University of Wisconsin Press.

Galileo Galilei. 1979. *Galileo's Notes On Motion*. S. Drake. ed. Annali dell'Instituto e Museo di Storia della Scienza. Florence: Italy.

Gorman, M.E. 1992. Using experiments to determine the heuristic value of falsification. In M.T. Keane & K. Gilhooly. *Advances in the Psychology of Thinking* Vol. 1. Hemel Hempstead, Hertfordshire, UK: Harvester-Wheatsheaf. In press.

Hill, D.K. 1988. Galileo's early experiments on projectile motion and the law of free fall. *ISIS*, **79**,646-68.

Klahr, D. & Dunbar, K. 1988. Dual space search during scientific reasoning. *Cognitive Science*, **12**:1-48.

Klahr, D., Dunbar, K. & Fay, A.L. 1990. Designing good experiments to test bad hypotheses. In J. Shrager & P. Langley eds. *Computational Models of Scientific Discovery and Theory Formation*, 356-402. San Mateo, CA: Morgan Kaufmann.

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

Naylor, R.H. 1974. Galileo and the problem of free fall. *British Journal For The History Of Science*, 7(26):105-34.

[3]E.g., the mystery key might function in different ways when its numeric parameter is, or is not, a prime.