CYBERNETIC THINKING AND SHARE-PRICE PREDICTION

BY ROGER T. HARTLEY

Submitted for the degree of
Doctor of Philosophy to

The Department of Cybernetics,
Brunel University

1974

CYBERNETIC THINKING AND SHARE-PRICE PREDICTION

The thesis presents the application of cybernetic thinking to the central problem of investment analysis; that of share-price prediction. Cybernetics is seen as an inter-disciplinary study (as opposed to multi-disciplinary) in which the barriers between living and non-living systems are ignored.

Suggestions from two independant studies in investment analysis are taken up and a theory of investment is proposed with a view to utilising the suggestions. The theory is formalised using the simple linear perceptron and methods based on logical calculi are used to analyse the perceptron formulation. The theory is then tested by allowing the perceptron to make predictions and the results of these predictions are discussed in the light of the theoretical analysis. Finally suggestions are made for alternative approaches to investment analysis which could lead to better results.

CONTENTS

PREFACE

Most writing on Cybernetics starts, or at least includes somewhere in its length, a run-down on the history of Cybernetics as the author sees it. This thesis does not, except to highlight my firm belief that Cybernetics is not so much a subject matter as a way of thinking; not so much a body of knowledge as a way of life. This view is certainly not new; in fact most authors allude to a similar tendency of thought. Because of this it seems to me that there is a concensus of opinion growing rapidly on this very point.

With reference to this opinion, it seems that the wheel has come full circle, or will do so soon. The Greek philosophers and scientists were all capable of expressing opinions on any aspect of life. It is true that they are mostly remembered now as specialists, each one having his own special study, but they had an attitude of intellectual curiosity towards the whole of life and, of course, made many great discoveries and displayed much expertise in thought.

In the new world (as opposed to the ancient world) the great scientists were natural philosophers and could turn their minds to almost any aspect of science. Some were moral philosophers as well. Newton, Leibnitz, Leonardo, Gallileo, Gauss - the list is too long to mention more than a few. It has been said that the last great 'all-rounder' was Whitehead and after the turn of the century the specialist scientists really took over. The last vestiges of this era can perhaps be seen in the natural and moral sciences triposes at Cambridge University.

However, a more detailed study reveals that there is little attempt at a general scientific education. What has happened then? The great scientists of today seem to spend a lifetime researching one particular problem in one area of one particular scientific discipline. Perhaps the whole phenomenon of specialisation can be explained using the "skyhook and skyscraper" metaphor of scientific study. The great scientists of the past were establishing the hooks and could turn their mind to many such hooks in a lifetime. Today we are stuck in one of a myriad of parallel skyscrapers (with occasional sorties back to the sky to establish firmer hooks) and it is difficult to see the wood for the trees, if I may be allowed a sub-metaphor. In the race downwards we forgot to look around us as the old philosophers had done.

There is now I think, a concensus that we should take a look around and try to build bridges between the various skyscrapers. What though is there to look at? So much has been written in answer to this question, that I think it is unwise to attempt it here. It is equivalent to asking the question "What is Cybernetics?" which invariably tends to be the opening shot of any discussion, whether with University people or not. There is no one answer, at least not yet. However, in my view the nearest one can get to a single topic is the study of the living/non-living dichotomy. In particular, it is the question "Are human-beings (and animals) different in principle from machines?", and, if they are, "What are those principles?". The term living/non-living dichotomy, although entirely accurate, is rather cumbersome and so I propose to call it the measurement of life, as if life were a variable with small values for insects (how much for rocks?) and higher values for animals. Where would the computers of today fit on such a scale? The mere formulation of the question like this opens up immediate avenues of exploration, having passed through the inevitable preliminary discussion of semantics.

A general study like this must, of course, be inter-disciplinary. That is, it must ignore the boundaries set up by scientific specialisation. Many writers have mentioned this (not that it really needs mentioning at all), from Wiener onwards. "This spirit of generalisation distinguishes Cybernetics from other branches of science." * Unfortunately, much of the work is often merely multi-disciplinary, in that it is "the application of Cybernetic ideas to ....". (This accounts for possibly half the published work on Cybernetics). Perhaps this is inevitable in a young science, as Cybernetics undoubtedly is, but it is then, not surprising that there is a difference in emphasis, sometimes quite marked, between different workers. Arbib has said that if Cybernetics is not

* Trask "The Story Of Cybernetics" 1971.

to be a pseudoscience, workers in it must be competent in one and really conversant with one or more other branches of science.** If this is true, even in the future, then Cybernetics will never be really interdisciplinary. If, as is suggested at the start of this thesis, it is more an attitude of mind, a way of thinking, then this point of view is not necessary. People can be trained to think and work like Cyberneticians and a deep study of a particular science or sciences can come later, if necessary. It is also unfortunate that Cybernetics has become fashionable (can anyone imagine chemistry being fashionable?). That is, the word is fashionable; the aims and hopes, let alone any of the current work are largely unknown. Even CDC* have called their latest creation, which is allegedly the most powerful computer in the world, the "Cyber 76". Perhaps this is a compliment to Cybernetics and shows great understanding and insight. This is doubtful. Publicity can be useful, but too much of the wrong sort can kill anything, especially something so vulnerable as a young scientific discipline.

Returning to the problem of the measurement of life, how is the dichotomy to be studied? One way of course is to ignore the split altogether; look at living things as if they were non-living and vice-versa. This paraphrases the twin approaches adopted by many workers - those of simulation and synthesis. In perhaps more explanatory terms these could be called model-building and artificial intelligence (or artificial behaviour to include machines which reproduce animal-like movements). They are both properly part of Cybernetics. It seems likely that artificial intelligence will be the only way to surpass human intelligence, but I think it will inevitably use simulation methods, and theories of brain functioning as a basis. The opposite way of studying the measurement of life is to try to find factors which clearly distinguish living from non-living things. One which comes to mind immediately is the fact that living things seem to possess a purpose; a goal or a set of goals. Non-living things seem to be "inert" in this sense. However, what is purpose?

** Arbib. "Brains Machines and Mathematics.
* Control Data Corporation.

Can we give machines purpose? and if we can does the machine really possess purpose or is it merely exhibiting purposive behaviour? Cybernetics says that we can because in principle there is no structural or material difference along the measurement of life scale. The difference, instead, is one of organisation which ultimately yields a necessary difference in behaviour.

Thus there is no Cybernetic subject matter unless it is life itself; there is no one Cybernetic method. Instead there is a philosophy of attack, a way of thinking which says that we are now capable of looking at scientific problems (and artistic ones too, it can be said) from all angles simultaneously, each one contributing something to the overall picture.

I can think of no better paradigm of the Cybernetic way of thinking than the one contained in the life and work of Edward Ichnatowicz. Here we have an engineer/ sculptor, a self-taught industrial craftsman who is studying the measurement of life problem by discarding all his preconceptions and scientific and artistic prejudices. He has built a thirty foot high sculpture which behaves, although in an admittedly trivial fashion, in a very life-like manner. Its movements (servomechanisms and hydraulics) are animal-like and it reacts to sounds in the room in a typically animal-like inquisitive fashion. It is, simultaneously, a work of art of great beauty (in the field of kinetic sculpture), and an interesting adventure in the infant technology of robot ics. The fact that he is unhappy with his creation means that Ichnatowicz will go on to endow further sculptures with more "intelligent" features like recognition of patterns (people's faces perhaps) and true purposive behaviour, the life-blood of Cybernetics.

GLOSSARY

This is a list of the rather more technical terms used in the thesis. Some are newly introduced here, some are everyday words which, in context, have a precise meaning; others are words with no everyday menaing past the obvious one.

Art/Science

An art/science is a study wherein an attempt is made to achieve rigour in the analysis of an aspect of human behaviour, with a view to improving that behaviour, or, at the very least understanding it. In these terms physics, chemistry, geology are sciences; mathematics, music and strategic studies are arts; cybernetics, biochemistry, psychiatry are art/sciences.

Black Box

Imagine a machine, i.e. a chunk of matter which makes responses to stimuli applied to it (e.g. if you kick it, it screams; if you give it a pound, it laughs). You have no idea what goes on inside the machine i.e. there is no obvious visible mechanism connecting stimulus and response. This is a black box. The black box problem is one of observing inputs and outputs and hypothesising a mechanism to predict further outputs after specified inputs have been applied.

Decision-Table

A well-known device for simplifying computer programming where a large number of inter-related tests are made throughout the program. In fact it is isomorphic to a flow-chart but instead of concentrating on the flow of control which the flow-chart's name implies, it focuses on the decisions or tests controlling and directing the flow.

Discounting

In this context discounting means that action whereby the value of an amount of cash promised in the future is less than the face-value by an amount which reflects the loss (great or small) incurred by not having the cash now. If I was offered £1000 one year from now, I should discount it to (say) £900 because if I had £900 now, I could invest it at the current rate of interest and have £1000 in a year's time anyway. In the money markets, and in all financial institutions, there is not much cash about - most of the money is invested (or lent) elsewhere. Clearly in this situation, and in many others, discounting is an essential practice.

Discriminant Analysis

A widely used method in all science where objects are distinguished, one from another, by the application of a test based on a mathematical model of the objects in terms of several attributes or parameters. The values of the parameters are combined in a function (usually linear) and the value of the function compared with another value for a different object or against a fixed threshold. Medical diagnosis has been done like this - diseases are distinguished by the differences amongst a number of symptoms; heart-rate, temperature, blood-pressure etc.

Discrimination Net

This has similar properties and uses as discriminant analysis, but whereas that technique treats all attributes at the same level, the discrimination net uses each attribute in a separate test and arranges them sequentially, or hierarchically in groups. The net is similar to a flow-chart, except that it can have many exit-points - a flow-chart usually only has one.

Flow-chart

A graphical representation of a computer program (or indeed any behaviour involving individual activities and decision points). Each box in a flow-chart is either a command - "add four to X" or "read in a punched-card" - or a test such as "if X is equal to one, go to the box labelled 'A'". The boxes are joined by arrows indicating the flow of control (i.e. the box which the computer has to process next).

Fundamentalism

A philosophy of investment analysis which has, as its basis, the argument that the "fundamental" variables (earnings, dividends etc.) somehow determine what the share-price <u>should</u> be. The <u>actual</u> share-price is either above or below this and so the share is either a bad or a good buy. There are many, more or less enlightened, variations on this throughout a majority of the investment profession.

Game against Nature

A two-person game in which one "player" (Nature) is not trying to win the game, but is, essentially, unpredictable except over short periods or in very general terms. It is a paradigm of the problem of prediction.

Heuristic

A much-used word, no single definition of which ever seems to be satisfactory. Essentially a heuristic is used in a goal-seeking situation where the goal is hard (or impossible) to reach by a single transformation of state descriptors. The phrases "short-cut" and "rule-of-thumb" are often used but are not accurate enough definitions for further study. No satisfactory theory exists for producing good heuristics, or indeed for the reverse - evaluating heuristics against each other. However, "heuristic' programs exist in abundance and have been shown to work well.

Intrinsic Share Price

An invention by fundamentalists (see above) to justify their analysis of the workings and financial records of companies in order to discover whether a particular company's shares are a good investment or not. There is no direct evidence for the existence of such an intrinsic price, or, if it does exist, whether the majority of investors are influenced by this thinking.

Investment Theory

A theory, in general, of the Stock Market and its relationships with investors which enables one to invest more soundly than any other theory. It is essentially a pragmatic device, and its object is making money. One example is the Dow Theory of the New York Exchange.

Issuing House.

A financial firm specialising in selling of new stocks or shares to the general public, for the first time. It undertakes to sell all of new issue, and if not, the issuing house is the loser, having unsold shares on its hands (although it can insure itself against this.)

Logical Net

A network of simple logical units (AND, OR, NOT etc.) interconnected in such a way as to compute a particular logical function. In other words the inputs to the network are combined in such a way as to produce the desired output for different values of inputs. It relates very much to actual physical networks used in computers and logical machinery in general.

Market

An organisation in which buyers and sellers come together in large numbers to transact business. The key point is that they more or less agree as to the price of the commodity, because of rivalry amongst them. It is this which makes a market different from a simple supply and demand system.

Measurement of Life

Imagine a one-dimensional scale with non-living things towards one end and living things towards the opposite end. This is the measurement of life scale. Inanimate objects such as rocks, crystals, metals are at the "low" end; living things such as bacteria, plants etc. come about half-way, whereas the higher animals and especially man, come right at the "high" end. The interesting problem is where to place human artifacts, especially those exhibiting purposive behaviour.

Model

In Braithwaite's terms, a model is an <u>interpretation</u> of a theory. More broadly, it is a useful explanatory tool, whereas a theory is more descriptive. A model, whether it is in nuts and bolts, mathematics, or a computer program, attempts to make clear the interactions between various relevant entities under discussion and is able to reproduce behaviour of the particular modelled system. "Turning the handle" of a model to produce simulated behaviour is a key feature.

Perceptron

A term introduced by Rosenblatt to describe the machine he developed as a pattern-recogniser. It can be used to describe (loosely) any threshold device used to classify objects into sets, but is strictly the simple linear machine defined precisely in the body of the thesis.

Portfolio

Simply a collection of shares held by any one investor (private or institutional). Portfolios are selected first of all, and then reviewed regularly (or not as the case may be). The composition of a portfolio clearly reflects the investment policy of an investor. Selection of portfolios to provide a required overall rate of return is a separate problem from that of prediction of prices.

Protocol

Frequently used in psychology to denote a transcript of the exact words used by a subject in an experiment. A tape-recording would serve equally well as a record, but examination of a written protocol can yield many useful cross-links which are lost in the aural dimension.

Redundancy

Redundancy can mean many things, but here relates to the information content of data. The most accessible example is perhaps written natural language, such as English. If there are 150,000 words in such a language, and each word has, on average, 5 characters, then in theory, we only need 11 distinct types of character (say A to K) to produce the required number of words. The fact that we use 26, or more in some other languages, means that English is highly redundant. Put another way, we are wasting information, but, of course, gaining in ease of recognition - vitally important for efficient communication.

Screening

A screen, or sieve, is a device for separating large undesirable objects from small desirable ones, or vice versa. Screening in investment analysis is the the process of selecting the "probables" from a list (usually very long) of "possibles" in shares considered worthy of investment. Each share can be screened quickly by using a set of simple heuristics to evaluate their worth and so a large number of shares can be kept under continual review.

Simulation

The philosophy behind the idea of simulation is central to the Cybernetic approach. Most living systems are too complex to permit detailed, overall mathematical analysis. An alternative approach is to model as much of the system's behaviour and structure as is possible and to build in facilities for "running" the model. In this way the model will hopefully simulate the behaviour of the system in a way which is close enough to the "real" system to permit valid predictions to be made. The problem of validation i.e. how close the model is, is a tricky one here.

Technicalism

The prediction of share price movements by studying graphs of past share-prices and sundry other variables. The recognition of repeating patterns in past share-prices is the crucial point in Technicalism, but this has yet to be shown to be possible in the long run. The belief that history repeats itself is the philosophy behind it. This may be true or may not be true (" How does history repeat itself?" is a good question to ask) but the difficult question is "When will history start repeating itself?". Technicalism has been much maligned and entirely refuted by random walk theory, but has yet to be studied from a subjective standpoint.

Theory

See Model. A Theory is descriptive or even prescriptive in that it lists essential features of a system in a concise way, without attempting to teach or explain what the system is all about. It is more rigorous and thus usually subject to strict mathematical analysis.

Theory of Investment

See Investment Theory. In the terms of this thesis, a theory of investment is a theory which states how and why investors invest on the Stock Exchange. It is not so much concerned with making money, but with, in the end, explaining this facet of human behaviour, and hopefully linking it with other studies in psychology and related disciplines.

Variety

In the words of Ashby:" Variety, in relation to a set of distinguishable elements, will be used to mean either (i) the number of distinct elements, or (ii) the logarithm to the base two of the number." Since the logarithm function is a monotonically increasing function, either meaning is accurate in a qualitative context. In a quantitative context, the logarithm form is used, the unit being "bits", as in information theory.

White Box

See Black Box. A white box is simply a black box in which the mechanism can be observed directly. An example is a motor car, whose behaviour can be explained by examining its behaviour and its internal workings and evolving a model which connects the two.

INTRODUCTION

This thesis is not nearly so romantic as the work of Ichnatowicz, and nor is it as close to the fundamental problems of the measurement of life as the preface insists that Cybernetic writing should be. However, it does apply the methods of analysis, of theory, of formalisation, to a single field of human activity with the single philosophy that such an activity can be formalised. Most of the non-living systems we know have been, or can be in principle, formalised in the laws of the physical sciences. We know roughly how to calculate the properties (one might almost say the behaviour) of all known physical substances, ranging in size from the elusive (and illusory ?) neutrino to the incomprehensibly vast masses of stars and their birth-places. There are mysteries still unsolved, and there are gaps in our knowledge, but our scientific method and philosophy tell us that this must always be so. Many of the micro-properties of living systems are also subject to similar physical (i.e. non-living) laws. The sciences of biochemistry and neurophysiology are two prime examples of the use of physical laws in the living world. However, the macro-properties of living systems are the subject of the very young life-sciences; psychology, sociology, ecology, anthropology, and are understood much less. One could here draw an analogy between the physical and the life-sciences. Once the working of atoms was understood, the development of solid-state physics, surface chemistry and the like could follow, i.e. there was a development (an extrapolation in simple terms) from micro-properties to macro-properties. In the life-sciences the same development should take us from neurophysiology through psychology to sociology, again a transition, this time a two-staged one, from micro to macro. However, the analogy breaks down because in traversing the neurophysiology - psychology barrier, one is also transcending the non-living/living split. The change from purely physical laws to psychological laws may prove troublesome. Cybernetics, of course, hopes to show that it is really not troublesome at all; that a theory exists which denies the existence of the barrier at all.

My field of study in this thesis could properly be lodged in the study (in the practice) of economics, which should be considered a part of sociology, but which is an art/science **on its own. However, since I knew little about classical economic theory, I started off with absolutely no preconceptions from this side. Marx was a mystery, Keynes a conundrum, and to some extent this is still true. Neither of these gentlemen had much to say about the Stock Exchange however, and so I do not feel deprived.

In my early general reading I gained an impression of the subject of Investment Analysis as an art which was dragging itself (being dragged?) into the art/science class. When I ask the question why, I get no reply other than that there is a general lessening of the gap between Lord Snow's two cultures; science is no longer a dirty word.* Good scientific works in Investment Analysis stick out like sore thumbs; so much so that for a while I was wondering whether it was wise carrying on, since I was working on the premise that there must be a lot of scientific work in such an obvious field. Now, of course, I know that there has been a large amount of statistical studies done, giving rise to much discussion both in the investment profession and in academic circles. The amount of other scientific work is small, but significant and this will be discussed later. Yet there is one question which I feel has not been discussed properly (at least not in the investment literature). It is "Why does investment analysis exist at all?" This of course begs the question "Why do people invest on the Stock Exchange?" To answer this a brief resumé of the raison d'être and the workings of the Stock Exchange is necessary.

The Stock Exchange (and here I speak exclusively of the London Stock Exchange, although many features are similar throughout the world's exchanges) is a market, quite simply. The commodity which is bought and sold is, primarily, the share certificate which entitles the owner to partake of the profits of the company in whose name the share was issued. Ordinary shares are nothing more, although the owner of ordinary shares might possibly get something if the

* The anti-technology and not anti-science —

* The anti-pollution epidemic is anti-technology and not anti-science

** This is not to denigrate economics, for Cybernetics itself is an art/science i.e. a study which applies itself to the measurement-of-life problem. In these terms, mathematics is an art, physics is a science.

company goes into liquidation. Preference shares are really "a share in the company" but are more scarce. The main reason for holding shares, apart from the annual profits share-out (the dividend) is that the owner hopes he can sell his shares at the market for a higher price than he paid, and thus make a profit. Buyers and sellers of shares (usually stock-brokers) are brought together through the stock-jobbers who, in fact, comprise the stock-exchange. There are about 700 jobbers who are members of the stock-exchange and it is they who actually fix the share prices. (Rather like a book-maker on a race course). Thirty-six of the members make up the council which makes the rules of the exchange, admits new members, reviews offers for sale of new shares and carries out its administration. If a company wishes to raise new capital it can go to the council (probably through its bankers) and ask that a new share-issue be made. If all is well, an issuing house will buy up the share certificates (thus the company gets its money) and offer the shares for sale to the general public. The company then has no further interest in the shares (except that owners of certain shares may have voting rights at board meetings). The public investor thus buys shares either for the annual dividend, or for the possible future profit when he sells them, or a mixture of the two. In other words he is out to make money. Put this way, the gambling aspect of the Stock Exchange is highlighted. In fact, many writers have put forward this aspect as one of prime importance and it is a view with which I agree. To put it bluntly, the Stock Exchange is a "moral" casino, where fortunes are made and lost in the same way as on the roulette tables of Monte Carlo. Thus the investor is a gambler, and as we shall see later on, a gambler with about as much hope of making a profit as a roulette player. The bank always wins in roulette and in the same way it is the members of the Stock Exchange who make the biggest profits from the Investment business.

It is thus easy to see why so many investment systems exist and why there is so much conflicting advice about which shares to invest in and also why the investment profession is so large. Just as there are hundreds of roulette "systems", all equally futile, so there are hundreds of different investment methods, techniques and philosophies all pretty much the same as far as profit goes. The question of how good these systems are is a major one and will be discussed in what follows. It appears that certain people do make small, consistent profits from the Stock Exchange and it must be asked how is this done.

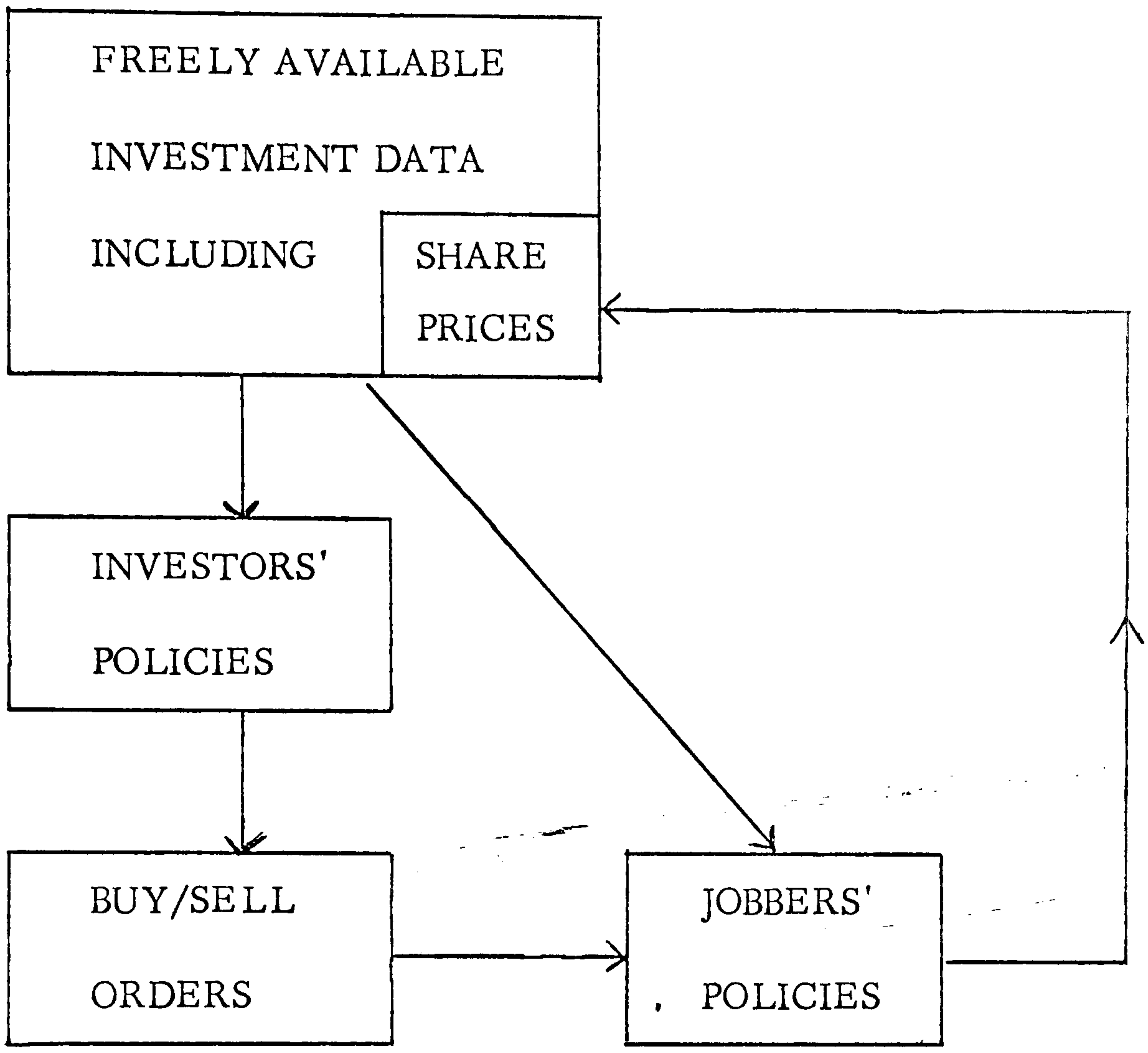
The central problem of Investment Analysis is one of share-price prediction. If one could predict which shares will go up in price and which will go down, intelligent buying and selling of these shares would lead to a consistent profit. It is a fact that share-prices fluctuate fairly rapidly (a week could produce a major change) and the reason for this fluctuation is our immediate problem. As already mentioned, it is the jobber who actually adjusts the share-prices and he is, in principle, a free agent. However, he is severely constrained in his actions by the profit motive of the company for which he works. Changing a share-price by too much could mean that there is a huge demand for shares he has not got or he has a large number of shares which he cannot sell. Either way he loses. Thus the jobber changes the price very carefully, all the while doing a complicated juggling act to maximise his profit. Unfortunately, outsiders can only guess at what goes on inside the jobber's head. There is no data from the London Stock Exchange which records individual price changes corresponding to a "buy" or "sell" transaction. If detailed knowledge of these changes, coupled with the reasons for making them were known, prediction would be much simplified and the jobber's profits would go down as a result of the improved predictions. Thus the jobbers keep themselves to themselves very much, in their own interest, and this avenue of attack is closed. The jobber makes changes in response to buy and sell orders from stock-brokers who handle shares for clients from the public outside. If these orders were known in detail we could make intelligent guesses as to the jobber's psychology and estimate what price changes he would make in response to changes in the buy and sell orders he receives. However, once again this information is neither recorded nor made available.

We are left then with the investors themselves, acting almost invariably through a middle-man (the stock-broker) who usually acts as advisor as well. Let us group all these together and call them investors. The buying and selling orders given by investors, of course, depend on the policies and the psychological make-up of each individual. Since there are hundreds of thousands of investors it is an impossible task to classify individual policies and hope to be accurate. However, we can say something about the existence of major psychological trends. It is a truism that if all investors think that a certain share will rise in value, eventually, but not straight away, the share price will rise, unless conditions change. This self-fulfilling property of the Stock Market is well known and is a crucial point in any study. It could thus be postulated that prices tend to follow major psychological trends in the investors' minds. One much quoted example is the death of President Kennedy which very quickly depressed the New York market. This happened merely because most investors thought that prices would fall because of general uncertainty in the future. In fact prices did fall, almost without exception, thus fulfilling the investors' prophecy.

It might now be worth a brief summary of the above line of thinking. If we are to predict share-prices, let us look at what directly affects them. The stock-jobber adjusts prices according to his own (secret) policies and psychology. As inputs the jobber receives buy and sell orders, given to him by, in effect, individual investors. These are issued by investors according to their individual investment policies and psychology. We cannot study the jobber directly and so we must study the individual investor, and investors en masse. There at least we have a chance of obtaining some data, since the inputs on which the vast majority of investors work are freely available to the general public. The jobbers' inputs and outputs are not available in anything like sufficient detail for a useful study. In a diagram of immediate effects * it looks like this:

(illustration overleaf)

* A useful explanatory device introduced by Ashby in "Introduction to Cybernetics" 1956.



Therefore, in conclusion, the most profitable line of attack seems to be to study the box labelled "Investors' Policies". A policy is the expression of a held opinion and the next chapter deals with opinion formation with the object of forming a useful theory to use in predicting share-prices.

OPINION FORMATION AND PROBABILITY

The subject matter covered by this short study is everyday speech, in particular statements of the form A was, is or will be B where A represents any noun, concrete or abstract, and B any predicate. Popper has said that "every statement has the character of a theory". Theory is really too strong a word to use in the present context because it connotes a rigorous application of scientific principles, so let us slightly change the wording to "every statement has the character of an opinion".

Opinions can be many things: value-judgements, forecasts, advice or informative statements. All these types of opinion are context-dependant since any statement, however simple, can be any or all of them. A statement like "water is wet", although meaningful in itself (as information) could be put to many uses:

All liquids are wet. Water is a liquid. Therefore water is wet. (Forecast)
Earth is dry. Water is wet. (Information)
Mercury is dry but water is wet. (Value-judgement)
You are wearing your best suit and the water is wet, so don't go rowing. (Advice).

The mere use of the word 'opinion' in connection with a statement indicates that there is dome doubt, however small, about its truth. Some statements are definitions, axioms of the physical and social systems we live in. "Elizabeth is Queen of England" is one example but even here some pre-restoration fanatics might contest it as merely an opinion. Other examples are "The sun rises in the east". "Iron is denser than aluminium".

These are facts by definition and are accepted by the vast majority as such. Some statements about the past also come into this category. e. g. "Christopher Wren designed St. Paul's Cathedral". (....was the designer of ...). However, the erosion of time can eat into any historical "fact as we well know.

"Shakespeare wrote Hamlet" and "Columbus discovered America" are two obvious examples that come readily to mind.

This sort of statement is not my concern here since the degree of doubt is either small or irrelevant or the statements are untestable by any known means.* My main concern is with statements about the present and the future both of which can have any degree of uncertainty between the limits of absolute truth or falsity. I want to examine statements like: "London is overcrowded"; "animals have no feeling"; "capatalism will inevitably break down"; "three thousand people will die on British roads next year".

Firstly, we need to separate clearly statements about the present and the future and I shall do this by calling them opinions and forecasts respectively. A forecast clearly has the nature of an opinion because its truth is necessarily in doubt but they are essentially different animals. Concentrating first upon forecasting, it is easy to see that these statements can be expressed in probabilistic terms. A statement such as "It will rain tomorrow" is clearly less likely than "It will rain sometime next year". The point to make straight away is that simple forecasting like this is a subjective process. No conscious techniques have been used to arrive at these statements (unless uttered by a professional weather-forecaster from the basis of his forecasting Method) and yet some unconscious counting of wet and dry days with similar characteristics; some weighing of different factors over a long period of time has gone on in the brain. In other words, there are no subjective opinions; there are only conscious and sub-conscious opinions. The difference in the method of computation is the distinguishing factor.

It is now easy to see how probability theory comes in;to each opinion which I have called a forecast we can attach a probability of it being true. This is no more, no less than a measure which enables us to compare the likeliehood of pairs of statements being true.*

* However much historical researchers may dig into the past their conclusions must contain a large proportion of blind faith.

The question of how probability is measured is central to my discussion and so a little time will be spent here. There are, broadly speaking, two conflicting yet complementary theories of probability. The classical objective theory*says that what we must do to measure probabilities is to count up the past occurrences of the event under consideration and divide by the total number of possibilities. This can then be used to predict future occurrences of the event. It assumes that elementary events are equally likely; an example is throwing a chosen number with a die, and is open to obvious criticism about verification of equally likely events. The answer to this criticism is, instead of talking about the probability of a particular occurrence, to talk about the probability of a set of occurrences of a similar nature. In other words probability is only defined in relation to some fixed set of events and the probability of an elementary event can vary with the set chosen. As a simple example, take the familiar, well-worn six-sided die. Take the set of possible throws i.e. one to six; the probability of throwing a six is one sixth. However, take the set of 10,000 tosses. There may well be only 160 sixes amongst these 10,000 tosses due to unknown (and irrelevant) properties of the die. The probability of throwing a six is then only one in six and a quarter.

Our job here is not to discuss the merits of the objective theory in itself (there are objections connected with verification) but to compare it with the subjective theory. Keynes defined probability as "the degree of rational belief". Popper amended this to read "the degree of rationality of a belief" and simply says that the probability of an event occurring is how likely it seems on the basis of the evidence available. A six-sided die, on first appearance, looks and feels symmetrical and so each number on the six faces is equally likely to be on the one thrown. This sounds vague and qualitative and, put like this, it is; but taken a little further it provides the necessary filling to plug a hole in the objective theory.

The objective theory relies heavily on the frequency of occurrence of an event or of a set of events.

* or, rather the union of all such theories.

If there is, or will be, only one event the theory says nothing, or rather gives trivial results if carried through rigorously. Take an event which has never happened before; there is no set one can define which will include it so one is forced to conclude that its probability is zero, or even worse, unknown. In a sense, of course, every event is unique, but some can be classified into sets, some not. The probability of snow on Christmas Day 1975 can be estimated on the objective theory; the probability of men landing on Mars on Christmas Day 1975 cannot.

If we are to assign probabilities at all to unique events, it must be on something like the subjective theory. Here I would like to draw what I think is a useful and very close analogy between the objective and subjective theories of probability on the one hand, and the laws and models of the scientific method on the other. The objective theory says, in effect, that we must find laws governing the occurrences of events by considering suitable sets of these, and counting each individual occurrence in an experimental sort of situation. The subjective theory is trying to say that we must build a model of the situation and estimate probabilities by "running the model" i.e. causing the model to produce the set of events considered in the objective theory. In simpler terms probability depends upon past occurrences; where none exist, build a model to provide them.

It is now plain what Popper meant by his phrase about statements having the character of theories. Forecasts, in particular, are made on the basis of internally stored information and almost certainly of models of unique events. An opinion like "It will rain tomorrow" is the verbalisation of one of numerous laws formulated in the light of experience. An opinion like "Britain will enter the Common Market" * must be based on an internal model. Possibly the same mechanism in the brain is involved in the two processes, although they are not really compatible. Laws come under various disguises when used by human beings. Prejudices, proverbs, old-wives' tales are just some of them that we use in every day life.

* Now a fact by definition! (March 1973)

Model-building is an exercise everyone indulges in in order to make predictions about unique events, although it is doubtful whether this is an on-going process. It usually requires conscious effort and does not come easily to less intelligent people. The mere remembering of a proverb or "what happened last time?" requires less mental effort (cognitive strain in the words of Bruner et al)* and is the path usually taken.

This brings us back to investment and to the investment policies of individuals. Postulating that people are no different in forming investment policies than they are in forming opinions and making forecasts about a host of things in their normal lives, we could further postulate two distinct methods of making investment predictions. One would be to use past occurrences of share-prices and to formulate laws governing their behaviour. Using these laws one should be able to predict future price behaviour. The other would be to build a model of the situation surrounding the share-price and to "run the model", producing predictions from its outputs. These two postulates in fact broadly describe the twin methods of investment analysis, always in conflict and almost never utilised simultaneously in the same investment policy. Corresponding to using laws to make predictions we have <u>technicalism</u>, a method much abused and criticised but still holding strong for reasons I shall not discuss here. Technicalism means that one only considers the technical variables in making predictions i.e. the history of the share-price and the corresponding <u>volume</u>** measurements. The laws are formulated in terms of the patterns which allegedly exist in the graph of the share-price history. One merely looks at the graph, chooses one of the numerous patterns which best fits it and blindly makes predictions accordingly. Perhaps this is not quite fair on an undoubtedly sophisticated method, but it is a technique well suited to modern computer technology. The best exposition of technicalism so far encountered is by Cohen and Zinbarg.*** They list and discuss in great detail some fifteen types of pattern and their corresponding actions.

* Bruner, Goodnow, Austin "A study in thinking."

** The total number of shares traded in any one day.

*** Cohen and Zinbarg "Investment Analysis and Portfolio Management" 1967

Corresponding to model-building in forecasting we have fundamentalism. This, roughly speaking, aims to cover everything that the technicalists leave out, i.e. all published information about the company whose shares are being traded, together with more general information concerning the state of the economy and similar items of immediate interest. It is a fact that, as we shall see later, the fundamentalists (and they form the bulk of investment analysts) scarcely ever make their models explicit. They keep their models and the corresponding investment policies in their heads; possibly this is to avoid leakage of information but is more probably because this is the way it was always done.

It is not our place here to discuss the two techniques of technicalism and fundamentalism further; even less to discuss their relative merits, except to say that the majority of the statistical studies have been directed against technicalism and, basically speaking, show that the method is practically worthless; one cannot predict share-prices on the basis of past prices alone, with or without volume figures.

However, with the thought that most investors are, more or less, fundamentalists and use internal (implicit) models I shall present a theory of investment which, if it represents a majority of investors (or in some sense represents an average over an ensemble of investors) should yield worthwhile predictions. Before this comes a discussion of different theories of investment and investment theories. The distinction between the two, and an important one from the point of view of Cybernetics, will soon become clear.

THEORIES OF INVESTMENT AND INVESTMENT THEORIES

The discussion in the previous chapter also yields a broader distinction which it is useful to consider when attacking the investment problem. Throughout the investment literature, whether old or new, the professional practice of investment is described as an art. This is such a common statement that it has become a cliché. (It is also true of much of the management techniques in industry). However, it is interesting to consider why the statement is made, especially in the face of seemingly sophisticated analytical techniques in current use in professional circles. The twin approaches of technicalism and fundamentalism described in the last chapter can indeed be paralleled with the discovery of laws and the building of models respectively. These are terms extracted from scientific activities and it is true that many "scientific" techniques are utilised in investment analysis. However, they are only techniques within an overall approach which can only deserve the term "art". Both technicalism and fundamentalism are arts because, in a sense, they are inward-looking - they are contained within themselves. The key point about a scientific approach, as opposed to an artistic one, is that the scientific approach demands a dedicated attempt at rigorousness. Whether complete rigour is achieved is, to some extent, irrelevant, but at the very least it should be attempted and the shortfall discussed. This point has been made many times, but it was made very forcefully by Stafford Beer in trying to persuade managers to accept scientific techniques into their everyday activities.* It is often thought that the scientific approach is logical, or rational, but logical thought is not the exclusive prerogative of the scientist. In fact everyone uses logical thought processes - all sane people are rational, although not all the time. Thus the artistic approach certainly lacks rigour but what it loses in rigour it gains in the lack of cognitive strain. In other words human-beings are very happy to exist in a very fuzzy, unprecise environment where they can place their own, subjective interpretation on themselves and their role within the environment.

* Stafford Beer "Cybernetics and Management" 1956.

Science, insisting on rigourousness, provides no such mental cover - there is nowhere to hide. The objectivity which the artistic approach lacks is also lacking from investment analysis. In general, the analysts make no statements about the investment process - all they produce are investment theories, i.e. how to invest well. A scientific approach to investment analysis must say something about theories of investment; there must be a discussion about how and why people invest.

However, if one gives the professional analysts the benefit of the doubt, both technicalism and fundamentalism have, submerged underneath the vast mass of data, the basic elements of a theory of investment. All too often, many analysts are unaware of these postulates and work on without really understanding the justifications for their approach. The technicalists' theory can be thought of as typified by Cohen and Zinbarg's comments:

"It is the influence of ....rumours, facts, and statistics that causes men to buy and sell their stocks. It is their actions that build the familiar chart patterns. You are not interested in why they are doing what they are doing. So far as your trading is concerned you are interested only in the results of their actions." They then continue with:

"The habits and evaluative methods of people are deeply ingrained. The same kinds of events produce the same kinds of emotional responses, and hence the same kinds of market action."

This well illustrates the sort of confusion that can arise when rigorousness is not insisted upon. If the theory states that certain emotional responses produce certain chart patterns, then it is vitally important that the reasons for these particular responses are studied; not ignored as Cohen and Zinbarg suggest. It might well be that the emotional response - chart pattern theory has a lot to commend it. The technicalists, however, have not studied the psychology of investors in any sort of rigourous detail. Their theory is thus untested (it may also be untestable - a very important point in formulating any theory).* It might also be said that the recognition of "patterns" is done entirely subjectively. A study of human pattern recognition would thus yield valuable results for the technicalists.

The fundamentalists, overall, have an equally short-sighted view of investment. We think it can be said that they make no attempt at a theory of investment but the basic postulates form a theory of the stock-market. Their view is that the share-price reflects the average share-holders opinion of the company, especially of its future earnings. They also say that these opinions are influenced by the past performance of that company. This is fine as far as it goes, except that they go on to say that the share-price of each company has an <u>intrinsic</u> value; a value dependent only on the past, present and future prospects of the company in terms of the so-called fundamental variables. There seems to be no general agreement about which variables are necessary, but plenty of advice (much of it conflicting) about important variables. The fundamentalist theory falls down in mentioning an intrinsic share-price, as they rightly say, is dependent on the aggregate of investors' transactions, but their professional advice is based on the relationship of this price to the intrinsic price, reflecting the "real" value of the company.** The "intrinsic" value theory which pervades much economic writing has long been refuted in academic circles, but still persists in professional investment.*** It is again an example of an untestable theory - the intrinsic price is unmeasureable directly and its existence could only be inferred from the activities of investors. There is no evidence which supports the theory from this source.

The very short descriptions of the basic postulates of technicalism and fundamentalism given above are, of course, gross simplifications of the real picture which is immeasurably more complex and hard to describe. It could be said that

---

* See P. Freeman "Some Aspects of Technical Analysis" The Investment Analysis October, 1967.

The postulate that there are patterns has, of course, been subjected to rigorous statistical analysis.

** Fama "Random Walks in Stock Market Prices" from the Investment Analysis December 1965.

*** See Granger and Morgenstern "The Predictability of Stock Market Prices". 1970.

a hundred different investors, taken at random, have a hundred different investment policies. Some investors, and this is especially true of the big institutional investment departments, are not as short-sighted as to stick steadfastly to one method or another - the sums of money involved are too large for that. However, we believe that one general comment is entirely valid when applied to the vast majority of professional investors. It is that there does not seem to be enough emphasis on the psychology of the people comprising the social group we call investors. It seems that most analysts see investment as a game against nature. In other words they accept the stock-market and its behaviour as something as remote and monolithic as the motion of the seas or the unpredictability of earthquakes. They are loath to accept the plain fact that the stock-market is a totally human edifice and is in principle, understandable in terms of human behaviour. Thus we claim that there is much too much emphasis on investment theories rather than the ultimately more rewarding approach based on theories of investment.

If we could split the investment literature into two - those that discuss investment theories and those that discuss theories of investment - the bias is heavily in favour of the former. We shall have an attempt to discuss the literature covering the latter, although our discussion is certainly not comprehensive. It is certain, however, that there is only one work in the investment field which could be said to contain a purely cybernetic approach. It is by Geoffrey Clarkson in "Portfolio selection: A Simulation of Trust Investment" 1962. Since the author also includes a discussion of theories of investment, the work is clearly of great relevance to this thesis. The reason why his work could be called cybernetic is, we think, mainly because he has approached an essentially human problem (portfolio selection) in as direct and rigorous a manner as possible. His model, being a simulation, does not rely on any theory of investment as its basis, and so it is free from any inaccuracies due to wrong assumptions in this line. Clarkson does infer the basis of a theory from his model, but this is a much safer process than the usual method of hypothesing directly from observations of behaviour or from data inputs and outputs. His work will be described in some detail, with the object of extracting from it a basis for presenting our own theory of investment in the next chapter.

First of all, Clarkson clears the ground and states exactly where his aims lie. The title of the book alludes to this but the following makes it clear: "We shall not be directly concerned with what an optimal portfolio or a "good" security is or ought to be. Rather, we shall address ourselves to quite a different question - namely, how do investors actually select portfolios from the set of available securities. (The work was done in the U.S.A., hence the use of the word "securities". "Shares" or "equities" are more common in England.) A portfolio is nothing more than a collection of shares held by one investor. It is usual for an individual investor to have his investments handled by a stock-broker and so we speak of his portfolio of shares being managed by the stock-broker. Clarkson is concerned with not only the managing of portfolios (the continual review of the performance of the shares contained in it) but also with the initial selection of shares which the stock-broker thinks should comprise the portfolio. In fact, the field is limited further since the number of different types of investor is so large as to be unmanageable. Granger and Morgenstern make the point that the distribution of types of investors is very uneven, and there is "no way of even remotely describing it." Thus Clarkson concentrates his study to the trust investment officer. This is because trust investment (in the U.S.A.) is constrained legally as to the types of shares which can be bought and the financial liability of the trust managers (usually a bank). These restrictions make the study simpler, without losing much validity when considering the wider implications.

Although the model simulates one particular investor, Clarkson included in his model as much experience in similar fields as possible. Thus instead of putting forward a narrow theory in one field of study he made an attempt at rigorousness by including his study as a particular case of general human decision-making. The psychological postulates were taken from contemporary research in decision-making, especially from the work of Newell, Shaw and Simon.*

On the subject of simulation, Clarkson compares the technique with techniques from classical econometrics used in the same problem area. The difference, he says, is not so much in the form of any model but in the manipulation of the model. Classical econometrics tends to use a "one-period change" method-in other words values of economic variables have an associated time subscript which is increased in steps to "run" the model. Simulation, on the other hand, uses decision techniques and threshold tests to "run" the model. Certainly this means that simulation models are somewhat closer to human behaviour; this, after all, is the aim of simulation.

The model itself is built upon three postulates of decision-making taken from Newell, Shaw and Simon. They are that there exist:

A. A <u>Memory</u> which contains information relating to individual companies, industries and the economy in general.

B. Search and Selection processes which analyse the information stored in the memory (much as a professional analyst would do)

C. A set of rules whereby the investor makes the decisions which guide the processes in B.

Concerning the last of these postulates, which is certainly the most important one as far as constructing an operational model goes, Clarkson makes the point that the rules must be defined unambiguously if the model is to run on a computer. The computer provides a very versatile medium for this sort of research but it has one drawback that Clarkson pinpoints. Any system, such as a computer,

...ber of such papers, including a shortened version of Clarkson's own book which has been constructed precisely and logically can only exhibit precise behaviour. Unfortunately, from the point of view of simulation of human behaviour, human beings do not, in general, exhibit such precise behaviour. Human precision only comes about through imprecise thinking processes. This is a criticism of all computer simulations since the days of Clarkson and the answer is only just being found. The work of L. A. Zadeh shows how a rigorous mapping from a precise system (the computer) to a precise system with an imprecise interpretation can help simulation research.* If the insistence on simulation of processes as well as the outcomes of decisions is dropped then, of course, such imprecision is not necessary; the outcome of a yes/no decision is precise whether produced by precise or "fuzzy" means.

The building of the model then becomes a problem of discovering which data is used by the trust investment officer (postulate A); which search procedures he uses (postulate B); and what rules he uses to choose the final portfolio of shares (postulate C). For the collection of his data, Clarkson mainly used the "protocol" technique, whereby the investment officer was asked to "think aloud" while making decisions and the form of the decisions was inferred from a record of his verbalisations. It is interesting to remark here that this whole study of Clarkson's could only have been carried out in the light of bona fide academic research. As we have already seen, if this process of discovering how investors actually invest could be carried out for the major investment groups (whatever they are), we would then have a means to predict future transactions and could, of course, anticipate them. For these very reasons, investors tend to keep their methods to themselves; this is another point in favour of Clarkson's work.

The model itself consists of three parts:

A. The analysis and selection of shares suitable for current investment.

B. The formation of an investment policy from information given by the client.

* Zadeh's concept of fuzziness, especially fuzzy algorithms.

C. The selection of a portfolio.

The data is basically in the form of a list of companies (the "B" list) which the investment officer considers as being worthy of investment at some time. At any particular time a subset of these is formed (the "A" list) which contains companies worthy of current investment. This is the process of screening very common in the investment world. The attributes listed for each company are as follows:

Sales
Earnings
Cash flow per share
Profit margin
Working capital
Price earnings ratio
Dividend payout ratio
Dividends per share
Dividend yield
Share price

From these basic attributes additional lists are formed of:

the average of the last ten yearly values
the fractional yearly change
the average yearly change over the last ten years

In addition to basic company data, the model accepts forecast information on economy and industry attributes (on a "below", "above", "equal to" scale) and forecasts of the following five company attributes:

Sales
Earnings per share
Cash flow per share

Profit Margin

Dividend per share

It could be concluded from Clarkson's description of the data input, that the investment officer considers just about every published piece of information which he can easily keep under constant review.

The quality of the data (i.e. how good each company is) is evaluated by forming the "Relative Performance List" which contains the following:*

1. The average value of each attribute over the last ten years and whether this is above, below or equal to the average of this for all companies in the same industry.

2. The average yearly change over the last ten years and an associated value on the three-point scale as in 1.

3. The expected yearly change over the coming year and the same three-point value as in 1.

4. The average expected yearly change over the coming three to five years and the same three-point value as in 1.

In addition to this, the "Relative Value List" contains similar information as the Performance list except that it relates exclusively to the price-earnings ratio. This concludes the "analysis" part of section A of the model. The selection of the "A" list from the "B" list is made by scanning the two lists described above and using a set of simple criteria to determine whether a particular company should be included for further examination of investment prospects. This is mainly done using the Relative Value list i.e. companies with above average P/E ratios are considered good investment prospects.**

The other part of the model which is interesting to the present discussion is C,

* It is interesting that Clarkson's "equal to" takes into account the inherent inaccuracies in investment data. It is in fact evaluated on a ± 5/% basis - much more of a "don't know".

** Clarkson was not, of course, concerned with whether they are good prospects - hence the presence of the dubious P/E ratio.

the selection of the portfolio. It is this part which yields the basis of the theory of investment which Clarkson inferred from his model. The tests carried out to select a number of companies consist of a number of binary decisions on the performance and forecasts of the most important attributes. Furthermore, the decisions are ordered so that some are clearly more important than others. Clarkson describes this as a discrimination net, but equivalently it could be described as a flow-chart or decision-table, terms taken from computer terminology. A company is selected if the following set of conditions holds:

1. Average growth in price over last ten years 20%.
2. Average growth in earnings per share over last ten years is not "below".
3. Forecasted growth in earnings per share (1 year) is not "below".
4. Increase in P/E ratio is not "above".
5. Forecasted dividend > 0.

There are other sets of conditions which result in acceptance and other sets which result in rejection of the company.

The results of "running" the model are exceptionally good. Clarkson tested his model against various random and naive selection procedures and the model was better (though not perfect) than any of them. In fact comparisons of several portfolios produced by the investment officer and by the model reveal little difference. What is more, comparisons of the investment officer's "protocols" and a trace of the model's workings reveal a distinct similarity- Clarkson's simulation is excellent in almost every respect.

In the penultimate chapter Clarkson discusses theories of investment and what contribution his model makes. He discusses the Shackle-Angell theory of investment that investors act as if they minimised the "potential surprise" that an investor expects if a particular outcome is actually realised. The theory (as all optimising theories do) demands that the investor acts rationally. Ellinger claims that most analysts deny that investors act rationally - in fact investment procedures are "about as rational as marriage".* Clarkson makes the point that such theories demand the formation of a function which connects economic variables with different weightings. In his model there is no evidence which supports this view. The Shackle-Angell theory also demands that the investor look at all the investment possibilities and to choose the best. Clarkson's model indicates that, in fact, investors process lists one item at a time, accepting the first one which passes a series of tests (the discrimination net). Clearly, the order of such a list is important. (the model - and in fact the investment officer - produces different portfolios for different orderings of the "A" list). There is considerable support for this view from other research in human decision-making, as Clarkson says.

We now come to an important point that a pure theory of investment, such as Clarkson's model suggests, must not only produce good decisions but it must produce them in a way which parallels the human thought processes. An optimising theory can produce the former but not the latter (at least it cannot produce the explicit thought processes, as verbalised in a protocol). However, as Clarkson says, precisely the same outputs (in this case the selected portfolio) could be produced by a number of different techniques. In an earlier paper, Clarkson and Meltzer used a process that assigned weights to individual attributes and combined these in a linear form to produce a number which was used to distinguish between different companies.** The results of this technique were similar to the model of the investment officer but provide no insight into the workings of his mind. We have here a dichotomy of approach which could be described as the black box technique versus the white box technique. The black box approach concerns itself with correctly identifying the inputs and outputs associated with the system under consideration, and then inventing some

---

* A. G. Ellinger "Security Price Behaviour" in Investment Analysis. May 1966.

** Clarkson and Meltzer "Portfolio Selection: A
Journal of Finance, December 1960

mechanism which will connect the two such that, when a suitable representation is chosen, a particular combination of inputs produces the correct outputs. The box is black because, for one reason or another, its internal workings are unavailable for close study. In other words, the actual mechanism which connects the inputs to the outputs is unknown (or unknowable). In contrast the white box approach attempts to represent this actual mechanism, or at least as much of it as possible, in an explicit model. The connection between inputs and outputs is thus demonstrated via the model instead of merely hypothesised as in the black box technique. Clarkson's model is clearly of the white box type and the theory of investment inferred from the model describes inputs, outputs and the intermediary mechanism (the decision rules and evaluation techniques). However, a theory of investment of the black box type is no less a theory than one which postulates a connecting mechanism; it is, perhaps, a weaker theory in that it is less informative than a white box theory, but conversely it is stronger in that it is more easily testable. In general, we might say that white box theories are explanatory, whereas black box models are predictive. Thus we come back once more to our suggested split between theories of investment and investment theories. The former aim at explaining how people invest, the latter at predicting share-price behaviour through aggregate investor activity.

In the next chapter we propose a theory which attempts to do both, but before this there is a discussion of an example of a black box, predictive theory from the literature. The work of Weaver and Hall is probably the best to discuss here in that it highlights the opposite approach to Clarkson's and, although slightly inconsistent in the justification of the theory, the predictive aspects are highly impressive.*

As perhaps might be expected, Weaver and Hall carried out their work in the course of their professional duties with a large stock-broking firm in London - hence the emphasis on prediction as the main aim. The authors start by out-

* Weaver and Hall, "Evaluation of Ordinary Shares Using a Computer" Journal of the Institute of Actuaries 1967.

lining their view of the investment problem. They split this into two parts: firstly, how the investor makes an estimate of a company's likely future growth and secondly, how the market will react if the forecast actually happens. Unfortunately they then mention that the second part of this procedure describes "how much of the expected growth is already discounted in the price". As Granger and Morgenstern point out, the much quoted ability of the stock-market as a whole to discount future events (in fact to discount anything at all) has no possible justification since nothing like a suitable theory exists to explain the phenomenon. Individuals can, and do, discount future events in interest rates, but this is another matter. However, Weaver and Hall then rightly say that the subjective process of making forecasts is the more difficult part of the investment problem to deal with. Instead they concentrate on the second part which involves a large amount of pure data-processing, something the computer can do much better than the human-being. It is at this point that the authors' theory of investment comes in. Instead of attempting to predict future share-prices as is usual, the predicted variable is the dividend yield. Basically the dividend yield is simply the ratio of dividend payout per share and the current share-price. It is thus an inverse function of the share-price. The formation of a forecast dividend yield involves, say the authors, the following comparisons:

A. The dividend cover against the average for all shares.

B. The future prospects, both short and long-term against the average.

C. The company's record against that for the industry.

These are the postulates for the theory and clearly involve a set of necessary data inputs. The authors say little about any mechanism in the investor's behaviour to connect these inputs to the forecasted yield (remember this is an example of a black box theory) and, in fact, say little about which "investor" or set of investors the theory is meant to relate to. Still, the comparisons listed above are put forward as the processes they "believe the investor carries out in his mind when choosing an appropriate yield for a share". The only hint at providing insight into the mechanism of the thought processes comes when the authors mention the subjective assignment of weights to the input variables in some sort of functional relationship. This is exactly the sort of the thing that Clarkson's model provides contrary evidence to and perhaps reflects a statistical background and training.

Having proposed a theory, Weaver and Hall describe a model which interprets the theory. The black box mechanism they chose was multiple regression analysis, a thoroughly understood technique from statistics. As mentioned already, the predicted variable is dividend yield, the predictor variables are:

1. dividend payout ratio.
2. forecast earnings growth rate - short-term.
3. forecast dividend growth rate - long-term.
4. historical earnings variability.
5. historical earnings growth rate.

The precise definition of these variables is not important, but the philosophy behind the definitions is. The first three variables are all based on forecasts of dividends and earnings. Thus they represent the subjective element in the investment problem referred to in the first phase of investment as Weaver and Hall see it. The other two variables, the "historical" variables, also involve a subjective element. The earnings variability attempts to measure the consistency of the company over the previous five years. Its value is the standard deviation of the earnings about the trend line defined by the historical earnings growth rate which is assumed, as implied, to be a straight line.

Thus all the variables have a subjective element in their actual formation. Not only are the actual values involved subjective, but the choice of the function to combine these subjective forecasts is also based on subjective assumptions.

This contrasts with the variables and pre-processing used in Clarkson's model. Weaver and Hall's choice of variables could be described as a subset of the Clarkson inputs but the way they are handled is slightly different, in the emphasis on the subjective aspect. Other factors were considered by Weaver and Hall, but they were thought to be of lesser importance than the five listed above. Among those considered were: the long-term prospects for the industry; the relationship of the current price to the asset value and the all-time high share-price; management; prospects of take-over; political considerations, such as possible nationalisation. One reason given for disregarding these is that they are difficult to quantify with any degree of certainty. The emphasis on comparison in Clarkson's model and also the use of the discrimination net is one way in which variables such as these could be handled, since no quantification is necessary.

We shall not dwell on a description of the application and analysis of the regression techniques except to say that any discussion on the significance of the predictor variables is admitted by the authors to be difficult. (No attempt is made to attach any meaning to the magnitude or sign of the regression co-efficients beyond common-sense - this is only correct in a method which is essentially insensitive to such subjective considerations). It is difficult to reconcile such a discussion of significance which relates to the arbitrary choice of the black box mechanism, with the choice of input variables as proposed by the theory. However, as the authors say, they are not testing the choice of the input variables using the regression analysis (a doubtful process in the light of the above discussion), but are content to accept the regression coefficients and use them in the subsequent predictions.

The actual predictions from the model take the form of dividend yields, or, in conjunction with the forecasted dividend, a forecast share-price. However, the authors mention later that the object is not to produce actual numerical estimates which, in any case, are subject to "fairly wide limits of error," but to rank the shares in order of attractiveness. The numerical estimates form the basis of the ranking and the whole ordered list is then split into five groups according to the shares' "relative cheapness" (apparently the forecast percentage gain in dividend yield). Portfolios were selected using these categorisations and compared with each other and with the Financial Times share index (representing the average performance of the stock-market). Gains of up to fifteen percent are reported for the best portfolios over an "aggregate" portfolio formed from all the shares ranked. Comparison of the computer method and other methods reveals that the computer model fares less well than the "in-house" comprehensive analysis, but better than other simpler methods (similar to Clarkson's discussion of "naive" methods). Since the computer predictions take much less time than the standard method, the company adopted the system and use it as the basis for their day-to-day transactions - a measure of the company's faith in the methodology. This faith seems to be over-optimistic in the light of other work done using similar statistical methods. As Granger says, "this type of analysis generally gives misleading results...... spurious correlations and regression coefficients are highly prone to occur".* However, one cannot disagree with the excellent and consistent performance of the portfolios over a period of three years. It is not our place to discuss why such practitioners as Weaver and Hall disagree with a mass of contrary evidence from statistical studies - that has been done elsewhere extremely comprehensively. The main reason for discussing their work is to illustrate an example of a theory of investment used as the basis for predicting share-prices. The final comment provides an interesting tie-up between a suggestion by Weaver and Hall as to an alternative method for ranking shares, and the method used in early work by Clarkson and Meltzer to select portfolios. Weaver and Hall (who seem not to have read Clarkson's work) suggest a form of discriminant analysis which places shares either in the "dear" or in the "cheap" category - the discriminant is simply a linear function of various investment variables. Clarkson & Meltzer reported good results in the simulation of portfolio selection using this method but Weaver and Hall propose it as a predictive technique. The next chapter takes up this suggestion and proposes a similar technique as a black box theory useful, perhaps, as a predictive tool.

*C.W.J. Granger "Analysis of Stock Market Price Data" Journal of Institute of Mathematics and its Applications" August 1972.

A THEORY OF INVESTMENT FOR PREDICTING SHARE-PRICES

We have seen that if there is a major policy among investors as a whole then their policy will be self-fulfilling, since there are sufficient numbers of investors with the same views to affect the market in a positive way. Thus if we were able to anticipate major psychological trends in the mass of investors, the movement of those particular shares which they consider would be predictable, more or less according to the size of the agreement. The factors to use in making the prediction would be simply the factors which influenced the psychological trend in the first place. The problems are clear; recognising a trend before it has got under way and estimating the size of it. The technical analysts reckon that such trends appear, somehow, in the patterns that the share-price movements make. However, on pure information-content grounds it is an extremely unlikely effect. It is true that the share-price graph is a distillation of the workings of investors minds (since it is they who cause it), but the variety* in a graph like this is is much less than the variety in the factors which fundamentalist investors study. Technical analysis would only be possible if most investors were technicalists. Most investors are not, so the approach falls down. As previously mentioned, the statistical evidence against technical analysis is quite overwhelming.

An example of self-fulfilment of investment ideas which readily springs to mind because of its wide publicity, is the Poseidon saga. In essence, the details are as follows. Poseidon was a small mining company in Australia and its shares stood at something like fifteen or twenty pence on the London Stock Market. The company announced, out of the blue, that they had discovered nickel on one of the sites, an unusually rich finding as well. Purely on this report (with mineralogists' preliminary reports for authenticity) the shares began to shoot up. As more and more glowing reports filtered through from the mine itself the momentum of the upward surge increased and finally the share-price rocketed. Within days

* Variety in the sense used by Ashby.

the price went to approximately eighty pounds and finally after reaching a peak of one hundred and twenty pounds* , collapsed exhausted, back to a few pounds. This, of course, is an extreme example, but highlights the fact that the mere prediction of vast company profits in the future can influence investors to buy. The extent of the acceptance of these predictions was converted into an upward surge in the price and this in turn attracted more investors. This sort of positive feedback is usually heavily damped in normal share-trading by a jumble of conflicting opinions. With Poseidon there was only one opinion and that was that the company would make huge profits and/or the share price would go up. The reason for the equally rapid fall in price is largely the reverse of this; everyone changed their minds about both the company and the shares. Ironically, the early mineralogists' reports were shown later to be far too optimistic, but by then a number of people had made a lot of money out of Poseidon, and the fate of the company could not have been further from their minds.

However, major psychological trends such as this are few and far between. The everyday trends, if they exist at all, are much more slight and, consequently, harder to detect. Small investors can only be influential in large numbers and then only if they are in general agreement as to investment policy. Because of their large numbers, however, their views tend to cancel each other out. Large investors, and this means the big financial institutions (such as insurance houses, banks, hire-purchase houses) and the large companies (petrol, tobacco, chemicals) can clearly have a positive effect in much smaller numbers since they have much larger sums of money available. (A favourite way of taking over a small company is to buy a large proportion of its shares, more or less quickly and then dictate its future actions and policies to it). So here we have a likely way of predicting share-prices; if we could study the investment procedures of the large institutions we could predict their next move and anticipate them. However, once again this is wishful thinking because (understandably) they operate behind closed doors and, as mentioned before, individual market transactions are not recorded.

* or was it £170 ?

The direct approach of anticipating major psychological trends seems to be doomed to failure due to lack of available information. We are left with having to theorise about such trends and use the scientific method to test our theories out. Professional investors tend to develop a theory, or have it thrust upon them, and then stick to it. They very rarely make their theory explicit and very rarely use something like the scientific method to test the theory and to modify it accordingly. A quote from Cohen and Zinbarg bears this out:

"The reader will probably acknowledge readily that most individual investors have little notion of what rate of return they have earned on their investment ....But most readers will probably find it difficult to believe that professional investors have almost as little knowledge as the proverbial man-in-the-street. Yet our experience suggests that this is truly the case." The authors, being themselves technicalists, seem not to follow their own advice, since they virtually ignore the many statistical studies which disclaim their methodology.*

Any theory of investment must, I think, incorporate readily available data, and say something about how people think in general, not only about investment. Thus the average investor will look at easily available data and use it in a fairly simple way, in a fairly simple implicit theory. Professional investors may use more sophisticated techniques, but cannot, if their theory is implicit, use complex computational methods or explicit decision rules. To this end, our theory, presented here, differs in one crucial aspect from Clarkson's discussed in the last chapter. Clarkson discovered the investment officer's implicit rules and made them explicit in the form of his "discrimination net", an ordered sequence of binary tests. The "average" investor, however does not, it is hypothesised, use such large quantities of data or such comprehensive decision-rules as a professional investor. Thus our theory does not contain explicit decision procedures to evaluate shares, except at the level of pre-processing the data. (This is much as in Clarkson's model).

*In Britain it is no better. That sensible magazine "Money Which" once said: "Assuming that skill in investment management consists in choosing shares which will do better than the average, it seems fair to conclude that Unit-Trust managers, collectively, do not show any noticeable investment skill. (Sept. 1970).

Furthermore, it is intended also to use the theory of investment as an investment theory, in the sense of the meanings of these two phrases as outlined in the last chapter. Clarkson's model was never intended to be used to attempt to predict share-prices - it was wholly a simulation model.* This is as it should be since it was a simulation of one man, and as we have seen, one man cannot hope to influence the stock-market on his own. However, if a theory of investment attempts to describe the policies of the average investor, then it is valid to test whether the theory is a good one or not by testing its predictions of actual share-price behaviour. It could perhaps be said that this is the only way to test such a theory, since there is no such person as "the average investor"; especially as the average is meant to be taken over all types of investor including stock-brokers and institutions.**

The form of the theory will be presented later, but a few comments are necessary before coming to a statement of the theory. As in Clarkson's model, the basis of the theory is comparison, a procedure which numerous psychological experiments have shown to be an easy one to do. It involves far less cognitive strain than an absolute determination. (Given a ball, it is hard to say how heavy it is, but given two balls, it is much easier to say which is the heavier.) However, comparisons are made directly between shares, rather than between one share and the average of the rest. Thus the theory predicts the relative behaviour of share-prices, one with another, permitting an overall ranking of the shares under consideration. Several authors have mentioned the necessity for a discussion of the relative performance of shares; our theory is an attempt to study precisely this. As will be seen later, the formulation of the theory includes Weaver and Hall's suggestion for the study of discriminant analysis to predict share-prices. The tie-up between their suggestions and Clarkson's model was discussed in the last chapter.

*It is interesting to note that Clarkson's "real" investor - the Mutual Fund Officer is so constrained by laws that his investments perform no better than the market average - from West & Tiniç "The Economics Of the Stock Market", 1971.

** Technicalism can be looked upon as a similar sort of theory - its predictions are "tested" by the continual discovery of similar "patterns" in share-price graphs.

Thus, in conclusion, a statment of the theory. The baldness of the statement leaves out some of the detail since the theory concentrates on the ultimate prediction i.e. the relative performance of share-prices and is not so concerned with the internal workings or how they are represented in a working model.

We shall hypothesise that the average investor, when taking an investment decision, considers two shares A & B and studies the current and historic values, for these shares, of several "investment variables" relating to the shares. He then evaluates whether, for each variable, Share A is better than Share B and then combines these binary comparisons in an implicit way to produce an overall comparison of Share A against Share B. He makes no judgement as to how good each share is individually and ignores factors which influence both e.g. the state of the economy, time of year etc. He only compares variables which can be compared without a lot of computation and which are freely available.

Discussion of several points arising from this will accompany the presentation of the results of testing the predictions of the theory, but they include:

Which variables are chosen?
How are historical values evaluated?
What is the nature of the implicit formation of the overall comparison?

However, in order to progress, it is clear that some technique of handling comparisons will be needed. Furthermore, strict <u>binary</u> comparison is something the human-being is loath to do. Many times he will say (sometimes he <u>has</u> to say) "I don't know." (The "don't knows" in an opinion poll comprise some "don't cares" but also many genuine expressions of an inability to make a decision, one way or the other.) Thus the development of a <u>ternary</u> logic, with the interpretation "yes, no, don't know" was necessary. The formulation of the theory in terms of this logic prepares the way for a re-interpretation later in terms similar to the discriminants mentioned by Weaver and Hall. The next chapter describes the workings of the logic.

A 3-VALUED LOGIC

This particular formulation of a three-valued logic happened to be the one most relevant to the job in hand. It happens also to coincide with Tarski's modification of Lukasiewicz' original formulation, in 1920. In particular it relies heavily on what Tarski called the "modal function". However, they talked exclusively about truth-value systems in which the meanings attached to the values of the variables (or propositions) are "certainly true", "doubtful", and "certainly false". In this formulation the semantics are different (as are the actual values chosen). The presentation of the logic is slightly formal, as it should be, the use of it being described in the next chapter.

OPERANDS

All operands are represented by upper-case letters, thus: A, B, C etc. The values any one operand (or expression) can take are three in mumber: -1, ±1, 1. The reason for choosing the symbol ±1 for the third value is to emphasise its "don't know" nature. Any three symbols could have been chosen, but the particular interpretation required is one of "less than", "don't know", and "greater than". The "natural order" of the values is -1, ±1, 1 and this order will be important later in the usage of the logic. The corresponding arithmetic values of the three symbols will also be important i.e. the numbers -1, ±1, 1.

The letters X, Y, Z will be used to signify any expression of operands A, B, C etc.

OPERATORS

There are four operators in all, two unary operators and two binary operators; (these qualifications refer to the number of operands needed for each operator to be meaningful).

The two unary operators, applied to any operand A, are:

$\overline{A}$ called "A over-bar"

A* called "A star"

Their effects are given in the following table.

A	$\overline{A}$	A*
-1	1	-1
±1	±1	1
1	-1	1

The effect of "over-bar" is similar to negation in two-valued logic, except that its effect on "don't know" is nil.

The effect of "star" is to eliminate any uncertainty i.e. its results for any value of A are absolutely certain. It is this operator which Tarski called the modal operator, and A* would be read as "A is possible", but he conceded this stretches the language as far as it will go. Here the operator is introduced as an expediency to simplify the workings of the theory. This will become clear in the next chapter.

These two operators can be combined in several ways; the table below lists all the non-trivial ones.

A	$\overline{A}^{*}$	$\overline{A^*}$	$\overline{\overline{A}^{*}}$
-1	1	1	-1
±1	1	-1	-1
1	-1	-1	1

Note that $\overline{\overline{X}} = X$ and $X^{**} = X^*$

The two binary operators are the familiar ones from two-valued logic, conjunction and disjunction. Here they are represented by the Boolean algebra symbols "+" and "." (or, almost exclusively, just nothing as in normal algebraic multiplication). Their results for all combinations of the value of two operands are given at top of following page.

A	B	A + B	AB
-1	-1	-1	-1
-1	±1	±1	-1
-1	1	1	-1
±1	-1	±1	-1
±1	±1	±1	±1
±1	1	1	±1
1	-1	1	-1
1	±1	1	±1
1	1	1	1

The results in boxes will be seen to correspond to Boolean algebra. It is also clear by inspection that:

$$A + B = B + A \quad \text{and} \quad AB = BA$$

as in Boolean algebra. This correspondence not only simplifies future analysis, but provides a firm base for it.

FUNCTIONS OF ONE VARIABLE.

There are a great many such functions; the only ones given here are ones which have proved useful or are interesting in themselves.

With constants:

$$A + 1 = 1 \qquad A.1 = A$$

$$A + -1 = A \qquad A.-1 = -1$$

Producing constant results:

$$\overset{*}{\overline{A}} + A^* = 1 \qquad \overline{\overset{*}{\overline{A}}}\ \overline{A^*} = -1$$

$$\overline{A}^* + A^* = 1 \qquad A^*\overline{A}^* = -1$$

$$\overline{A} + A^* = 1 \qquad A\ \overline{A}^* = -1$$

Important ones involving "star" are:

$$X\,X^* = X \quad \text{and} \quad X + X^* = X^*$$

"absorption" functions:

$$\overset{*}{\overline{A}} + \overline{A^*} = \overset{*}{\overline{A}}$$

$$A^* + \overline{\overset{*}{\overline{A}}} = A^*$$

$$A^* + A = A^*$$

$$A + \overline{\overset{*}{\overline{A}}} = A$$

$$\overline{A} + \overline{A^*} = \overline{A}$$

(There are corresponding conjunctive forms)

There are twenty seven different combinations of the three values -1, ±1, 1 taken three at a time. All of these can be expressed as functions of one variable, some with constants as well. They are listed in the Appendix.

FUNCTIONS OF TWO VARIABLES

Corresponding to Boolean algebra we have rules covering DeMoivre's laws:

$$\overline{XY} = \overline{X} + \overline{Y} \quad \text{and} \quad \overline{X + Y} = \overline{X}\,\overline{Y}$$

Also as in Boolean algebra:

$$X + XY = X \quad \text{and} \quad X(XY) = XY$$

Very important rules are:

$$(X + Y)^* = X^* + Y^* \quad \text{and} \quad (XY)^* = X^*\,Y^*$$

Corresponding to the law of absorption in Boolean algebra ($a + \bar{a}b = a + b$) we have:

$$A^* + \bar{A}B = A^* + B$$

FUNCTIONS OF THREE VARIABLES

The laws of distribution and association are the same as Boolean algebra.

$$X(Y + Z) = XY + XZ$$

and $$X(YZ) = (XY)Z = Y(XZ)$$

APPENDIX

The 27 combinations of three values -1, ±1, 1 expressed as functions of one variable A.

-1	-1	-1	$A\overline{A^*}$
-1	-1	±1	$\overset{*}{\overline{A}}.\pm 1$
-1	-1	1	$\overset{*}{\overline{A}}$
-1	±1	-1	$A^*\overset{*}{\overline{A}}.\pm 1$
-1	±1	+1	$A.\pm 1$
-1	±1	1	A
-1	1	-1	$A^*\overset{*}{\overline{A}}$
-1	1	±1	$A^*\ (\overset{*}{\overline{A}} + \pm 1)$
-1	1	1	A^*
±1	-1	-1	$\overline{A^*}.\pm 1$
±1	-1	±1	$(\overline{A^*} + \overset{*}{\overline{\overline{A}}}).\pm 1$
±1	-1	1	$\overset{*}{\overline{\overline{A}}} + \overline{A^*}.\pm 1$
±1	±1	-1	$\overline{A}.\pm 1$
±1	±1	±1	$(\overline{A} + \overset{*}{\overline{\overline{A}}}).\pm 1$
±1	±1	1	$A + \pm 1$
±1	1	-1	$\overset{*}{\overline{A}}\ (A^* + \pm 1)$
±1	1	±1	$A^*\ \overset{*}{\overline{A}} + \pm 1$
±1	1	1	$A^* + \pm 1$
1	-1	-1	$\overline{A^*}$
1	-1	±1	$\overline{A^*} + \overset{*}{\overline{\overline{A}}}.\pm 1$
1	-1	1	$\overline{A^*} + \overset{*}{\overline{\overline{A}}}$
1	±1	-1	$\overline{A}$
1	±1	±1	$\overline{A} + \pm 1$
1	±1	1	$\overline{A^*} + \overset{*}{\overline{\overline{A}}} + \pm 1$
1	1	-1	$\overset{*}{\overline{A}}$
1	1	±1	$\overset{*}{\overline{A}} + \pm 1$
1	1	1	$\overline{A} + A^*$

THE THEORY FORMALISED AND ANALYSED

This chapter describes how the logic presented in the last chapter is utilised in formalising the theory outlined earlier. Its use is purely analytic and no attempt is made to postulate the existence of such a logic in the workings of the brain. From this aspect, the theory is very much a black-box or input-output theory. The logic and its associated computation rules can then be thought of as a description (in Braithwaite's terms a model*) of the mechanism which activates decisions such as the theory attempts to cater for. As we shall see later, another, more concise (perhaps more Cybernetic) means of arriving at the same result from the same data will play a very important part. Again this is only a computational device and not an attempt at brain theory. Both techniques, the logic and the simple perceptron described later have the right connotation with our belief of the brain's workings. This is why they are such good descriptions of mechanisms in the brain which are still not understood at all clearly. (It may be that because of the brain's vast complexity, any theory of the brain will be a mere description of its workings. It will be none the less a useful predictive tool.)

THE INPUT DATA

The choice of data with which to test the theory was difficult, mainly due to unfamiliarity with the investment world. The data had to be readily available and yet sufficiently comprehensive to permit worthwhile testing. The source eventually chosen was the popular investment information service Moodie's and, in particular their "Investment Handbook". The listed variables at that time (1971) were eleven in number. (The Handbook has since been enhanced). They were:

1. Sales - the turnover of the company in one year.
2. Total profits - in one year, before tax etc.
3. Earned for Ordinary - the amount of profit set aside for the dividend on ordinary shares.

*Braithwaite "Scientific Explanation" 1953.

4. Earned income - the amount, as a percentage, of the maximum profit that could be paid out as dividend (based on the latest year's capital.

5. Dividend - the actual percentage paid.

6. Net assets per share - the theoretical value of the company on liquidation averaged per share.

7. Average share price. - average of the highest and lowest prices during the year.

8. Yield ratio - the ratio of the dividend yield (the percentage of the share-price paid as dividend per share) to the average yield on Moodies Equity index at the same date.

9. Ordinary capital - the amount of capital issued as ordinary shares at the end of the financial year.

10. Capital employed - the total capital reserves of the company.

11. Pre-tax profits - profits before tax etc. but after capital depreciation has been accounted for.

Values of these eleven variables were given for the past ten years in yearly intervals.* Thus for each company there were ten values for each variable. All of this information was taken and used as data input.

Ten companies were chosen for testing the theory and they were all chosen from one industrial grouping, mainly on a vague assumption that they would be more comparable. In fact it is very common to consider groupings such as these when making investment decisions, probably for this very reason. The ten chosen were all chain-stores:

* 1961 to 1970 inclusive, values taken March each year.

1. Boots.
2. Foster Brothers.
3. Curry's.
4. Marks and Spencer.
5. Dorothy Perkins.
6. Austin Reed.
7. W.H.Smith.
8. H.Samuel.
9. Army & Navy Stores.
10. Debenhams Stores.

Ten may seem a small number with which to test a theory, but remembering that the basis of the theory is comparison, in fact there are forty five individual pieces of information. Whether even this is adequate, and several other points about the choice of data will be postponed for a discussion of the results.

For the following explanation, let V (i, j, k) represent the value of the jth variable in the kth year of the ith share. The range of each subscript is from one to ten. P (i, k) represents the share-price of the kth year of the ith share.

PRE-PROCESSING OF THE DATA

In order to be predictive in nature, the data must be organised to reflect this nature. Clearly, we shall be predicting something about share-prices, since this, as has been mentioned before, is the core of the investment situation. Absolute prices are, of course, not comparable, thus ratios of the prices of adjacent years were taken. This reflects, for each company, the percentage increase in share-price over the previous year.

Symbolically this is

$$P'(i,k) = P(i,k+1) / P(i,k)$$

P' is now comparable between shares and is the predicted variable. A similar situation exists with other variables, namely sales, profits of one sort or another, and the two capital values. (variable nos. 1, 2, 3, 9, 10, 11). The remaining variables are ratios and so are, in principle, comparable between shares. For this reason, a second set of data was generated from the first by taking ratios with all the variables in a similar fashion to the share-price as shown above. The new variables can all be interpreted as "percentage increase over last year" and are, being ratios, directly comparable between shares.

They are derived thus:

$$V'(i, j, k) = V(i, j, k) / V(i, j, k-1)$$

This reduces the number of years for which data is available from ten to nine but this is not serious. This point will also be discussed later. The variables which were originally ratios (i.e. nos. 4, 5, 6, 8) are now ratios of ratios but this is not a serious fault for two reasons. One is that there are now two sets of data for analysis and in either one or the other each variable is comparable (but not both). The other reason is that if Moodie's see fit to include comparable variables with non-comparable variables (i.e. ratios and absolute values), and if many investors typically use such data, then perhaps comparison of absolute values (i.e. essentially non-comparable ones) is something that must be accounted for. (It is certainly true that absolute price increases are considered important, because all changes are recorded in this way. If a share rises by five pence it is considered exceptional whether the shares are worth twenty pence or two hundred.*)

COMPARISON OF TWO SHARES

Once the data for each share has been pre-processed we can now consider a comparison of two shares and then go on to the problem of forming a prediction.

*This has implications for a theory of the mechanism of the Stock Exchange. i.e. the behaviour patterns of the jobbers.

The two shares used to illustrate this are in fact numbers 2 and 3, Foster Brothers and Curry's. The formation of the comparison array is done quite simply by subtracting corresponding variables and testing whether the result is greater than, less than, or equal to zero. The presence of a 1 means less than; a 2 means equal to (don't know); a 3 means greater than.* The array is then, for these two particular shares:

VARIABLE	1	2	3	4	5	6	8	9	10	11	7
YEAR											
1970	1	1	1	1	1	3	3	1	3	1	?
1969	1	1	1	1	1	3	3	2	3	1	3
1968	3	1	3	3	1	3	1	3	1	3	1
1967	1	3	3	3	2	3	3	2	3	3	3
1966	3	1	1	1	1	1	1	2	3	1	3
1965	3	3	3	3	3	3	1	1	3	3	1
1964	3	1	1	1	1	3	3	3	3	1	1
1963	2	1	1	1	2	3	1	2	1	1	1
1962	2	3	3	3	1	3	3	1	3	3	3

The question-mark against year 1970's share-price (variable seven) signifies that this is the value to be predicted.

FORMATION OF THE PREDICTION

This will be presented in two different ways, as mentioned already. One way (the logical-net approach*) is purely analytic; the other way (the perceptron approach**) is purely computational. In fact the logical-net method of analysis will be used to analyse and formalise the perceptron method of computation, so the two are really equivalent. It must be stressed again, though, that neither attempts to explain the workings of the brain. They are both useful descriptions (at a meta-level) of possible aspects of the brain's behaviour.

---

* All of the computational work was done by computer. Much of the representation of results and the terms used to describe them reflects this.

** A concise description is given in Klir and Valach "Cybernetic Modelling" 1967 (they use binary logic).

*** The supreme analysis of perceptrons is in Minsky and Papert "Perceptrons" 1969.

It will be shown, by analysis, that the data described above is linearly separable, to use pattern-recognition jargon and so the simplest perceptron, one of order one, can be used to compute the predicted value.

THE LOGICAL-NET APPROACH

This, to my knowledge, has always been described using binary logic and so a little time must be spent in studying what effects the three-valued logic has on it. Let us take a simple example in binary logic and, from the data which I shall invent, show how an expression is formed which describes the date set. We shall have three variables as input to a black-box. Each input can take the values 0 (off) and 1 (on) as in Boolean algebra. The output is a single variable with the same possible values. Various combinations (not all the possible ones) of input values gave the following table.

Variable row	a	b	c	output
1	0	0	1	1
2	1	0	1	0
3	1	1	0	0
4	0	1	1	1
5	1	1	1	0
6	0	1	0	1

Each row represents a state of the black-box at any particular instant in time. The formation of the descriptive expression is achieved by taking, for each row with a 1 in the output column, the conjunction of variables (negated if necessary) which could produce a 1 as output. These conjunctions are then formed into one disjunction to give the final expression.

Thus:
row 1 gives $\bar{a}\ \bar{b}\cdot c$
rows 2, 3 are ignored
row 4 gives $\bar{a}\ b\ c$
row 5 is ignored
row 6 gives $\bar{a}\ b\ \bar{c}$

The expression is thus $\bar{a}\,\bar{b}\,c + \bar{a}\,b\,c + \bar{a}\,b\,\bar{c}$

The reason why this algorithm works is clear when the final expression is studied. Such an expression can only take two values, 0 and 1. If it takes the value 1 then either $\bar{a}\bar{b}c$ or $\bar{a}bc$ or $\bar{a}b\bar{c}$ must have the value 1. This is the meaning of the disjunction sign "+". For rows 1, 4, 6 at least one of $\bar{a}\bar{b}c$, $\bar{a}bc$, $\bar{a}b\bar{c}$ is 1 and so the whole expression has value 1. Any other combination of values of a, b, c gives a value of 0 for the expression. This is the reason why this method is useless as a predictive tool. It describes the data already presented exactly but says that any other input not already presented will give value 0 at the output. Since I take it that prediction is really about unique events in the future i.e. events which have not happened before, then the method is inadequate.

It is interesting to note that we can incorporate the '0' output rows by, in a sense, inverting the algorithm. i.e. take a disjunction of variables negating where necessary, and form a conjunction of the expressions thus formed.

i.e. for rows 2, 3, 5 we get

$$(\bar{a} + b + \bar{c})\,(\bar{a} + \bar{b} + c)\,(\bar{a} + \bar{b} + \bar{c})^{*}$$

The total expression is thus:

$$(\bar{a}+b+\bar{c})\,(\bar{a}+\bar{b}+c)\,(\bar{a}+\bar{b}+\bar{c})\,(\bar{a}\bar{b}c + \bar{a}bc + \bar{a}b\bar{c})$$

If we set $(\bar{a} + b + \bar{c})\,(\bar{a} + \bar{b} + c)\,(\bar{a} + \bar{b} + \bar{c}) = X$ and

$(\bar{a}\bar{b}c + \bar{a}bc + \bar{a}b\bar{c}) = Y$ then

the expression is XY.

However, if we consider each term in Y taken with each term in X, as if we were expanding the expression, we can see why the X group is superfluous. Either the product disappears altogether (e.g. $(\bar{a} + b + \bar{c})\,\bar{a}\bar{b}c = 0$), or only the Y group expression remains. (e.g. $(\bar{a} + \bar{b} + c)\,\bar{a}\bar{b}c = \bar{a}\bar{b}c$). This is a simple proof of the algorithm, using the two laws $aa=a$ and $a\bar{a}=0$. It again emphasises the paucity of

* This is identical to $\overline{a\bar{b}c}\;\overline{ab\bar{c}}\;\overline{abc}$ so the previous analysis holds for this set as well.

this approach for prediction purposes, but illustrates the basis of a similar algorithm for the three-valued logic.

Before that, however, it must be noted that as often as not, the expression resulting from the above procedure can be reduced to a simpler form, by using the reduction rules of Boolean algebra. e.g. in the example given:

	$\bar{a}\,\bar{b}\,c + \bar{a}\,b\,c + \bar{a}\,b\,\bar{c}$	
=	$\bar{a}\,(\bar{b}\,c + b\,c + b\,\bar{c})$	using the laws of distribution and association
=	$\bar{a}\,(\bar{b}\,c + b\,(c + \bar{c}))$	using the laws of distribution and association
=	$\bar{a}\,(\bar{b}\,c + b)$	since $c + \bar{c} = 1$ and $b.1 = b$
=	$\bar{a}\,(b + c)$	since $c + \bar{c}b = c + b$ (law of absorption)

This simplified expression is a better description because it has condensed the information contained in the data set into fewer symbols.

Now consider the same problem but with the following data set:

VARIABLE ⟶	A	B	C	output
row				
1	±1	1	1	-1
2	1	-1	-1	-1
3	-1	1	±1	1
4	±1	-1	±1	1
5	-1	-1	±1	-1
6	1	1	-1	1

The nomenclature is as in the chapter introducing the three-valued logic. The algorithm to form the descriptive expression is as follows:

Take each row with output 1 and form conjunctions, with over-bars and stars as necessary to produce the output of 1. Form a disjunction of these to give expression Y. i.e. $Y = \bar{A}BC^* + A^*\bar{B}C^* + AB\bar{C}$ from rows 3, 4, 6.

The expression X is formed similarly but in a converse fashion, just as in the two-values example.

$$X = (\overline{A}^{*} + \overline{B} + \overline{C})\ (\overline{A} + B + C)\ (A + B + \overline{C^{*}})$$

Note the use of the star to turn ±1 into 1.

The final descriptive expression is not XY as in two-valued logic but X*Y i.e.†

$$(\overline{A^{*}} + \overline{B}^{*} + \overline{C}^{*})\ (\overline{A}^{*} + B^{*} + C^{*})\ A^{*} + B + \overline{C^{*}})\ (\overline{A}\, B\, C^{*} + A^{*}\, \overline{B}\, C^{*} + AB\overline{C})$$

This, as mentioned before, is the prime reason for choosing the operator* as it is. The reason is as follows: each of the terms within Y can take one of three values and so can Y itself. If anyone of the expressions in Y has value 1, then Y is 1 and so must be the whole expression. If none of the terms within Y has value 1 then Y can take a value of either ±1 or -1. In this case, one of the expressions with X (e.g. $(\overline{A^{*}} + \overline{B} + \overline{C})$) may be -1, meaning that X is -1 and so is the whole expression. Thus if Y= -1 or ±1 and X = -1 everything is fine. If, however, Y = 1, X could be either 1 or ±1. It is this case of Y = 1, X = ±1 which must be avoided since 1. ±1 = ±1‡, and not 1, as required. This rather complicated reasoning can be summarised by the table below, giving all possible values of X and Y and the required outcome.

X	Y	outcome
-1	-1	-1
1	1	1
-1	±1	-1
±1	1	1

This table only includes cases where either X = -1 or Y = 1 but not both. (This is impossible since a single term must be in X or Y but cannot be in both simultaneously). It can be seen that the expression describing the outcome is XY except for the last row. An operator on X is needed to convert only that row so as to give the correct output of 1. This operator is the star and so X*Y is the required expression. The other combinations of values give different.

† Note that (A+B)* = A* + B* and (AB)*. Also that $\overline{A}^{*} = A$
‡This approach does not allow for outputs of ±1.

results, corresponding to predictions for inputs not yet presented. They are:

X	Y	outcome = X*Y
±1	-1	-1
1	±1	±1
±1	±1	±1
1	-1	-1

It can also be seen that different inputs, apart from giving outputs of ±1 or -1, as expected, can give outputs of 1. e.g. the input A = -1. B = 1, C = 1 gives X* Y= 1.

Thus a distinct gain over two-valued logic is obtained in that new inputs can produce all three values as output, depending on their form.

As with two-valued logic, there is a possibility of reduction of the final expression X*Y. In particular, the terms contained in X usually disappear since they only need a "match" in one variable with each of the expressions in B.

e.g. $$(\overline{A^*} + \overset{*}{\overline{B}} + \overline{C})\, A^* \overline{B}\, C^* = A^* \overline{B}\, C^*$$

since $A^* \overline{B}\, \overset{*}{\overline{B}}\, C^* = A^* \overline{B}\, C^*$ and $(\overline{A}^* + \overset{*}{\overline{C}})\, A^* \overline{B}\, C + A^* \overline{B}\, C^* = A^* \overline{B}\, C^*$

carrying out this manipulation in the example above gives:

$$(\overline{A}^* + \overset{*}{\overline{B}} + \overset{*}{\overline{C}})\, \overline{A}\, B\, C^* + A^* \overline{B}\, C^* + A\, B\, \overline{C}$$

since the other terms in X disappear.

This is a reduction from 81 (3^4) terms in the original expression, down to 5. We will now look at the example taken from the investment data and look at the reduction problems there. The variables 1 to 6 and 8 to 11 will be re-named. A to J, the share-price will be S. It can now be seen that each year, i.e. 1962 to 1969, yields one expression, either in the X group if the value of S is 1 (in the comparison array) or in the Y group if the value of S is 3.

Forming the two groups as demonstrated above gives:

$$X = (\bar{A} + B + \bar{C} + \bar{D} + E + \bar{F} + G + \bar{H} + I + \bar{J}) \quad (\bar{A} + \bar{B} + \bar{C} + \bar{D} + \bar{E} + \bar{F} + G + H + \bar{I} + \bar{J}) \; (\bar{A} + B + C + D + E + \bar{F} + \bar{G} + \bar{H} + \bar{I} + J) \quad (\bar{A}^* + B + C + D + \overline{E^*} + \bar{F} + G + \overline{H^*} + I + J)$$

$$Y = \bar{A}\,\bar{B}\,\bar{C}\,\bar{D}\,\bar{E}\,F\,G\,H^*\,I\,\bar{J} + \bar{A}\,B\,C\,D\,E^*\,F\,G\,H^*\,I\,J + A\,\bar{B}\,\bar{C}\,\bar{D}\,\bar{E}\,\bar{F}\,\bar{G}\,H^*\,I\,\bar{J} + A^*\,B\,C\,D\,\bar{E}\,F\,G\,\bar{H}\,I\,J$$

When X + Y is formed it can be seen that all the members of X are absorbed by each member of Y, there being at least one variable in common in each partial product. Thus $X^*Y = Y$ and apart from a common factor of I, this cannot be reduced further.

This final expression is typical of the ones that result from the investment data and, as perhaps could have predicted, gives a result of "less than" when applied to the figures for 1970. To give a result of 1 the 1970 figures would have to be the same as a set of previous figures except for the variables with values of 2 (corresponding to ±1 or "don't know"). To give a result of ±1 one of the terms would have to be ±1, again meaning only a small change from a previous expression. Both of these events are fairly unlikely because of the small number of inputs (8 from a maximum of 3^{10} or 59, 049) and so we would expect a prediction of -1. The figure of 8 from 59, 049 is misleading because it disregards the possibility of regularities in the data. If the sets of figures change slowly then the probability of obtaining a 1 or ±1 output clearly increases. Since this is "real-world" data one would hope that this is indeed the case. If it is not then the predictions will be spurious and the results overall must be approached with caution. This is discussed more fully when the actual results are presented.

In any case, there are problems with this method of analysis. The amount of computation, whether by computer or by hand, is almost certainly not worth the effort for the reasons outlined above. Another approach, which is shown to be equivalent (in a restricted sense) and involves much less computation is presented in the next chapter. The chapter after that presents a closely allied method of analysis and encompasses the results obtained by the simple <u>perceptron,</u> a discussion of which now follows.

SOME PERCEPTRON THEORY

The perceptron is here presented as a computational device; specifically to reproduce the results produced by the sort of logic utilised in the last chapter. The connotations that the perceptron has of being a "pattern-recognition" device or a "learning machine" are not here intended to be any more than useful descriptive analogies. In the same way we talk, in physics, about "electrons jumping up and down energy levels". This of course they do not do, but every physicist talks in these terms because they are immediately graspable and so are easy to work with.

Thus we will talk about the figures for each year in the investment data (the comparison array again) as a <u>pattern</u>, and this pattern is to be <u>recognised</u> or <u>categorised</u> by the perceptron as meaning an increase or decrease in relative share-prices.

The following is a definition of the perceptron taken from Minsky and Papert but modified slightly in order to fit the three-valued logic.

A predicate is a function ϕ of some pre-defined part of a pattern. It can take the values -1, ±1, 1 according as the predicate is not true, undecided or true. In our case the predicate is "V_i is greater than V_j" where V_i and V_j are two values of a variable taken from corresponding years in two different shares. The set of predicates $\{\phi A, \phi B, \ldots, \phi J\}$, when applied to variables A to J for each share generate the comparison array already described. However, in general there will be a family Φ of predicates i.e. $\{\phi_1, \phi_2, \ldots, \phi\}$ which generate corresponding comparison valves when applied to a pattern X which is defined as any set of distinguishable values.*

A perceptron Ψ is then defined as producing values of -1, ±1, 1 when applied to the pattern X $[\text{thus } \Psi(X)]$, according as $\sum_{i=1,n} \alpha_i \phi_i(X)$ is less than, less than or greater than,

*Minsky and Papert talk in geometric terms about real patterns in the Euclidean plane. It is easy to map the investment data onto such a formulation.

and greater than θ respectively, where α_i and θ are numbers chosen to make the categorising correct. θ is called the threshold and the α_i are weights. Such a perceptron is said to be linear with respect to Φ.

Briefly, this is

$$\begin{aligned}\Psi(X) &= 1 \text{ if } \sum \alpha \phi(X) > \theta \\ \Psi(X) &= \pm 1 \text{ if } \sum \alpha \phi(X) > \text{or} < \theta \text{ (or } = \theta) \\ \Psi(X) &= -1 \text{ if } \sum \alpha \phi(X) < \theta\end{aligned}$$

As discussed in the chapter on the three-valued logic, the values -1, ± 1, 1 represent "less than", "don't know" and "greater than". An important step is to remove this restriction and let the values represent the numbers -1, ± 1, 1 so that we can use these arithmetic values in direct computations. In much the same way the true and false values of Boolean algebra can become the numbers 1 and 0 and expressions can be evaluated arithmetically, with the restriction that $1 + 1 = 1$ in Boolean arithmetic.

Much the same sort of thing can be done with perceptrons. Consider the predicate "a is in X and b is in X" where a and b are two points in the Euclidean plane, X is some subset of the plane. This can be written, in Boolean algebra terms, as the simple term (ab) since it takes the values 1 and 0 as the worded predicate demands. In arithmetic terms this can be written

$$\lceil a + b > 1 \rceil \qquad \text{instead of } ab$$

where the brackets "$\lceil \quad \rceil$" denote taking the value 1 if the inequality is true, and 0 if false. To make this clearer, it can be represented in table form thus:

a in X	b in X	a b	ab in X	a + b	a + b>1.	$\lceil a + b > 1 \rceil$
FALSE	FALSE	0 0	FALSE	0	FALSE	0
FALSE	TRUE	0 1	FALSE	1	FALSE	0
TRUE	FALSE	1 0	FALSE	1	FALSE	0
TRUE	TRUE	1 1	TRUE	2	TRUE	1

By inspection, this formulation fits the definition of a linear perceptron, with weights equal to 1 and threshold also 1.

$$\text{i.e.} \quad \alpha_1 = 1 , \alpha_2 = 1 , \quad \theta = 1$$

Thus we have the basis of the parallel nature of logical forms and perceptrons. Nothing will be said in this chapter about the computational problems of the perceptron, but it is basically one of calculating the weights α_i and the threshold θ when presented with a set of patterns with known categories.

However, we must look at the logical formulation of the perceptron in order to know that we are giving the perceptron a good chance of finding the weights and threshold i.e. that such numbers do exist at all. The reason is that we have restricted our perceptron to one which is linear in the predicates ϕ_i *. We have seen how such predicates can be combined in Boolean algebra expressions to produce perceptrons with different functions. It is necessary to ask if all the possible Boolean expressions can be formulated as linear perceptrons.

It turns out that this becomes less and less true as the expressions become more complex, i.e. involve more simple predicates. Let us again take the two-predicate expressions, as we took the expression a.b above.

There are sixteen Boolean expressions of two variables and, in fact, fourteen of them are capable of representation by linear perceptrons. The two which are not are given below:

a	b	$a \oplus b$	$a \equiv b$
0	0	0	1
0	1	1	0
1	0	1	0
1	1	0	1

*We only consider "single-point" predicates. M & P call these perceptrons "order one" perceptrons

As a simple proof that they are not linear, assume that a $\oplus$ b <u>is</u> linear and that

$$\lceil \alpha a + \beta b > \theta \rceil$$

then we must have, for each combination of values of a and b,

$$0 \leqslant \theta, \quad \beta > \theta, \quad \alpha > \theta \text{ and } \alpha + \beta \leqslant \theta$$

Clearly we have $\alpha + \beta > 2\theta$ and $\alpha + \beta \leq \theta$ which is impossible thus $a \oplus b$ is not linear. The proof for $a \equiv b$ is similar.

To reach a perceptron formulation for $a \oplus b$, consider that it can be written $\bar{a}b + a\bar{b}$. Minsky and Papert show how this can be written as a perceptron:

$$\lceil (1 - a)\, b + a\, (1 - b) > 0 \rceil$$

i.e. $\lceil a + b - 2\, ab > 0 \rceil$

Thus $a \oplus b$ is not linear but of order 2.

When expressions with more variables are considered, the fraction which are capable of linear representation decreases rapidly. However, there is a class of expressions (discovered during the course of this work) of any number of variables, which are linear, although at first sight they do not appear to be so. Minsky and Papert show that expressions of the form abc...and a + b + c.... are linear but it is true also that expressions of the forms:

$f(a, g) = a + g(b, f)$ and

$f(a, h) = a.\ h(b, f)$ where a and b are any simple predicates are linear

Note that these definitions are recursive and that f can be null (i.e. $f(a, 0) = a$). These two forms can be seen to include abc... and a + b + c... as special cases.

A slight generalisation of these special cases is the expression:

$a_1 + a_2 + .. a_m + b_1 b_2 b_3 ... b_n$ which has the perceptron form:

$$\left\lceil \sum_{i=1,n} b_i + n \sum_{j=1,m} a_j > n - 1 \right\rceil \equiv \lceil E > 0 \rceil \quad \text{where}$$

$$E = \sum b_i + n \sum a_j - n + 1$$

To prove this, consider the two cases $\lceil E > 0 \rceil = 0$ and $\lceil E > 0 \rceil = 1$.

If $\lceil E > 0 \rceil = 0$, all the a_j must be 0 <u>and</u> at least one of the b_i. Then:

$$\text{Max} \left\{ E \;\middle|\; \lceil E>0 \rceil = 0 \right\} = n - 1 - n + 1 = 0$$

Since $0 \not> 0$, $\lceil E>0 \rceil = 0$ as was assumed at first.

If $\lceil E \not> 0 \rceil = 1$ at least one of the a_j must be 1 <u>or</u> $b_1 b_2 \ldots b_n = 1$

then: $\text{Min} \left\{ E \;\middle|\; \lceil E>0 \rceil = 1 \right\} = n - n + 1 = 1$

Since $1>0$, $\lceil E > 0 \rceil = 1$ as was assumed. Thus the postulated perceptron form is verified..

To prove the general forms $a + g(b, f)$ and $a.h(b, f)$, first consider the simple expression ab which we know is linear, of the form $\lceil a + b > 1 \rceil$.

Now let $b = \lceil \beta > \theta \rceil$ where β is an arithmetic expression with maximum value n (i.e. $0 \leqslant \beta \leqslant n$) and θ is an integer.

Thus we have: $ab = \lceil a + b > 1 \rceil = \lceil a + \lceil \beta > \theta \rceil > 1 \rceil$

The aim is to allow b to be replaced by a Boolean expression with perceptron form $\lceil \beta > \theta \rceil$. We will then show how the brackets around $\lceil \beta > \theta \rceil$ can be removed.

Now, $0 \leqslant \theta < n$ since β must be capable of being both greater than and less than θ for $\lceil \beta > \theta \rceil$ to have two possible values. Thus $0 \leqslant \theta/n < 1$ and this means that the following inequalities are true:

$$-\theta/n < 1 \quad \text{or} \quad (0) + (0 - \theta/n) < 1$$
$$1 - \theta/n \leqslant 1 \quad \text{or} \quad (0) + (1 - \theta/n) \leqslant 1 \text{ and } (1) + (0 - \theta/n) \leqslant 1$$
$$2 - \theta/n > 1 \quad \text{Or} \quad (1) + (1 - \theta/n) > 1$$

These inequalities can be seen to fit the expression

$\lceil a' + \beta' - \theta/n > 1 \rceil$ where a' and β' are ordinary Boolean variables.

If we now put $\beta' = \beta/n$ and $a' = a$ and compare with the form:

$\lceil a + \lceil \beta > \theta \rceil > 1 \rceil$ we can derive:

$$ab = \left\lceil a + \frac{\beta - \theta}{n} > 1 \right\rceil \qquad \text{where } b = \lceil \beta > \theta \rceil$$

In similar fashion it can be shown that

$$a + b = \lceil a + \frac{\beta - \theta}{n} > 0 \rceil \quad \text{where } \lceil \beta > \theta \rceil = b \text{ and } 0 \leqslant \beta \leqslant n$$

We thus have an algorithm for making substitutions in expressions and preserving the linearity of their perceptron form. For example:

consider E = a + b (c + def), an arbitrary expression of the correct form.

$$\text{def} = \lceil d + e > 1 \rceil \; f = \lceil \lceil d + e > 1 \rceil + f > 1 \rceil \quad \text{where } 0 \leqslant d + e \leqslant 2$$

$$\text{def} = \lceil \tfrac{1}{2} (d + e - 1) + f > 1 \rceil$$
$$= \lceil d + e + 2f > 3 \rceil *$$

$$\therefore \; c + \text{def} = \lceil c + \lceil d + e + 2f > 3 \rceil > 0 \rceil \quad \text{where } 0 \leqslant d + e + 2f \leqslant 4$$

$$\therefore \; c + \text{def} = \lceil c + \tfrac{1}{4} (d + e + 2f - 3) > 0 \rceil$$
$$= \lceil 4c + d + e + 2f > 3 \rceil$$

$$\therefore b(c + \text{def}) = \lceil b + \lceil 4c + d + e + 2f - 3 \rceil > 1 \rceil \quad \text{where } 0 \leqslant 4c + d + e + 2f \leqslant 8$$

$$\therefore b(c + \text{def}) = \lceil b + \tfrac{1}{8} (4c + d + e + 2f - 3) > 1 \rceil$$
$$= \lceil 8b + 4c + d + e + 2f > 11 \rceil$$

$$\therefore a + b(c + \text{def}) = \lceil a + \lceil 8b + 4c + d + e + 2f > 11 \rceil > 0 \rceil \quad \text{where } 0 \leqslant 8b + 4c + d + e + 2f \leqslant 16$$

$$= \lceil a + \tfrac{1}{16} (8b + 4c + d + e + 2f - 11) > 0 \rceil$$

$$\text{thus: } a + b(c + \text{def}) = \lceil 16a + 8b + 4c + d + e + 2f > 11 \rceil$$

Negations, such as in $\bar{a}b$, are easily catered for by the substitution of 1 - a for a in the expression without the negation. e.g. $ab = \lceil a + b > 1 \rceil$

$$\therefore \quad \bar{a}b = \lceil 1 - a + b > 1 \rceil$$
$$= \lceil b - a > 0 \rceil$$

Even though all the expressions mentioned cover a fairly wide range (there are other classes of expression such as "any two from three" i.e. ab + bc + ca**) it is still true that a great many expressions, especially those containing several variables, are not linear. It is now crucially important to realise exactly what is being said here. When we say that ab is linear we are really saying that <u>all</u> the patterns represented by <u>all</u> the combinations of values of a and b are <u>separable</u>

---

* A simpler form is $\lceil d + e + f > 2 \rceil$ but the one derived is no less accurate or valid.
** See Appendix for proof of this form.

into two classes, according as ab = 0 or 1 by a linear perceptron. There are four patterns for ab, namely (0, 0); (0, 1); (1, 0) and (1, 1) and all four can be presented to the perceptron which will then classify them correctly. If these four patterns are presented and the perceptron (linear of course) is told that the function is a + b, then it will fail to classify them correctly. Now consider what happens when only two of the four patterns are presented i.e. (0. 1) and 1, 0). If the perceptron is told that both of these are to be put in the same class then it will have no trouble with the computation. This is true, even though in the two cases the logical-net expression which describes the patterns presented is the same. A table will make this clearer:

All patterns presented:

a	b	outcome
0	0	0
0	1	1
1	0	1
1	1	0

logical-net expression: $\bar{a}b + a\bar{b}$
These two patterns <u>are</u> not linearly separable.

Only two patterns presented:

0	1	1
1	0	1

logical-net expression: $\bar{a}b + a\bar{b}$
These two patterns <u>are</u> linearly separable.

The reason why the perceptron can separate apparently inseparable patterns will be seen in the next chapter. However, it can be said here that if the number of patterns presented to the perceptron is less than the number of variables in each pattern, then the patterns are always linearly separable. This has been proved by Nilsson*, among others. Since in the investment data there are eight or nine patterns only and ten variables it would seem that complete success is assured in using a linear perceptron to compute the predictions. This is true if one remaining fact can be cleared. There must be

* Nilsson "Learning Machines" 1965. There is one condition relating to this statement, although not too serious. See the next chapter.

a way of extending the binary perceptrons used above to three-valued logic. To do this we shall look again at the choice of values in binary logic.

Strictly speaking, the simple perceptron described above cannot fit into the three-valued logic scheme. An expression in this logic can take one of three values. This is equivalent to putting a pattern into one of three different categories and so only one test is needed. With three values, two tests are needed and this involves the use of three functions, one corresponding to each category. These functions, usually called discriminant functions can have a perceptron-like form but the theory of their behaviour is much less understood. However, as previously mentioned, only outputs of 1 or -1 are considered i.e. a binary output for a three-valued input. Thus if we can find that subset of three-valued expressions which have a binary output, then we have a valid extension of the binary perceptron theory to cover three-valued inputs.*

The answer is clear because in the "star" operator we have a means of reducing the variety of input values from three down to the correct two (1, -1) at the output.

Earlier, a general form of binary function with a linear perceptron form was discussed. It will now be shown, using a simple mapping device, why such functions have a linear form. The extension to the three-valued case then is made obvious.

The mapping provides a method of ordering patterns in a consistent way. Each pattern is mapped onto the corresponding integer (binary, ternary, or, in fact, to any base) formed by juxtaposing the individual pattern elements (the values of the simple predicates) in a given fixed order. A simple example makes this procedure clear.

* See Appendix to this chapter for further discussion.

Take the set of (binary) patterns :

	a	b	c	n
pattern 1	1	1	0	6
pattern 2	0	1	0	2
pattern 3	1	0	1	5

If the order of the elements is chosen as abc then the corresponding numbers (converted to decimal) are as the fourth column. Thus the patterns can clearly be ordered 2, 3, 1 with increasing n. n will here be called the <u>pattern-integer.</u> If the order of the elements was c a b the pattern-integers would be 3, 1, 6 and the order of the patterns would then be 2, 1, 3.

If we now consider the set of all possible patterns with two binary elements, the patterns can be seen to have a <u>natural order</u> determined by the integer from 0 to 3:

a	b	n
0	0	0
0	1	1
1	0	2
1	1	3

Similarly with three elements:

a	b	c	n
0	0	0	0
0	0	1	1
0	1	0	2
0	1	1	3
1	0	0	4
1	0	1	5
1	1	0	6
1	1	1	7

Now consider the categorising of such sets of patterns. If we draw a horizontal line anywhere in the table, dividing the whole set of patterns into two categories, the perceptron function is clearly of the form:

$$\lceil 4\,a + 2\,b + c > n \rceil$$

[In writing out this expression we are here working exclusively in denary arithmetic, $4a + 2b + c$ is the denary equivalent of the binary integer abc]

For example take the division thus into categories "A" and "not-A"

a	b	c	category = A ?
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

The perceptron which provides this division is then:

$$\lceil 4\,a + 2\,b + c > 1 \rceil$$

If the categories are reversed then the inequality becomes

$$\lceil 4\,a + 2\,b + c \leq 1 \rceil \quad \text{or, equivalently}$$

$$\lceil 4\,a + 2\,b + c < 2 \rceil$$

It is now interesting to apply the logical-net approach to discover which function of a, b and c has the linear form above. The function can be seen to be:

$$(a\,b\,c + a\,b\,\bar{c} + a\,\bar{b}\,c + a\,\bar{b}\,\bar{c} + \bar{a}\,b\,c + \bar{a}\,b\,\bar{c})$$

This reduces to $(ab + a\bar{b} + \bar{a}b)$ which is the same as $(a + b)$

Of course we know that this has a linear perceptron from (and a much simpler one than the one previous) but the procedure described provides a technique for deriving a whole family of linear forms. There are seven such functions, each corresponding to drawing the horizontal line from after the first pattern to after the seventh pattern.

The function of this last is simply (using the logical-net technique again).

$$(a\ b\ c)$$

Successive functions are derived by "adding in"(using the logical "or" operator) the functions corresponding to the next pattern. The following sequence can be derived:

1. $a\,b\,c$
2. $a\,b\,c + a\,b\,\bar{c} = a\,b$
3. $a\,b + a\,\bar{b}\,c = a(b + c)$
4. $a(b + c) + a\,\bar{b}\,\bar{c} = a$
5. $a + \bar{a}\,b\,c = a + b\,c$
6. $a + b\,c + \bar{a}\,b\,\bar{c} = a + b$
7. $a + b + \bar{a}\,\bar{b}\,c = a + b + c$

These seven functions are, in fact, the complete set of functions of three variables which have the form shown earlier to have corresponding linear perceptrons* i.e. the recursive forms:

$$f(a,\ g) = a + g\,(b, f) \text{ and}$$
$$f(a,\ h) = a.\ h\,(b,\ f)$$

We thus have a technique for deriving complete families of such functional forms for any number of variables. Since the basis of the technique is the ability

*There is a corresponding set formed by the negation of these functions.

to form the pattern-integers and thus to order the patterns, it extends easily to cover the three-valued case. In describing this the three values -1, ±1, 1 are mapped onto the ternary arithmetic values 0, 1, 2. Nothing has been lost in so doing since the set -1, ±1, 1 are symbols and have no arithmetic meaning.

The logical-net method is, as demonstrated before, more complicated in the three-valued case since it involves the formation of the expression $X*Y$. Since the number of patterns increased for the same reasons, only the two variable case will be illustrated here. The derivation of the family of functions for higher numbers of variables is, although difficult in execution, no more complex than the method described above. There are eight functions in all.

	Patterns A B	Functions in X^*	Functions in Y	Pattern-integer
1	0 0	$(A^* + B^*)$	$\overline{A}\,\overline{B}$	0
2	0 1	$(A^* + \overline{B^*})$	$\overline{A}\,B^*$	1
3	0 2	$(A^* + \overset{*}{\overline{B}})$	$\overline{A}\,B$	2
4	1 0	$(\overline{A^*} + B^*)$	$A^*\,\overline{B}$	3
5	1 1	$(\overline{A^*} + \overline{B^*})$	$A^*\,B^*$	4
6	1 2	$(\overline{A^*} + \overset{*}{\overline{B}})$	$A^*\,B$	5
7	2 0	$(\overset{*}{\overline{A}} + B^*)$	$A\,\overline{B}$	6
8	2 1	$(\overset{*}{\overline{A}} + \overline{B^*})$	$A\,B^*$	7
9	2 2	$(\overset{*}{\overline{A}} + \overset{*}{\overline{B}})$	$A\,B$	8

Function 1: Patterns 1 to 8 in category "A" (output value 2)
Pattern 9 in category "not A" (output value 0)

$$(\overset{*}{\overline{A}} + \overset{*}{\overline{B}})\,(AB^* + A\overline{B} + A^*B + A^*B^* + A^*\overline{B} + \overline{A}B + \overline{A}B^* + \overline{A}\overline{B})$$

this reduces to $\overset{*}{\overline{A}} + \overset{*}{\overline{B}}$ since $\overline{B} + B^* = 2$

Function 2: Patterns 1 to 7 in category "A"
Patterns 8 and 9 in category "not A"

$$(\overset{*}{\overline{A}} + \overset{*}{\overline{B}})(\overset{*}{\overline{A}} + \overline{B^*})(A\overline{B} + A^*B + A^*B^* + A^*\overline{B} + \overline{A}B + \overline{A}B^* + \overline{A}\overline{B})$$

which reduces to $\overset{*}{\overline{A}} + \overline{B^*}$

If this is continued the following list of eight functions results:

1	$\overset{*}{\overline{A}} + \overset{*}{\overline{B}}$
2	$\overset{*}{\overline{A}} + \overline{B^*}$
3	$\overset{*}{\overline{A}}$
4	$\overline{A^*} + \overset{*}{\overline{A}}\,\overset{*}{\overline{B}}$
5	$\overline{A^*} + \overset{*}{\overline{A}}\,\overline{B^*}$
6	$\overline{A^*}$
7	$\overline{A^*}\,\overset{*}{\overline{B}}$
8	$\overline{A^*}\,\overline{B^*}$

Again there are corresponding functions formed by the negations of the above list and again they form a complete family of functions with linear perceptron forms. The relevant perceptron function is, since the conversion is now from ternary to denary numbers, $9A + 3B + C$.

The family of functions listed above are exactly that subset of three-valued functions of two variables which have only a binary output discussed above. The extension of the perceptron formulation to the three-valued case has thus been demonstrated. In the next chapter we return to practical reality and discuss the problem of <u>linear</u> <u>separability</u> and describe a method for determining whether or not a given set of patterns satisfies this criterion or not.

APPENDIX TO SOME PERCEPTRON THEORY

To demonstrate the linear perceptron form of functions such as "two at a time" it is sufficient first of all to look at the truth table of a simple example.

a	b	c	ab + bc + ca
0	0	0	0
0	0	0	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

By inspection, each pattern abc has its "inverse" pattern in the opposite category. By the inverse we mean the pattern formed by inverting each of pattern components. e.g. pattern 001 (category 0) has an inverse of 110 (category 1). The corresponding function produced by inverting all pattern components is $\overline{ab} + \overline{bc} + \overline{ca}$. Since the logical product of this function and the original is 0, this proves that component-inverse patterns always have opposite categories. The extension of this to any number of variables (e.g. three from four, four from five) is quite clear.

The linear perceptron form for ab + bc + ca is, again by inspection, $\lceil a+b+c > 1 \rceil$. The component-inverse of this is $\lceil 1-a+1-b+1-c > 1 \rceil$ or $\lceil a+b+c > 2 \rceil$. Since the value of a+b+c can increase or decrease by a minimum of 1, this is equivalent to $\lceil a+b+c \geq 1 \rceil$. Clearly, this is equivalent to the opposite category and so the component-inverse pattern is always in the opposite category.

The use of the three arithmetic values -1, ± 1, 1 is possible in certain cases, but is not practicable overall.

e.g. consider the simple expression a + b; the perceptron form can be seen to be $\lceil a + b > -1 \rceil$ in the following table.

a	b	a + b	$\lceil a + b \rceil$	$\lceil a + b > -1 \rceil$
-1	-1	-1	-2	No
-1	±1	±1	-2 or 0	?
-1	1	1	0	Yes
±1	-1	±1	-2 or 0	?
±1	±1	±1	-2, 0 or 2	?
±1	1	1	2 or 0	Yes
1	-1	1	0	Yes
1	±1	1	2 or 0	Yes
1	1	1	2	Yes

Where the value ±1 yields an uncertain value of a + b, one test against a threshold of -1 is insufficient to resolve the uncertainty. However, in general this sort of procedure is not possible and the substitution trick demonstrates with Boolean algebra is not possible either.

LINEAR SEPARABILITY AND PROGRAM COIN

This chapter contains a description of the computer program which provides the model which interprets the black box theory presented in an earlier chapter. The program is simply a version of the well-known perceptron training scheme* in several different forms. Program COIN (COmputerised INvestment) will be described later, along with the attempts at "perceptron steering", a concept to be defined with the description.

Before that, however, we must discuss the reasoning behind the "steering" attempts and its connection with the problem of linear separability. In the last chapter we saw that, in general, a complete pattern set, especially those with a large number of components, is not linear separable. As a reminder, here is the simplest example - four patterns each of two components:

a	b	outcome	
0	0	0	logical-net
0	1	1	expression
1	0	1	$\bar{a}b + a\bar{b}$
1	1	0	

Patterns with ten components, such as those from the investment data, are likely to be similarly non-separable, especially if there is a large number of them. This is a common problem when the perceptron is used as a classification mechanism. Minsky and Papert describe a number of alternative techniques which circumnavigate this barrier but they will not be discussed here because the perceptron is to be presented not to classify sets of patterns, but to carry this one stage further and to classify patterns not yet processed by the perceptron. This does not seem to have been discussed elsewhere - Minsky and Papert only mention it in passing as a possibility. If there are patterns which the perceptron has not seen, then it follows that the perceptron was trained only to classify a subset of

*The ~~calculation of the~~ weights α_i and the threshold θ - see definition of perceptron in the last chapter.

the total pattern set. This means that the linear inequality was calculated purely using the patterns in this subset; the coefficients and the threshold (the α_i and θ respectively) are only capable of separating these patterns. They may, (as Minsky and Papert say), be capable also of correctly classifying other patterns, and this is our concern here.

Since it is of great relevance, we repeat that Nilsson, amongst others, shows how a set of patterns with N components can be separated as long as there are fewer than N + 1 patterns in the set. If there are more, then the set might still be separable, but this is not certain - it depends upon the actual form of the patterns themselves.* It would be extremely useful if we had an analytic technique which would tell us when a given set of patterns is separable, and when not. One such technique suggests itself from the discussion of "pattern-integers" used in the last chapter to justify the derivation of a set of logical expressions with linear perceptron forms. An analytical method is described below which maps any set of pattern-integers (if possible) onto an ordered set which, as we demonstrated, has an obvious linear perceptron form.

To illustrate the technique we shall consider a simple example of four patterns, each having four components. The components can take the values 1, 2, 3 (as in the comparison array introduced earlier).

Pattern	a	b	c	d
1	3	1	1	3
2	1	2	3	3
3	3	1	3	3
4	2	2	3	1

Let us assume that patterns 3 and 4 are in category "A" and patterns 1 and 2 in category "not-A". If it is possible to perform the same isomorphic transformation on all four patterns such that the resulting pattern-integers are in "natural

*There are certain circumstances when even fewer than N + 1 patterns cannot be separated. However, these are very unlikely - see Nilsson.

order" then the pattern set is definitely linearly separable. Note that we have not assumed separability, our aim is to test for it. The procedure is as follows: Compare every pattern in category "A" with every pattern in category "not-A" and form an expression which describes whether each pair of pattern components is correctly ordered or not, relative to the classification scheme. Here we assume that the pattern-integers in category "A" should be greater than the pattern-integer in category "not-A". The expressions look like Boolean expressions but this is a mere illusion. The notation was chosen in this way as a convenient shorthand but also because the plus sign has the right Boolean connotation of "OR" and the negation sign also provides the right idea. Thus:

Compare pattern 1 with pattern 3. These are, again:

1:	3	1	1	3	"not-A"
3:	3	1	3	3	"A"

Only component c has the correct orientation. The other three components neither have nor have not - they are irrelevant and are thus ignored.

The expression for this comparison is thus, simply: (c). It could be interpreted as "these two patterns can be ordered correctly (in the pattern-integer sense) by c alone".

Now take patterns 1 and 4.

1:	3	1	1	3	"not-A"
4:	2	2	3	1	"A"

Components b and c have the correct orientation. Components a and d have the opposite orientation. Thus the expression for this comparison is $(\bar{a}+b+c+\bar{d})$. The interpretation is "these patterns can be ordered by b or by c or by the complements of a or d". Note that the inverse orientation is formed by taking the complement of the pattern component in the base of the pattern-integers.

Here the base is 3 and the complement is formed by taking the difference from 4. Thus 1 becomes 3, 3 becomes 1 and 2 remains as it is. In binary and ternary representations the operation of complementing can be seen to parallel the logical operation of negation - hence the choice of Boolean notation.

The other two comparisons give expressions thus:

patterns 2 and 3 : $(a+\bar{b})$

patterns 2 and 4 : $(a+\bar{d})$

The complete analysis can thus be written

$$(c)\ (\bar{a} + b + c + \bar{d})\ (a + \bar{b})\ (a + \bar{d})$$

No interpretation of this in Boolean terms is possible apart from the obvious fact that the individual terms in the brackets are to be combined to produce an overall expression relating to all four comparisons taken together.

The procedure continues as follows: scan the complete expression for variables which appear either <u>only</u> without negation or <u>only</u> with negation. In the example variables c and $\bar{d}$ result from this operation; a and b appear as $\bar{a}$ and $\bar{b}$ also and are rejected: write down each of these "un-paired" variables and <u>follow</u> them by <u>all</u> the individual terms which do <u>not</u> contain the chosen variables. Join all such derived expressions by a plus sign.

Thus in the example, take c and follow it by $(a + \bar{b})$ and $(a + \bar{d})$ $(\bar{a} + b + c + \bar{d})$ already contains c, as does the expression (c). Take $\bar{d}$ and follow it by (c) and $(a + \bar{b})$ - the other two contain $\bar{d}$ already.

Thus we derive:

$$\underline{c}\ (a + \bar{b})\ (a + \bar{d}) + \underline{\bar{d}}\ (c)\ (a + \bar{b})$$

The underlining indicates that this variable <u>must</u> come before any others.

It can now be seen that we are now left with two sub-analyses and we are required to reduce the expressions $(a+\bar{b})\,(a+\bar{d})$ and $(c)\,(a+\bar{b})$ separately.

$(a+\bar{b})\,(a+\bar{d})$ reduces to $(\underline{a}+\underline{\bar{b}a}+\underline{\bar{b}\bar{d}}+\underline{\bar{d}a}+\underline{\bar{d}\bar{b}})$, where $\underline{\bar{b}a}$ means $\bar{b}$ <u>followed</u> by a and similarly for the other pairs; combined with the $\underline{c}$ already derived we have:

$$(\underline{ca}+\underline{c\bar{b}a}+\underline{c\bar{b}\bar{d}}+\underline{c\bar{d}a}+\underline{c\bar{d}\bar{b}})$$

In a similar fashion $(c)\,(a+\bar{b})$ gives $(\underline{ca}+\underline{c\bar{b}}+\underline{\bar{b}c}+\underline{ac})$;combined with $\underline{\bar{d}}$ this gives:

$$(\underline{\bar{d}ca}+\underline{\bar{d}c\bar{b}}+\underline{\bar{d}\bar{b}c}+\underline{\bar{d}ac})$$

The final completely derived expression is the combination of these two:

$$(\underline{ca}+\underline{c\bar{b}a}+\underline{c\bar{b}\bar{d}}+\underline{c\bar{d}a}+\underline{c\bar{d}\bar{b}}+\underline{\bar{d}ca}+\underline{\bar{d}c\bar{b}}+\underline{\bar{d}\bar{b}c}+\underline{\bar{d}ac})$$

It can now be seen that each expression completely underlined provides a true <u>alternative</u> transformation to map the pattern-integers as they stand onto a correctly ordered set.

For example, take $\underline{ca}$. This says "take component c and follow it by component a". If we do this we get the following pattern set:

	a	b	c	d		c	a	conversion to denary
1	3	1	1	3	$\longrightarrow$	1	3	6
2	1	2	3	3	$\longrightarrow$	3	1	10
3	3	1	3	3	$\longrightarrow$	3	3	12
4	2	2	3	1	$\longrightarrow$	3	2	11

The derived pattern-integers are now correctly ordered - i.e. the integers for patterns 1 and 2 (category "A") are <u>both</u> less than both the integers for patterns 3 and 4 (category "not-A").

Similarly, consider $\underline{\bar{d}\,\bar{b}\,c}$

	a	b	c	d		$\bar{d}$	$\bar{b}$	c	denary
1	3	1	1	3	$\longrightarrow$	1	3	1	19
2	1	2	3	3	$\longrightarrow$	1	2	3	18
3	3	1	3	3	$\longrightarrow$	1	3	3	21
4	2	2	3	1	$\longrightarrow$	3	2	3	36

The justification for this simple procedure is equally simple and relies on the concept of the pattern-integer. The operations of ordering components and forming the complement are one-to-one transformations i.e. they are completely reversible. This means that the act of transforming the patterns into ordered pattern-integers does not destroy the correct categorisation of the original patterns. Furthermore, a well-known property of numerical representations, means that after one component has been dealt with, the others can then be considered independently of the one which has just been processed. Thus, if two numbers are correctly ordered by their first digit, the values of the second and subsequent digits do not destroy the ordering of the numbers. This simply says that the minimum value of the digit in column n represents a greater value than the value of the digit in column n + 1 (numbering from the left). The operation of choosing a variable which is not paired with its negation in any comparison, ensures that the comparisons which contain this variable are satisfied whatever variables follow (because of the property of numbers mentioned above) - the remaining comparisons are satisfied independently of the ones already dealt with.

Clearly each derived ordering has a corresponding linear perceptron form, similar to those mentioned in the last chapter. For $\underline{ca}$ the corresponding perceptron is:

$$\lceil 3c + a > 10 \rceil$$

For $\bar{d}\bar{b}c$ there are two possibilities because of the gap in pattern-integers between the two categories.

Thus we have either:

$$\lceil 9\,(4 - d) + 3(4 - b) + c \;>\; 19 \rceil \qquad \text{or}$$

$$\lceil 9\,(4 - d) + 3(4 - b) + c \;>\; 20 \rceil$$

All three perceptrons are equally valid in that they correctly classify the four patterns. In fact, since there are nine different orderings derived in the analysis, there are of the order of twenty completely different perceptron forms, and all are equally valid. Before discussing this further, it is necessary to examine the analytical procedure a little more closely. Consider the following set of patterns:

	a	b	c	category
1	1	2	1	not - A
2	3	1	3	not - A
3	2	1	3	A
4	1	3	2	A

An analysis following the lines given above yields the following expressions:

1 and 3 gives $(a + \bar{b} + c)$

1 and 4 gives $(b + c)$

2 and 3 gives $(\bar{a})$

2 and 4 gives $(\bar{a} + b + \bar{c})$

Since there are no "unpaired" variables, the analysis breaks down and produces no result. In other words the patterns <u>cannot</u> be mapped onto a set of pattern-integers using the transformations utilised above. However, the above pattern set <u>is</u> separable using the perceptron:

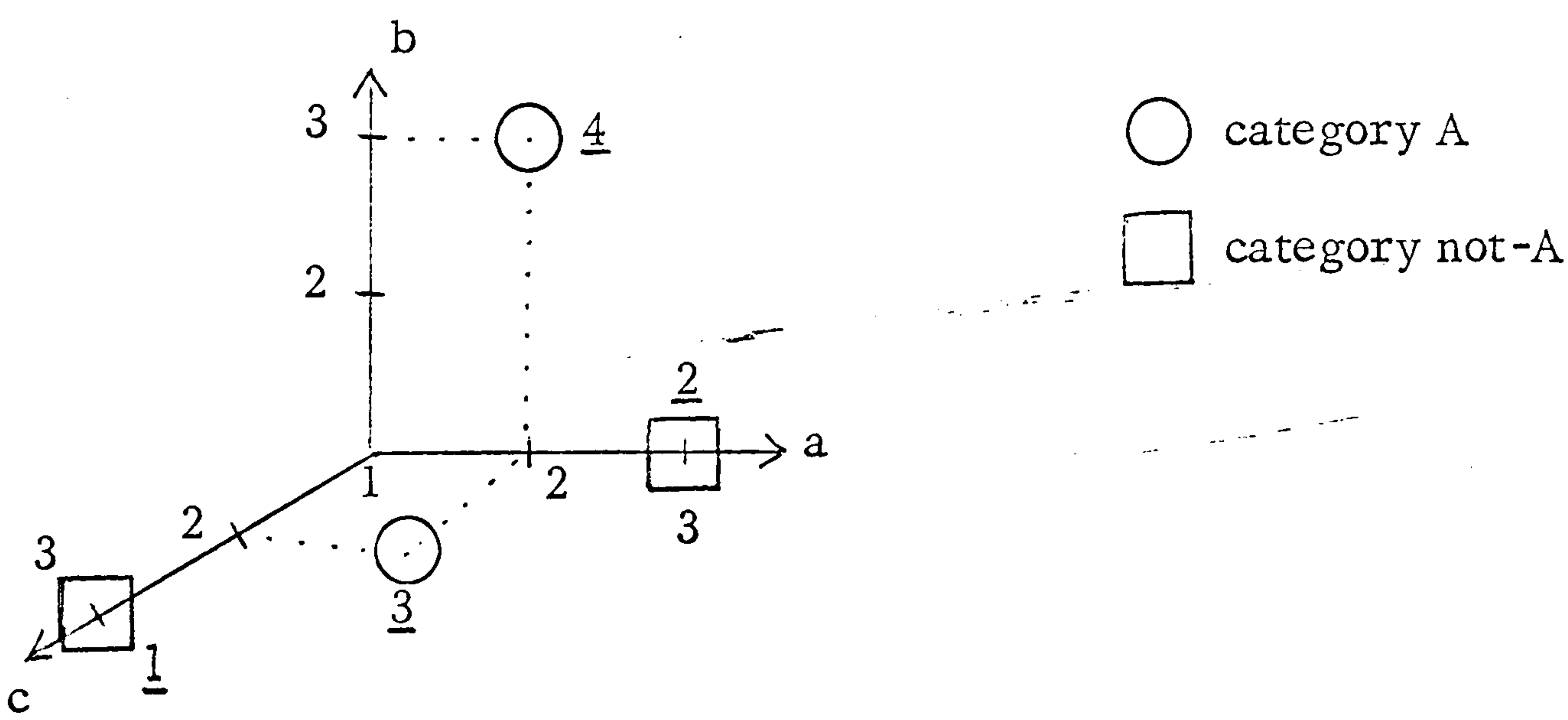
$$\lceil c - a > 0 \rceil$$

It seems that the analysis is not powerful enough. It errs in not telling us that a pattern set is separable, when in fact it is. This is however, preferable to concluding that a pattern set is separable, when in fact it is not. The analytic procedure, since it relies on the ordering of pattern-integers, could, fortunately, never do this.

To check whether it correctly identifies a set as non-separable consider the following simple set of four patterns:

	a	b	c	category
1	1	1	3	not - A
2	3	1	1	not - A
3	2	1	2	A
4	2	3	1	A

A graphical representation of these will convince that they are indeed non-separable. Take a, b, c as three orthogonal axes; the patterns can then be represented as points in this space.*



A linear function (as the perceptron form is) in 3-space is, of course, a plane. By inspection, no plane can divide the space into two regions with patterns 1 and 2 on one side and patterns 3 and 4 on the other side. The reason for this is that all four patterns are co-planar. This is an example of the condition for in-separability given in Nilsson's proof which was referred to before.

*This is well-known procedure, providing numerous insights from a useful visual repres~~entation - See Nilsson~~

The expression derived from the analysis is:

$$(a + \bar{c})\ (\bar{a} + b)\ (\bar{a} + c)\ (a + b + \bar{c})$$

From this we derive $\underline{b}\ (a + \bar{c})\ (\bar{a} + c)$ which gives no result since both a and c are "paired". Thus analysis predicts the non-separability which we can see by inspection.

Although this is far from complete, the examples given above indicate that our method of analysis is only in error sometimes when a pattern set is separable but analysis says not. Apart from that it is a good test of separability and also can directly provide a perceptron form. To conclude this section, it is of great importance to see the analysis at work on the investment data.

We shall analyse the data from the shares of Fosters and Reeds. The comparison array for these two shares is produced in exactly the same way as shown in the chapter on the formalisation of the theory. Rearranged into categories, the patterns are as follows:

VARIABLE ⟶ YEAR ↓	A	B	C	D	E	F	G	H	I	J	S
69	1	1	1	1	3	3	1	1	1	1	1
66	2	3	3	3	3	3	1	2	1	3	1
64	2	1	1	1	1	3	1	2	1	1	1
63	2	3	3	3	1	1	1	1	3	3	1
68	3	3	3	3	3	3	2	3	3	3	3
67	3	1	1	1	3	1	1	1	3	1	3
65	2	1	1	1	3	1	3	3	1	1	3
62	2	3	3	3	3	1	1	3	1	3	3

The analysis expressions are as follows:

$$(A + B + C + D + G + H + I + J)$$
$$(A + \overline{F} + I)$$
$$(A + \overline{F} + G + H)$$
$$(A + B + C + D + \overline{F} + H + J)$$
$$(A + G + I)$$
$$(A + \overline{B} + \overline{C} + \overline{D} + \overline{F} + \overline{H} + \overline{I} + \overline{J})$$
$$(\overline{B} + \overline{C} + \overline{D} + \overline{F} + \overline{G} + \overline{H} + \overline{J})$$
$$(\overline{F} + H)$$
$$(A + E + \overline{F} + \overline{H} + I)$$
$$(A + B + C + D + E + G + I + J)$$
$$(E + \overline{F} + G + H)$$
$$(B + C + D + E + \overline{F} + H + J)$$
$$(A + E + F + G + H)$$
$$(A + \overline{B} + \overline{C} + \overline{D} + E + \overline{J})$$
$$(\overline{B} + \overline{C} + \overline{D} + E + \overline{G} + H + \overline{I} + \overline{J})$$
$$(E + H + \overline{I})$$

The reduction of these expressions give the following list of possible orderings:

$\underline{A\ E\ \overline{F}}$

$\underline{A\ E\ H}$

$\underline{A\ E\ (\overline{B} + \overline{C} + \overline{D} + G + \overline{J})\ \ (\overline{F} + H)}$

$\underline{A\ \overline{F}\ (\overline{B} + \overline{C} + \overline{D} + E + G + H + I + J)}$

$\underline{A\ G\ \overline{F}\ (E + H + \overline{I})}$

$\underline{A\ G\ H}$

$\underline{A\ G\ \ (B + C + D + E + J)\ \ (\overline{F} + H)}$

$\underline{A\ H}$

$\underline{A\ \overline{I}\ \overline{F}}$

$\underline{A\ \overline{I}\ G\ \overline{F}}$

$\underline{A\ \overline{I}\ G\ H}$

$\underline{A\ \overline{I}\ G\ (B + C + D + E + J)\ \ (\overline{F} + H)}$

$\underline{A\ \overline{I}\ H}$

$\underline{G\ A\ \overline{F}\ (E + H + \overline{I})}$

$\underline{G\ A\ H}$

$\underline{G\ A\ (B + C + D + J)\ \ (\overline{F})\ \ (H + E + \overline{I})}$

List cont'd..

$\underline{A \bar{I} E \bar{F}}$

$\underline{A \bar{I} E H}$

$\underline{A \bar{I} E (\bar{B}+\bar{C}+\bar{D}+G+\bar{J})\ (\bar{F}+H)}$

$\underline{G E A (\bar{F}+H)}$

$\underline{G E F}$

$\underline{G \bar{F} (A+\bar{B}+\bar{C}+\bar{D}+\bar{J})\ (F+H+\bar{I})}$

$\underline{G \bar{F} E}$

$\underline{G \bar{F}\ (\bar{I}+H+E)\ (A+\bar{B}+\bar{C}+\bar{D}+\bar{J})}$

$\underline{G \bar{F}\ (B+C+D+H+J)\ (A+G+I)}$

These give a total of no less than ninety two possible orderings, each of which has an independent perceptron form, each of which is entirely valid, in that it correctly separates the pattern set! When the perceptron is used for prediction, the latest data set (the figures for 1970) are used. In this example, the "new" pattern (the one not included in the perceptron's training set of patterns) is:

VARIABLE → / YEAR ↓	A	B	C	D	E	F	G	H	I	J	S
70	3	1	1	1	1	3	1	3	1	1	?

The problem is now one of which perceptron to choose to produce the prediction variable S. With ninety two to choose from, this is no easy matter, for different perceptrons produce different predictions. To see this, let us choose, at random, four different orderings from those above and derive the value of the variable S i.e. discover which category the pattern for 1970 fits into. Let us therefore choose:

1. $\underline{A G \bar{F} \bar{I}}$ 2. $\underline{A H}$ 3. $\underline{G \bar{F} \bar{B} \bar{I}}$ 4. $\underline{G E \bar{F}}$

The transformed patterns produced by these three are as follows:

VARIABLES → / YEAR ↓	$A G \bar{F} \bar{I}$	$A H$	$G \bar{F} \bar{B} \bar{I}$	$G E \bar{F}$
69	1 1 1 3	1 1	1 1 3 3	1 3 1
66	2 1 1 3	2 2	1 1 1 3	1 3 1
64	2 1 1 3	2 2	1 1 3 3	1 1 1

cont'd.....

63	2 1 3 1	2 1	1 3 1 1	1 1 3
68	3 3 1 1	3 2	3 1 1 1	3 3 1
67	3 1 3 1	3 1	1 3 3 1	1 3 3
65	2 3 3 3	2 3	3 3 3 3	3 3 3
62	2 1 3 3	2 3	1 3 1 3	1 3 3

The corresponding "prediction" patterns are:

1. 3 1 1 3   2. 33   3. 1 1 3 3   4. 1 1 1

By inspection, the correct categorisations for these patterns are:

1. below   2. below   3. above   4. above

where "above" and "below" refer to the horizontal line drawn in the "transform" table above. It might have happened that the "prediction" patterns fitted in the gap between the two sets of pattern-integers, when no prediction would be possible using this method (e.g. 2132 with the ordering AG$\bar{F}\bar{I}$). However, this is very unlikely; a prediction can be produced for the great majority of new patterns.

We have here an apparent stumbling-block - a definite possibility of inconsistency in predictions depending on rather arbitrary circumstances. In other words, each ordering produced by analysis and its associated perceptron, as mentioned before, is <u>equally</u> valid - there is no <u>inherent</u> way of choosing the "best" one, even if such a thing exists. The inconsistency referred to above is certainly borne out by the results, as we shall see in the relevant chapter.

As a way out of this apparent dead-end, we need some way of introducing consistency into the perceptron training scheme which was mentioned at the beginning of the chapter. Thus the concept of <u>steering</u> was introduced, whereby the progress of the perceptron could be directed along pre-determined lines. To describe this concept, and to show what is meant by "training" a perceptron, we now turn to Program COIN.

Conceptually, a practical perceptron is a machine which "looks at" patterns which are presented to it, along with their correct classification, and calculates the weights α_i and the threshold θ such that it could produce the correct classifications without being told a second time, if it is presented with the same patterns again. The way it produces the weights is basically by a trial-and-error process using a trial set of weights. In other words, the patterns, with their classifications, are presented one at a time; the perceptron forms its linear function using the present pattern and the present set of weights, the α_i. The value thus derived is tested against the threshold; if the test is true then the next pattern is presented; if the test is false then the weights are changed according to some pre-set method. In this way, the perceptron cycles through the set of input patterns, called the training set, until the threshold test is true for every pattern in the set. The similarity of this scheme to a paradigm of a learning situation (such as teaching a child to recognise words), led to the description of the perceptron as a learning machine - hence the title of Nilsson's book. This aspect will not be discussed at all here, since Minsky and Papert have had something like the last word in this direction.

Instead, after a few relevant remarks about the nature of the perceptron's iterative training scheme, we shall describe the process in a little more detail. The most important point relates to the so-called perceptron convergence theorem. The trial-and-error scheme outlined above is clearly, a priori, subject to possible failure for a number of reasons. The crucial part of the process is the point at which the weights are changed. Unless this is done in a reasonable manner, as Minsky and Papert put it, the weights could be subject to rather odd, even disastrous behaviour, resulting in the perceptron not being able to solve the problem, even when the pattern set is known to be linearly separable. The convergence theorem basically says that if one particular weight-changing method, or a close relative of it, is used, then the perceptron will always separate the pattern set, i.e. satisfy all the inequalities for all the patterns, in a finite time. The proviso is that the pattern set is known to be separable in the first place. The particular weight-changing method can be described thus:

Assume each pattern is a set of numbers $\{X_i\}$

Assume the current set of weights is $\{\alpha_i\}$ and the threshold is zero*

If $\sum_{i=1}^{n+1} \alpha_i X_i > 0$ and it should be < 0 then put

$$\alpha_i' = \alpha_i - X_i$$

Conversely, if $\sum_{i=1}^{n+1} \alpha_i X_i < 0$ and it should be > 0 then put

$$\alpha_i' = \alpha_i + X_i$$

On this basis, the convergence theorem can be proved. It could be said that we are adding and subtracting the pattern components from the weights. There are minor variations on this, the main one being to add or subtract some multiple of the pattern components. i.e.

$$\alpha_i' = \alpha_i + kX_i$$

In another variation, k is calculated so that the inequality is just reversed i.e. the weights are changed so that the new weights correctly classify the current pattern, if its classification was incorrect with the old set of weights.

We shall not discuss the convergence theorem or the training schemes in any great detail. They can be found to be discussed and analysed in both of the books frequently mentioned in the last few chapters - Nilsson and Minsky and Papert. However, we shall remark on the general anonymity of these training schemes. What is meant by this is that they seem entirely arbitrary - we cannot say that the perceptron appears to be aware of what it is doing, where it is going. It seems to us, that in what is essentially a search procedure, there should be some direction of the search to a specific end - apart that is from the overall aim of correctly classifying the patterns in the training set. It is not our wish to digress here into a discussion of heuristic search procedures but the concept of steering has much the same object. The various

*If the perceptron is written $\sum_{i=1}^{n} \alpha_i X_i > \theta$ this is equivalent to $\sum_{i=1}^{n} \alpha_i X_i - \theta > 0$ or $\sum_{i=1}^{n+1} \alpha_i X_i > 0$ where $\alpha_{n+1} X_{n+1} = \theta$. By making X_{n+1} constant we can remove the need for θ.

versions of Program COIN all have some aim other than merely separating the training set of patterns. As we have seen above, different perceptrons (i.e. different sets of valid weights) may produce different predictions from the same "prediction" pattern. The overall aim of Program COIN is to "ring the changes", so to speak; to try to extract different predictions for the same pair of shares, wherever possible. The hope was that, after all the comparisons which cannot be decided upon because of differing predictions were accounted for, the remaining comparisons would yield a valid set of predictions. The ways in which the perceptron was "steered", the results of these attempts and the evaluation of the results are the subject of the next chapter.

RESULTS: ANALYSIS AND EVALUATION

Although the results presented in this chapter were produced mainly by various computer versions of the linear perceptron, it is hoped they will be viewed in the light of the several analytic techniques presented in the preceding chapters. The emphasis has been on the analytic side rather than on plunging directly into massive experimentation, for several reasons. Much of the work we have come across has concentrated almost exclusively on the direct evaluation of experimental results; this is only right when testing a predictive theory, but it seems that in some cases this has been done at the expense of an insightful discussion of the techniques used (especially where such techniques are essentially subjective as in much of technical analysis). This is less true of the numerous statistical studies, where the methodology has a firm foundation in probability theory, but even here there has been some optimism in the assessing of results. Spurious correlations are frequent in such statistical work and careful discussion and analysis is needed in all cases.

With this sort of consideration in mind, the results are here presented as bald facts and criticised to the utmost limits. Only then can we assess whether the theory has been confirmed to any extent, and in any case, heavy criticism and analysis of results can lead to improvements in the construction of the theory. The central aim of the experimentation was to check that the "sign-posts" set up by the analytic techniques point in the right direction. In general it might be said that they point down the road of pessimism, especially from the demonstration of a multiplicity of valid perceptron forms. At the very least the "sign-posts" point toward caution.

It is clearly essential when testing a predictive theory to know what actually happened after the prediction was made. Although this was all done in the past, i.e. the actual outcome was already known, the procedure is no less a valid one. (In fact it closely resembles the perceptron training scheme when viewed overall.)

Since data was taken from the years 1962 to 1970, the predictions were for the end of the financial year in 1971. Graphs of the share-prices for all ten shares will be found in the appendix to this chapter. Also shown are the graphs of the ratios of the price for one year to the price for the previous year. The shares are numbered 0 to 9 for the purpose of presenting lists of results in an easily digestible form. The correspondence of these numbers and the shares' names are as follows:

0	Reed (Austin)
1	Foster Brothers
2	H. Samuels
3	Debenhams
4	Army & Navy
5	Boots
6	Currys
7	Marks & Spencer
8	Dorothy Perkins
9	W.H.Smiths

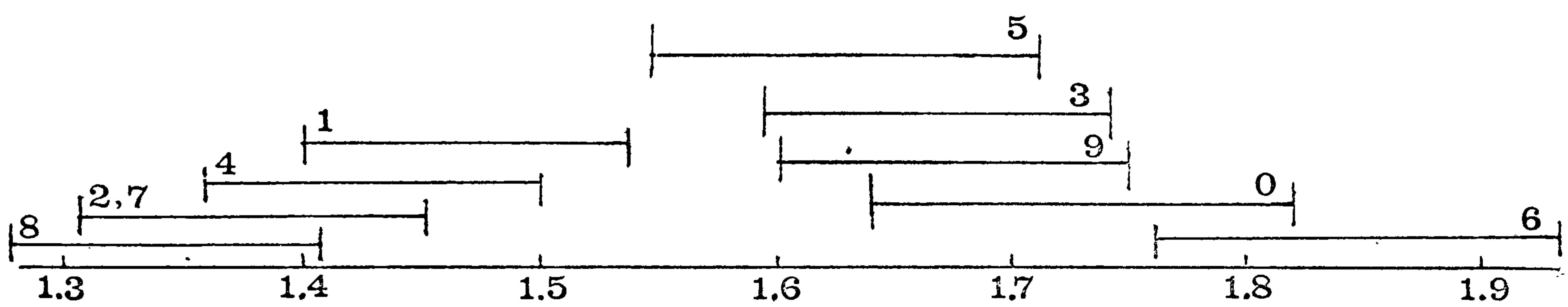
The ratios of the share-price for 1971 to the share-price for 1970 are as follows:

0	1	2	3	4	5	6	7	8	9
1.73	1.47	1.38	1.67	1.43	1.63	1.85	1.38	1.34	1.68

Since we are not at all concerned with absolute values these ratios will be ranked according to size and the ranked order will form the basis of the experimental testing. (It is interesting to note that all the shares chosen gained not less than one third in value during the year 70-71. The implications of this sort of "parallel" behaviour will be discussed later.) Thus the shares can be ranked in order of their growth performance during 70-71 :

$$6 \rightarrow 0 \rightarrow 9 \rightarrow 3 \rightarrow 5 \rightarrow 1 \rightarrow 4 \rightarrow 2 = 7 \rightarrow 8$$

The arrows indicate "better than", the equals sign indicates equal performance. This then is the order which it is the object of the theory to predict when applied to those shares contained in the ranking. This ranking implies forty five separate orderings of pairs of shares. The combination of the results of these forty five predictions should give an overall ranking as the actual one given above. However, there are two problems connected with such a procedure. The first, and the more general one, concerns the accuracy of forming a ranking like the one above. Several of the ratios, the ordering of which are being predicted, are rather too close to be absolutely sure of the individual ordering of a pair. In particular the ratios for shares 2 and 7 are identical to two decimal places. Even a slight error in one or other of the ratios could change round the ranking. If we look at the diagram below, we can judge how bad this problem is:



The lines thus : |———| give the limits of the ratios for each share on a ± 5% basis. This is an arbitrary error on the ratio, which implies something like a $\pm 2\frac{1}{2}\%$ error on the share-prices which produced the ratio. The source of such errors in the share-prices is rather obscure - certainly it is not an error of measurement, since those were the actual share-prices as recorded by the stock exchange. Instead we have to look at possible random fluctuations in the share-price around the time of measurement. This could be likened to trying to read a pointer on a dial which is fluctuating over a fairly wide range. Share-prices do not fluctuate so rapidly, but a week can bring about an unaccountable change, and our estimate of $2\frac{1}{2}\%$ for such a random change is not unreasonable. The justification for this comes from the many statistical studies based on the Random Walk theory which states that share-price behaviour is essentially random i.e. share-prices cannot be predicted on the basis of past price behaviour. (This does not mean, as Granger and Morgenstern and many others say, that share-prices are not predictable by other means. The findings of these workers is here accepted fully - our aim is to investigate these "other means".) If the $\pm 5\%$ error limits are acceptable it means that there may also be errors, of a more drastic nature, in the ordering of the set of ratios. As we can see from the diagram, the shares seem to split neatly into two groups, shares 6, 0, 9, 3, 5 in the "better" group and shares 8, 2, 7, 4, 1 in the "not so good" group. In other words, all the shares in the first group performed better than all the shares in the second group, even allowing for errors. Within the groups we can only say that 1 performed better than 8 in the second group and that 6 performed better than 9, 3 and 5 in the first.

Thus because of possible errors, the original forty five individual orderings are reduced to twenty nine, the remaining sixteen being, on the basis of the error estimate, undecidable*.

The results of Program COIN will thus be tested against these twenty nine and not the original forty five. They are:

$$(5,\ 3,\ 9,\ 0,\ 6) \quad > \quad (1,\ 4,\ 2,\ 7,\ 8)$$

$$1 \quad > \quad 8$$

$$6 \quad > \quad (5,\ 3,\ 9)$$

where the $>$ sign stands for "performed better than".

Before discussing how Program COIN produced its results it is worth mentioning a few facts which can be gleaned by a fairly detailed visual scanning of the comparison arrays for all forty five comparisons. These are shown in the appendix. Because the order of presentation of the variables, across the page, patterns of regularity in the ones, twos and threes emerge (there may be others, but they are not apparent visually). Chiefly this concerns the second, third and fourth variables which are, respectively, total profits, earned income

* A reduction in the assumed error to $\pm 3\%$ produces no change.

for ordinary shares, and earned income. By definition these variables are closely related and this fact is borne out by the way the patterns emerge. Often no information is gained by having all three variables since, for a particular comparison, all three have the same value for all ten years. e. g. the comparison of shares 1 and 4. Apart from that, one gets the impression that the distribution of ones, twos and threes in the arrays is more or less a random one. There are no obvious correlations with the pattern of behaviour of the share-price ratios and no smooth changing of orientations over the years. One would certainly not expect simple correlations with the relative share-price ratios, since no such correlations have been uncovered by much deeper statistical studies than this. One might, a priori, expect there to be evidence of smoothly changing relative variables, but this does not appear to be so. This seems to indicate that the performance of one company relative to another is erratic and, certainly in the shares chosen, no one company maintains superiority for very long. One final remark on the comparison arrays is that because of the lack of regularity (except for variables two, three and four), the chosen variables seem to be more or less <u>independent</u> of each other. Clearly in <u>any</u> sort of investment work, the choice of independent variables will lessen the work involved in company analysis and also help to lessen misunderstandings about the relevance of particular variables. Redundancy in investment data is very common (perhaps it gives a sense of security) but is not really justified on analytical grounds.

PROGRAM COIN

We discussed in the last chapter how and why the perceptron, when presented with a subset of all possible patterns, can produce different predictions using the same "prediction" pattern which it has not "seen" before. It could happen that the "prediction" pattern is identical with one in the training set, but with such a small set of patterns relative to the possible total, this is extremely unlikely. (It would be likely if there was much regularity in the patterns, but, as we saw above, this is not the case). To get the perceptron to produce these differing results, it is necessary to change its training scheme. This is the concept of "steering" as presented in the last chapter.

The training scheme can be altered in many ways, all of which were tried with varying degrees of success - measured by successful separation of the training set of patterns within a fixed number of cycles through the set. Remember that unless there are very good reasons, the training should conclude successfully in a finite time for all comparisons. This is because, as mentioned before, there are fewer patterns in the training set than the number of components of each pattern.

There are three basic ways of steering a perceptron, and they will be described briefly in turn.

Method 1 The constant k in the weight-changing procedure can be changed. The weights are changed according to the scheme:

$\alpha_i = \alpha_i \pm k X_i$ depending whether the sum $\sum \alpha_i X_i$ is < 0 or > 0

(X_i are the pattern components)

The convergence theorem has been proved for values of k between 0 and $2 \sum \alpha_i X_i / \sum X_i^2$, but the training may be successful depending on the nature of the patterns, for values greater than this. (Clearly $k < 0$ will always take the sum $\sum \alpha_i X_i$ the wrong way and lead to wild oscillations).

Either k can be fixed beforehand, or it can be calculated to just correct the inequality. It is shown in Nilsson that the value of k is then $\sum \alpha_i X_i / \sum X_i^2$

Method 2 Again, the convergence theorem can be extended to cover the case where the weights are initially set to zero) has an infouence on the time (or number of cycles through the training set) taken to reach a satisfactory set of weights -

if the initial set of weights were the final set there would be no training at all. The results of analysing several comparisons, using the methods described in the last chapter, seemed to favour variables 1, sales; 6, net assets per share; and 9, capital issued as ordinary shares. These featured prominently in a number of such analyses, and so an attempt was made to "favour" these variables by starting them off with a non-zero weight of the order of the usual finishing weight. (about + 20).

Method 3. Since the perceptron does not do complex juggling acts with the weights when the need arises to change them, it often happens that a pattern which was previously satisfied with the current set of weights can go back into the pool of unsatisfied patterns when the weights are changed. Hence the need to cycle through the training set of patterns, again and again, until all the patterns have been satisfied. The order in which the patterns are considered by the perceptron is clearly of great importance. At least it is clear that changing this order will change the final weights, although it is not clear how different they will be. The "normal"order of the patterns was in reverse chronological order i.e. from 1970 to 1961. Since the pattern for each year is equivalent to every other pattern (no one is more important than another), this order is, in fact, an arbitrary one. Other arbitrary orders were tried as well as a"natural" ordering which suggests itself from the analytical method. A priori, it might be expected that if the patterns are ordered as to their classification (i.e. all patterns with relative share-price ratio equal to one first, followed by all those with the relative ratio equal to three), then a solution of weights might be found which was close to one of the solutions suggested by analysis. Why this does not happen is discussed below.

We have shown that there are many solution sets of weights available where the weights are powers of three. However, the perceptron, whatever training scheme is used, never finds one of these solution sets. The reason is to be found in the nature of training schemes themselves. Whenever the test on an inequality fails i.e. when the current set of weights classifies a pattern from the training set incorrectly, all the weights are changed since all the pattern components are non-zero. It might well be that merely increasing or decreasing one variable by a reasonable amount would rectify the misclassification, but the perceptron is oblivious to this fact. Only by distinguishing between the variables to some extent can this sort of procedure be carried out. The simple perceptron cannot do this and in any case, any attempt to stray beyond the confines of the convergence theorem is fraught with danger. A few thoughts on such modified training schemes reveal that an approach, which fits the concept of "steering" much better, along these lines would be profitable. One thought is that by merely including an extra pattern in the training set with zero components except for selected ones would ensure that the selected components would remain either positive or negative, whichever is desired. Thus, in our case a pattern such as (0 0 0 $\frac{1}{2}$ 0 0 0 0 0 0 0) with a classification of '3' would ensure that variable 4 is always greater than zero. Another thought is that by insisting on the weights being powers of three, choosing which weights to increase and which to decrease according to some pre-set criterion, the "analytic" solutions could be reached. There are many such procedures waiting to be tried out, and if the theory is sufficient to handle them, the humble perceptron could become a much more useful and meaningful weapon in the prediction armoury.

DETAILED RESULTS

Over many computer runs, the "steering" methods described above were used to operate Program COIN. Three individual comparisons are of great interest because whatever training scheme was used, no solution set of weights was found after a reasonable number (fifty) of cycles through the training set. In fact an examination of the successive sets of weights, after each change, showed that either the training sequence was in a loop from which situation the perceptron could never extricate itself, or the weight set was hovering around some average set and never quite finding complete satisfaction. These three will be discussed in some detail in what follows.

Shares 6 and 9

The comparison array for this pair is:

VARIABLES → YEAR ↓	A	B	C	D	E	F	G	H	I	J	K	S
67	1	1	1	1	2	1	3	3	1	1	1	1
66	3	3	3	3	3	3	1	2	1	3	1	1
62	2	3	3	3	3	1	3	2	1	3	1	1
69	1	3	3	3	3	1	3	2	1	3	1	3
68	1	1	1	1	1	1	1	1	1	1	1	3
65	3	1	1	1	1	1	1	2	1	1	1	3
64	3	3	3	3	3	1	3	3	1	3	1	3
63	2	3	3	3	3	1	3	2	1	3	1	3

[K is the constant variable which replaces θ, the threshold in the perceptron in-equality.]

Without analysis we can see that the patterns for 1962 and 1963 are identical, but they are in opposite categories. This is clearly impossible, both from a common-sense point of view and from a perceptron point of view, since no inequality can be satisfied both ways by the same number. The next case is also similar.

Shares 5 and 6

The comparison array is:

Table continued overleaf/.

VARIABLE → YEAR ↓	A	B	C	D	E	F	G	H	I	J	K	S
68	3	1	3	3	1	3	3	3	3	1	1	1
65	3	3	3	3	3	1	3	1	1	3	1	1
64	3	1	1	1	1	3	3	3	3	1	1	1
63	1	1	1	1	1	1	1	2	1	1	1	1
69	1	1	1	1	1	3	1	1	1	1	1	3
67	3	3	3	3	2	3	1	2	3	3	1	3
66	1	1	1	1	1	1	1	2	1	1	1	3
62	1	3	3	3	2	3	3	2	3	3	1	3

Here it is the patterns for 1963 and 1966 which are identical, but again they are in opposite categories.

Shares 2 and 5

This case is different, because all the patterns are different.

The comparison array is:

VARIABLE → YEAR ↓	A	B	C	D	E	F	G	J	I	J	K	S
68	3	1	1	1	3	1	1	2	3	1	1	1
67	1	3	3	3	2	1	1	2	3	3	1	1
63	3	3	3	3	3	3	3	1	3	3	1	1
62	3	1	1	1	2	3	3	2	1	1	1	1
69	3	1	1	1	1	1	3	3	3	1	1	3
66	3	1	1	1	3	3	1	2	1	3	1	3
65	3	2	3	3	3	3	3	3	3	1	1	3
64	3	3	3	3	3	1	3	1	1	3	1	3

Analysis reveals that only variable A can provide a partial ordering; after that no variable can be chosen. Thus certainly no solution is available based on the ordering procedure which we have used to analyse the perceptron. There is a similarity between this case and the example of a non-separable set given in the last chapter. In both cases, a partial ordering is possible but the analysis breaks down when taken further. It is highly likely that the set of patterns is degenerate, just as in the example quoted.

Shares 4 and 8

Comparison Array

VARIABLE / YEAR	A	B	C	D	E	F	G	H	I	J	K	S
69	1	1	1	1	2	3	3	2	3	1	1	1
68	3	3	3	3	3	3	1	2	3	3	1	1
67	1	1	1	1	2	3	1	2	1	1	1	3
66	3	3	3	3	2	3	3	2	3	1	1	3
65	3	3	3	1	3	3	1	1	3	3	1	3
64	3	1	3	3	3	3	1	3	3	1	1	3
63	3	3	3	3	3	3	1	2	3	3	1	3
62	3	1	1	1	2	3	3	1	3	1	1	3

As with the first two, this is a case of identical patterns with different categories. Here it is patterns for 1968 and 1963.

Shares 6 and 7

Comparison array:

Table continued overleaf/.

VARIABLE YEAR	A	B	C	D	E	F	G	H	I	J	K	S
	1	3	3	3	1	1	1	2	3	3	1	1
	1	1	1	1	2	1	3	3	1	1	1	1
	1	3	3	3	3	1	3	2	1	3	1	1
	3	1	1	1	3	1	3	3	1	1	1	1
	1	3	1	1	1	1	3	1	1	3	1	3
	1	1	1	1	1	1	1	3	1	1	1	3
	1	1	1	1	3	3	1	1	3	1	1	3
	3	3	3	3	3	1	3	2	3	3	1	3

This case is similar to that for shares 2 and 5. Variable F provides a partial ordering but after that there are no variables for further orderings. The same remarks apply to this comparison.

This completes the discussion of comparisons which the perceptron cannot handle. We think it is safe to say that such coincidences are rare in data where the variables are independent of each other and since there are only five such cases, it confirms our (and Moodie's) choice of variables.

Two other cases deserve consideration, the one for its relationship to perceptron analysis, the other for its relationship to the whole idea of the perceptron.

Shares 1 and 5

Comparison array:

Table continued overleaf

VARIABLE / YEAR	A	B	C	D	E	F	G	H	I	J	K	S
	3	1	1	1	3	1	3	3	3	1	1	1
	1	1	1	1	3	3	1	2	1	3	1	1
	3	3	3	3	3	3	3	2	1	1	1	1
	3	3	3	3	1	1	3	1	1	3	1	1
	1	3	3	3	2	3	3	2	1	3	1	3
	3	1	1	1	3	3	1	2	3	1	1	3
	3	3	3	3	3	3	1	3	3	1	1	3
	3	1	1	1	3	1	3	2	3	3	1	3

Analysis reveals that no single variable can provide an ordering hence no pattern-integer type of perceptron can provide a solution. However, by choosing an arbitrary order for cycling through the pattern set, the perceptron did find a solution set of weights. Thus we have a case similar to the example in the last chapter where analysis said that no solution was available, but an obvious solution could be found by inspection. It is clear that, as we have said before, the method of analysis is inadequate in certain circumstances. No reason for this anomaly has been found yet, although it is thought that there is a much closer tie-up between the algorithm for deriving the perceptron form from Boolean expressions, presented in the chapter on perceptron theory, and the analytical method. The answer to why analysis is inadequate in these cases, and also greater insight into the actual weights derived by a practical perceptron, is believed to lie in this tie-up.

The other comparison to look at in detail is shares 3 and 7. Here share 3 performed worse than share 7 throughout all the years for which data was taken. Hence there is no training to be carried out and, if the initial set of weights is correct, the prediction must be the same. i.e. 3 worse than 7. However, the actual performance, as can be seen from the graphs, was the reverse of this - company 3 picked up and its share-price out-performed that of company 7. A prediction to match this could have been made, but only by choosing the initial weights in the correct way - rather like looking at the answer and then solving the problem. The perceptron, being essentially a pattern classifier, needs a fair number of patterns of each category to perform its job; when all the patterns presented are of the same category, it cannot help but conclude that all patterns are like this. This is another example of how the perceptron lacks "intelligence" and highlight the need for rigorous analysis.

OVERALL RESULTS

The combined results of all computer runs are summarised in the table below. A "greater than" sign means that the share on the top was predicted to out-perform the share on the left-hand side. A "less than" sign means the reverse of this. The star means that either no prediction was obtained, as in the cases discussed above, or that both predictions i.e. "greater than" and "less than" were obtained in different runs. These are the cases where the "steering" of the perceptron has produced the desired effect.

	0	1	2	3	4	5	6	7	8	9
0		*	<	*	*	*	<	*	<	<
1	*		<	<	<	<	*	>	*	*
2	>	>		<	*	*	<	<	*	>
3	*	>	>		*	>	>	>	*	*
4	*	>	*	*		>	*	<	*	<
5	*	>	*	<	<		*	>	*	*
6	>	*	>	<	*	*		*	<	*
7	*	<	>	<	>	<	*		*	*
8	>	*	*	*	*	*	>	*		*
9	>	*	<	*	>	*	*	*	*	

In the table, the forty-five comparisons are duplicated for ease of analysis.

The first statistic of importance is that of the forty five comparisons twenty one are predictions, twenty four produced no predictions or produced both. It may be that there are more "dual prediction" comparisons but further attempts at steering produced no further ones. If we refer back to the actual set of relative performances chosen to test against, we can see that there are six correct predictions, seven incorrect and sixteen undecided, making a total of twenty nine. It could be said that this is the major result from this work, and is frankly disappointing. It is disappointing not only in the low percentage of correct predictions (about 20%) but in the fairly low percentage of comparisons which produced predictions at all (about 45%). The conclusions and evaluation of this result will be left until later, but we shall now report the results of a short attempt at a less ambitious approach, but using all the techniques described above.

One of the major points of the theory as presented here is that comparisons are made between individual shares. This goes against the considered methods of both Clarkson and Weaver and Hall. Both of these studies base the comparisons (or suggest it in the case of Weaver and Hall) on evaluating shares, not relative to each other, but to some average; either of the "market as a whole", or of the shares under consideration (as is the case in Clarkson's simulation).

To this end, a fictitious share was created, here called "A" (for average) which has values which are averages of all the corresponding values of all ten shares under consideration.* The graphs relating to A are given in the appendix. The results of comparing all ten shares in turn, with A over a number of computer runs, using the same techniques of steering as already described, are as follows:

Table shown on following page:

*Unfortunately, variable A, sales, had to be dropped because some of the companies did not record sales until about 1966.

$0 > A$

$1 < A$

$2 > A$

$3 < A$

$4 < A$

$5 < A$

$6 * A$

$7 > A$

$8 * A$

$9 > A$

Of the two comparisons which gave no prediction, the comparison of shares 8 and A contains indentical patterns in each category, and so is undecidable; the comparison of 6 and A is of the type which gives nothing when subjected to analysis. However, unlike shares 1 and 5, discussed above, no amount of "steering" has yet produced a prediction.

Comparing each share with the average over all shares implies that a split is possible of the shares into two groups, those which are better than average, and those which are worse than average. If this is done, the two groups produced can be ordered thus:

$$(0\ 2\ 7\ 9) > (1\ 3\ 4\ 5)$$

Tested against the actual relative performances we see there are four correct predictions, four incorrect and twenty one which are undecided. If anything this is a worse result than the previous one, but this is partially a consequence of not testing against the correct set of comparisons. The average of the share-price ratios for 1970 is 1.56.. Relative to this average we get the following list:

Table shown on following page:

0 $>$ A

1 $<$ A

2 $<$ A

3 $>$ A

4 $<$ A

5 $>$ A

6 $>$ A

7 $<$ A

8 $<$ A

9 $>$ A

this implies the group ordering: (0 3 5 6 9)$>$(1 2 4 7 8).

Of these twenty five orderings, four are correct, four are incorrect and seventeen undecided. If the same procedure is adopted for reducing the number of actual orderings because of the possibility of errors in the ratios, the following group ordering is found: (0 6)$>$(2 7 8). Of these six orderings, only one is predicted correctly, two being incorrect and two undecided.

The results from the introduction of the "average" share are no better than those produced by comparing each share with every other one. We must now discuss why the results are poor, and suggest ways in which improvements could be made in theory and techniques. The implications for both theories of investment and investment theories will then be discussed and a critical comparison made of the results of similar work. Finally there is a conclusion widening out the scope and suggesting alternative ways forward.

EVALUATION OF RESULTS

In order to carry out an evaluation satisfactorily, we must return to the start and look closely at each chain in the link from initial idea to production of results.

The original idea of using linear weighted functions to choose shares was, we believe, due to Clarkson and Meltzer, as already mentioned. They used such a technique as an aid to the simulation of a trust investment officer and found good results. Independently, we believe, Weaver and Hall proposed a similar system as an investment theory. We have tried to distinguish Clarkson's emphasis on a theory of investment, and Weaver and Hall's emphasis on an investment theory. Our idea was to attempt a fusion through the similar technique and we tried to justify the approach through the well-established theory of the perceptron. Clarkson's point about a theory of investment requiring demonstration of the human decision processes as well as the decisions themselves, led to the restriction of the theory to a black-box type, but also allowed us to use the perceptron as the mechanism inside the black box. With such a black-box theory, the identification of inputs and outputs is clearly crucial. Although independence of the inputs has been demonstrated, implicitly, in the workings of the perceptron, it is entirely likely that not enough consideration was given to the choice of inputs. Certainly the results could, a priori, have been better had forecasts been included in the data set. Weaver and Hall's concentration on the inclusion of forecasted dividends and earnings seemed to give good results. A theory of investment is hard to justify without forecasts, especially in the light of Clarkson's model. There is also the possibility that different variables would be needed in a theory of investment behaviour as opposed to a predictive investment theory.* From this point of view, perhaps it was too ambitious to try a fusion of the two sorts of theory.

As to the distribution of the data, choosing yearly periods, we feel, is about right but perhaps using only ten years' data was insufficient. On the other hand, companies can change drastically in this sort of time period - the take-over is the obvious example - but these peculiar changes cannot, in general, be included except on an ad hoc basis.

The performance of the black-box mechanism, in this case the linear perceptron, can be discussed in isolation from the formation of the theory. Although the small

* As mentioned earlier, all the shares chosen moved more or less in parallel. Other, more "detailed" variables, could have highlighted the differences among the group much better.

number of patterns considered by the perceptron guaranteed predictions in the majority of cases (89%), it also meant that a large number of different sets of weights could separate the training set of patterns and this led to the distinct possibility of "dual" predictions. This fear is borne out in the results and is perhaps the most disappointing feature of them. There is simply no way of resolving this without adding more patterns for the perceptron to consider. On an information content basis, the perceptron can only produce more certain predictions if it has seen more patterns from the very large set of possible patterns. However, if there are more patterns than variables separability cannot be guaranteed. Only if the right choice of inputs has been made will this give better results.

The simple perceptron used in this study is not, as we have mentioned before, very "intelligent". What we mean by this is that there is more information contained in the data, as borne out by our analytical method, than the perceptron ever uses. A development of the concept of steering the perceptron, if accompanied by a parallel development in the theory, would vastly improve the meaningfulness of the perceptron training scheme.* In passing it might be remarked that Clarkson's discrimination net finds a parallel in the ordering of variables derived by analysis. The development of this into a more comprehensive theory has already been mentioned as a useful step forward.

COMPARISON WITH OTHER RESULTS

In a study menitoned in their book, Granger and Morgenstern describe an attempt to predict proportional changes in price in terms of proportional changes in earnings and dividends. This can be seen to be similar to our study here, except of course that we predict comparisons of proportional changes. Using regression techniques, they found no significant confirmation of the theory, and suggest instead a shift in emphasis to the prediction of relative price changes, as we have attempted.

Malkiel and Cragg tried earlier to carry out this suggestion, again using regression

* We were loath to develop the steering concept much further; the monitoring of the perceptron's behaviour during training using a set of heuristics is the obvious way, but this is a study in itself.

methods. They incorporated forecasts of growth in their model but still the results were unconvincing. The only study mentioned by Granger and Morgenstern to show any evidence of an ability to predict relative price changes is that of Weaver and Hall and this has been discussed in some detail. It must be remembered that they, unlike the other workers mentioned, are professional analysts and would try harder to produce evidence of success than an essentially disinterested academic worker (of whom the opposite might be said).

Our own study has produced results not dissimilar to those studies with which it can be compared, although it is clear that the future of scientific investment analysis lies in this direction. The prediction of relative price-changes is the key here, and there must be a shift of thinking if further negative studies are to be avoided.

CONCLUSION:

No excuses will be made for the presentation of negative results. We believe that we have justified the need for the study because not only is it based on two independent suggestions of a similar nature, but we think it also provides a demonstration of how techniques taken from Cybernetic thinking can provide a completely different angle in a traditional area, and produce equally valid results. One of the drawbacks of such traditional methods is that the interpretation of the models used is extremely difficult. Clarkson has demonstrated that a model using simulation techniques can not only have a valid and complete interpretation, but also yield excellent results. Hence our intended emphasis on a theory of investment - in other words, a theory which starts at the right place, with the investor.

APPENDIX: GRAPHS AND ARRAYS

The Key to the graphs is as follows:

All share prices are in new pence.

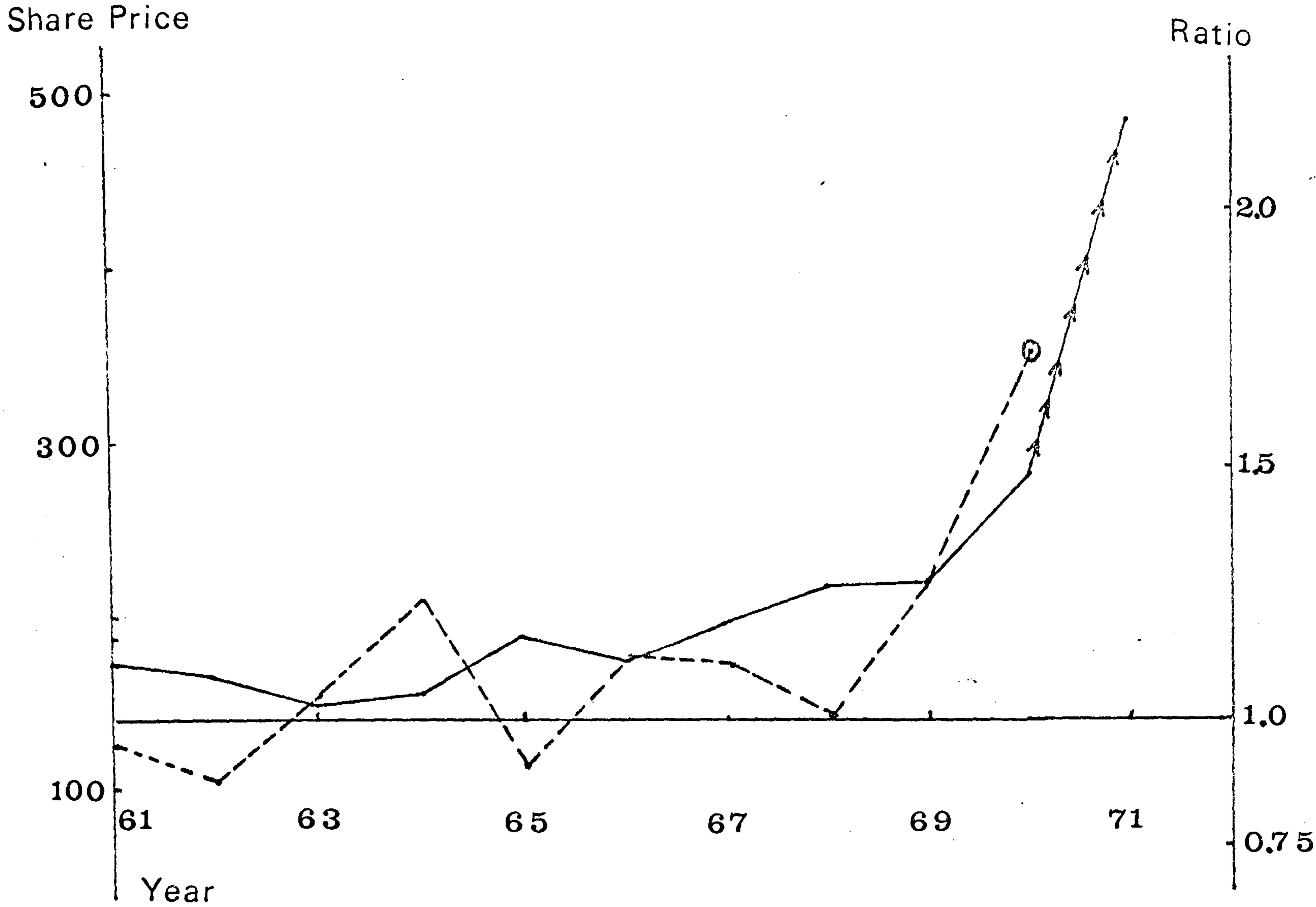
The solid lines connect average share prices over the year given. The dashed lines connect ratios of the year's average share price to the preceding year's average price.

The solid line with arrows represents the actual price movement which was being predicted. Similarly the small circle is the corresponding ratio.

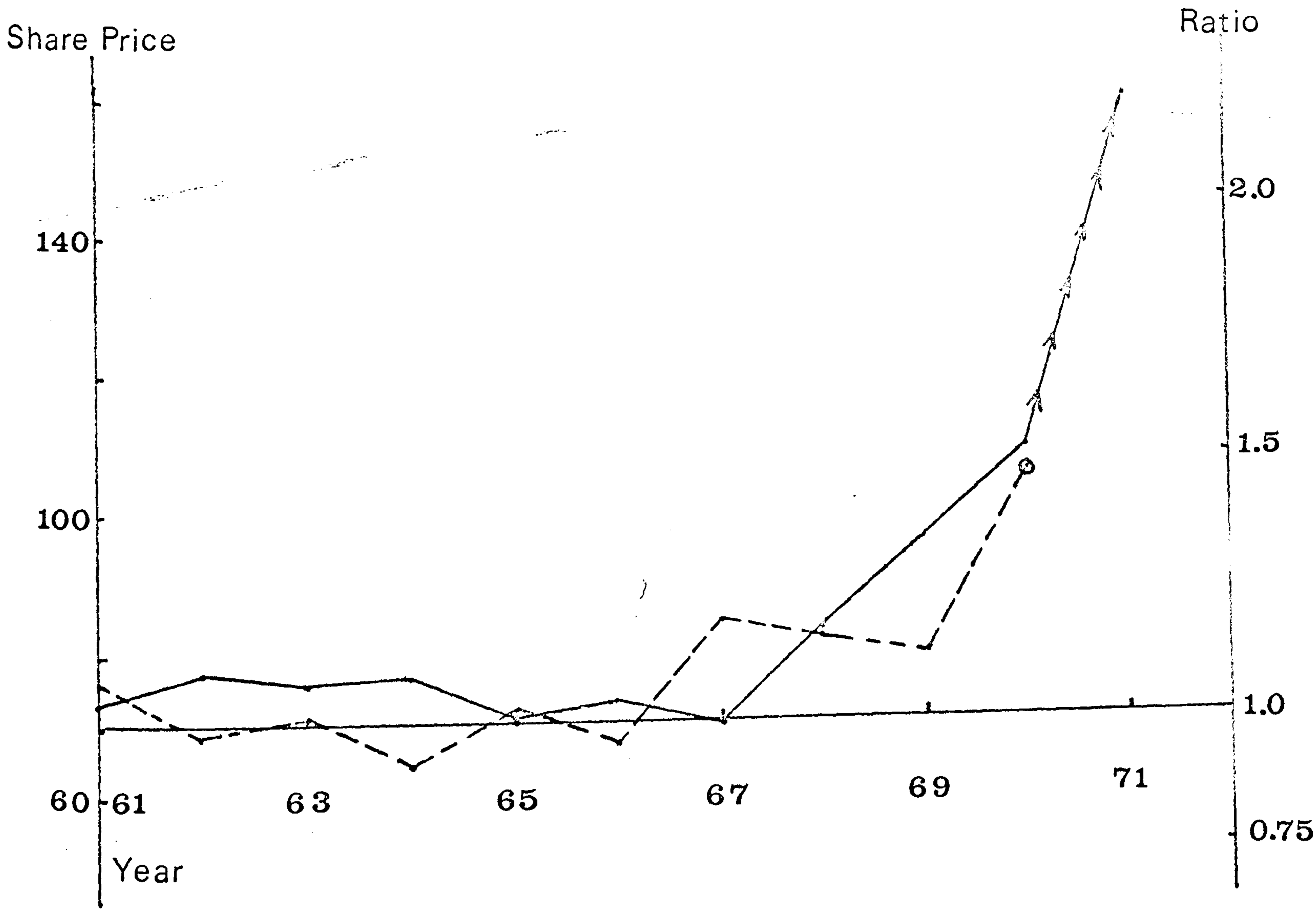
The arrays are the comparison arrays for each pair of shares as described and analysed in the thesis.

The format for each one is exactly as that on page five of Chapter six, except that the extra constant component needed by the perceptron has been added and a zero replaces the question-mark.

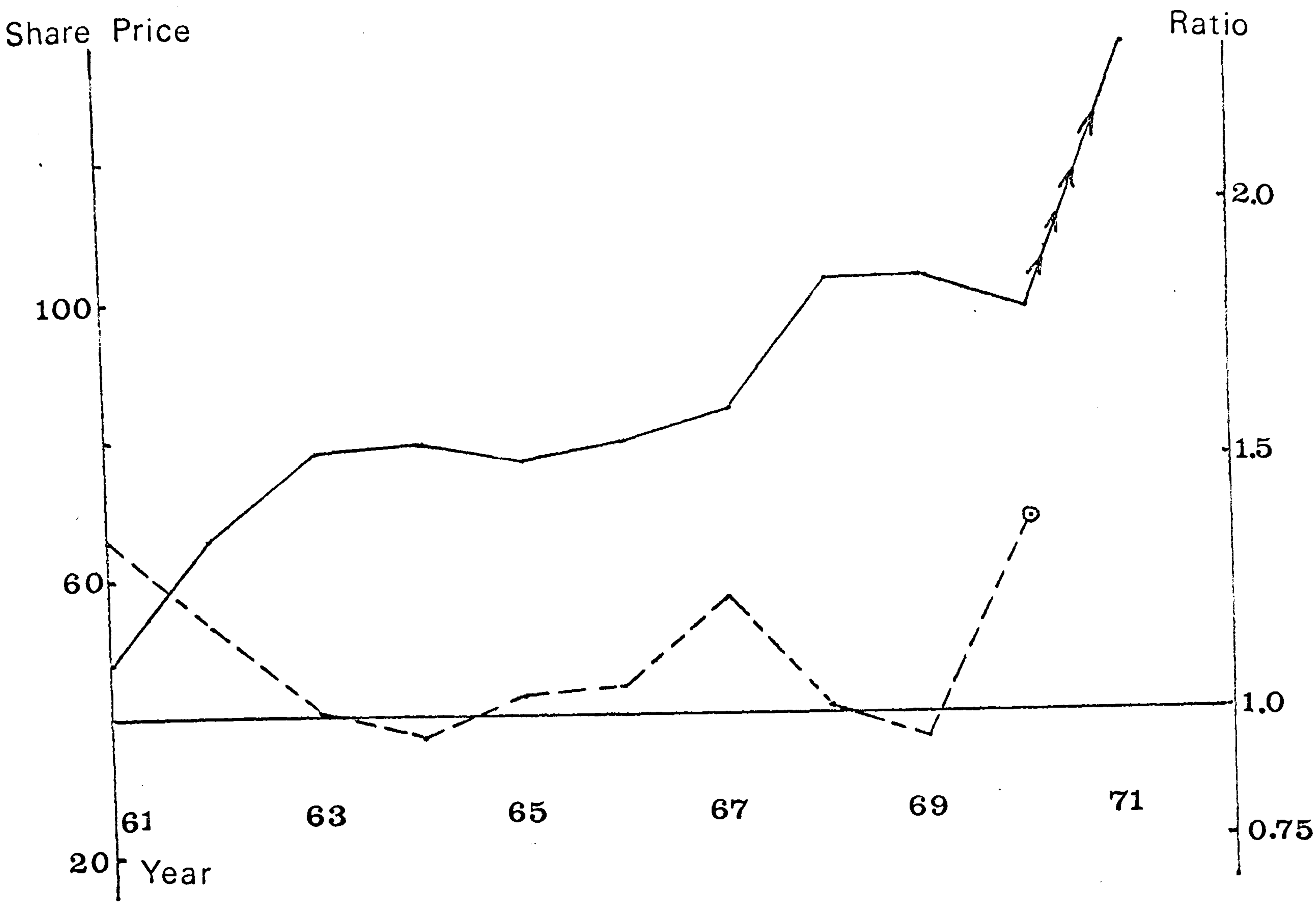
The shares are numbered from 70 to 79, but they are in the same order as quoted on page two of Chapter nine, (on results).



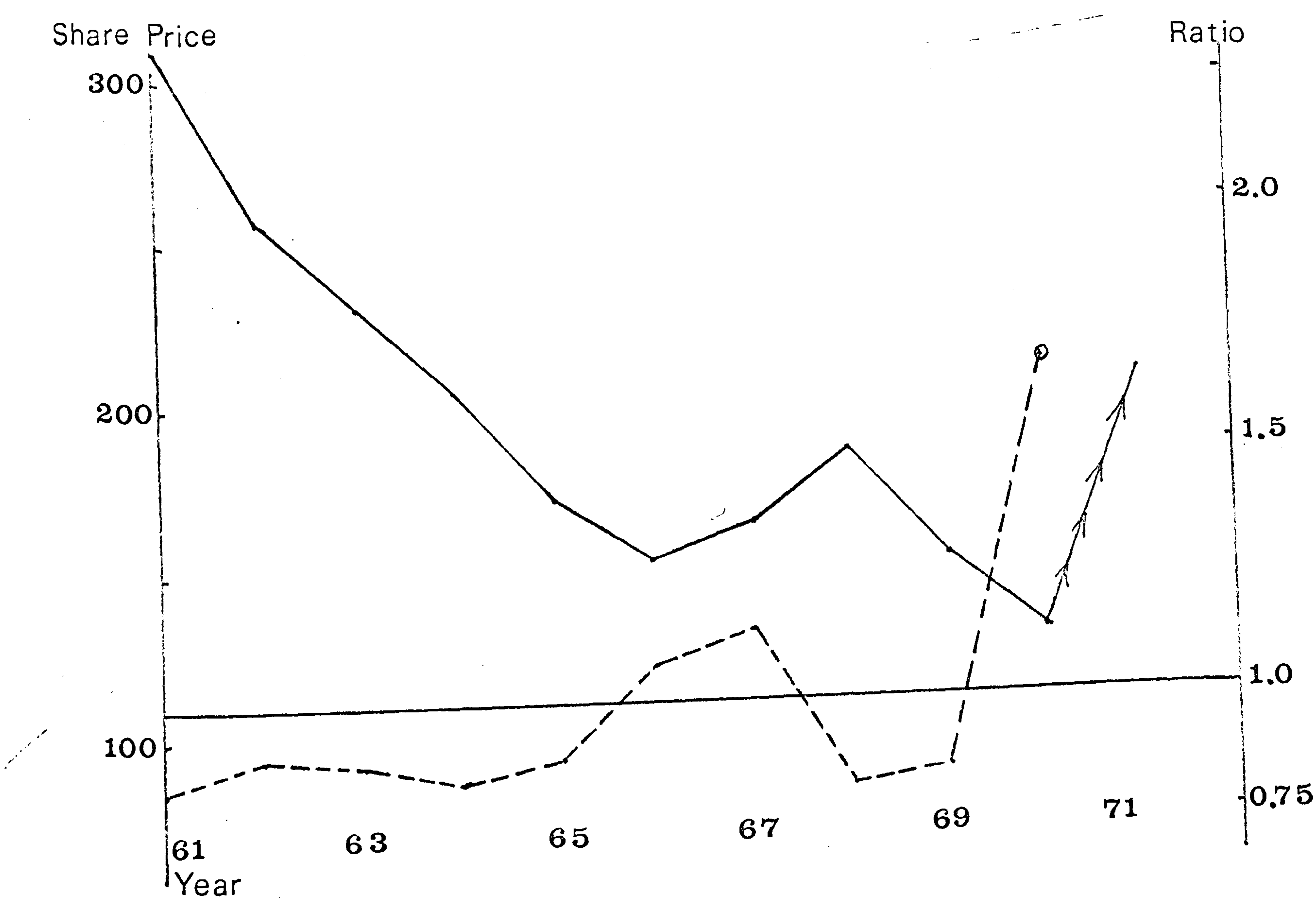
Share 0 Austin Reed



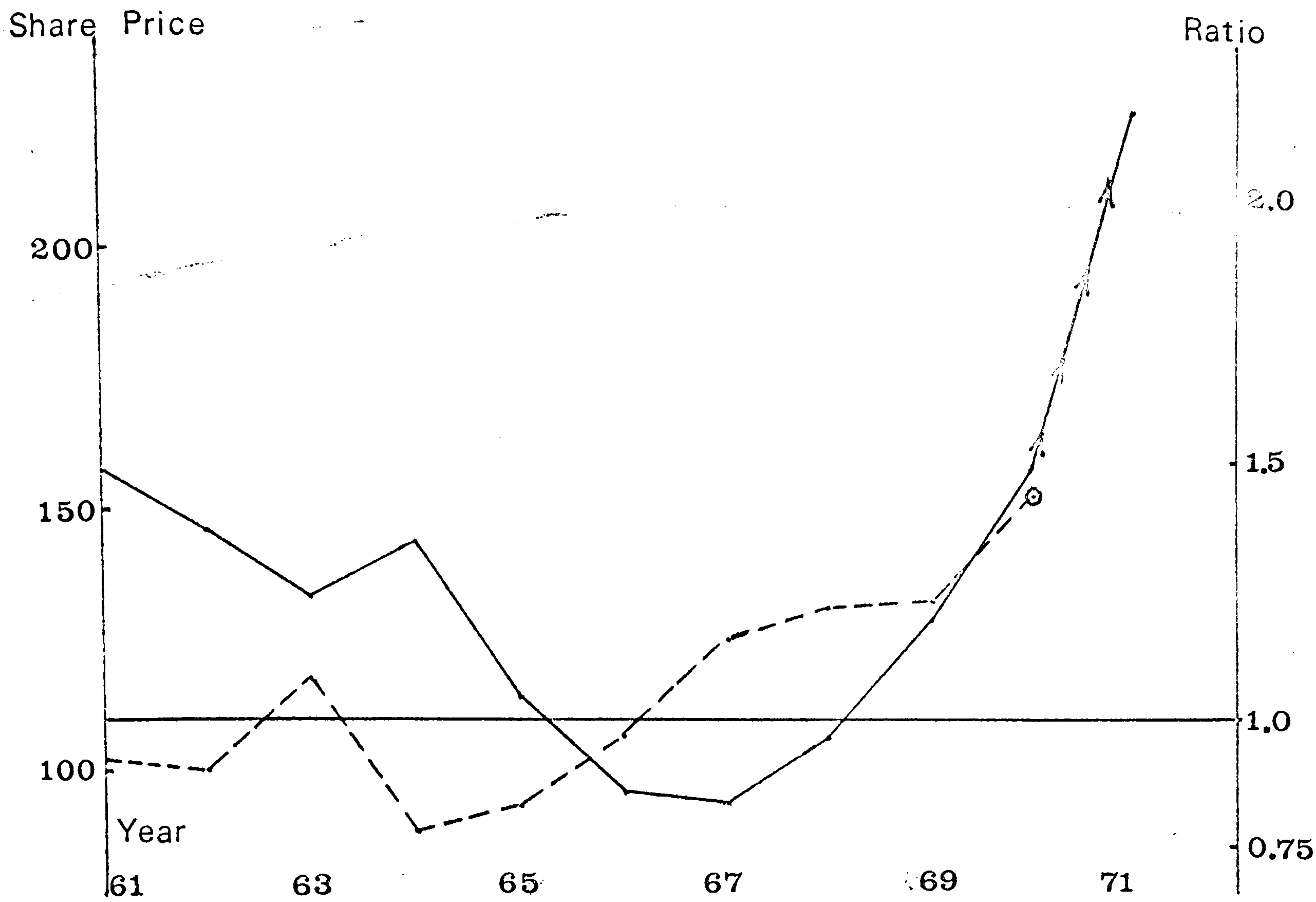
Share 1 Foster Brothers



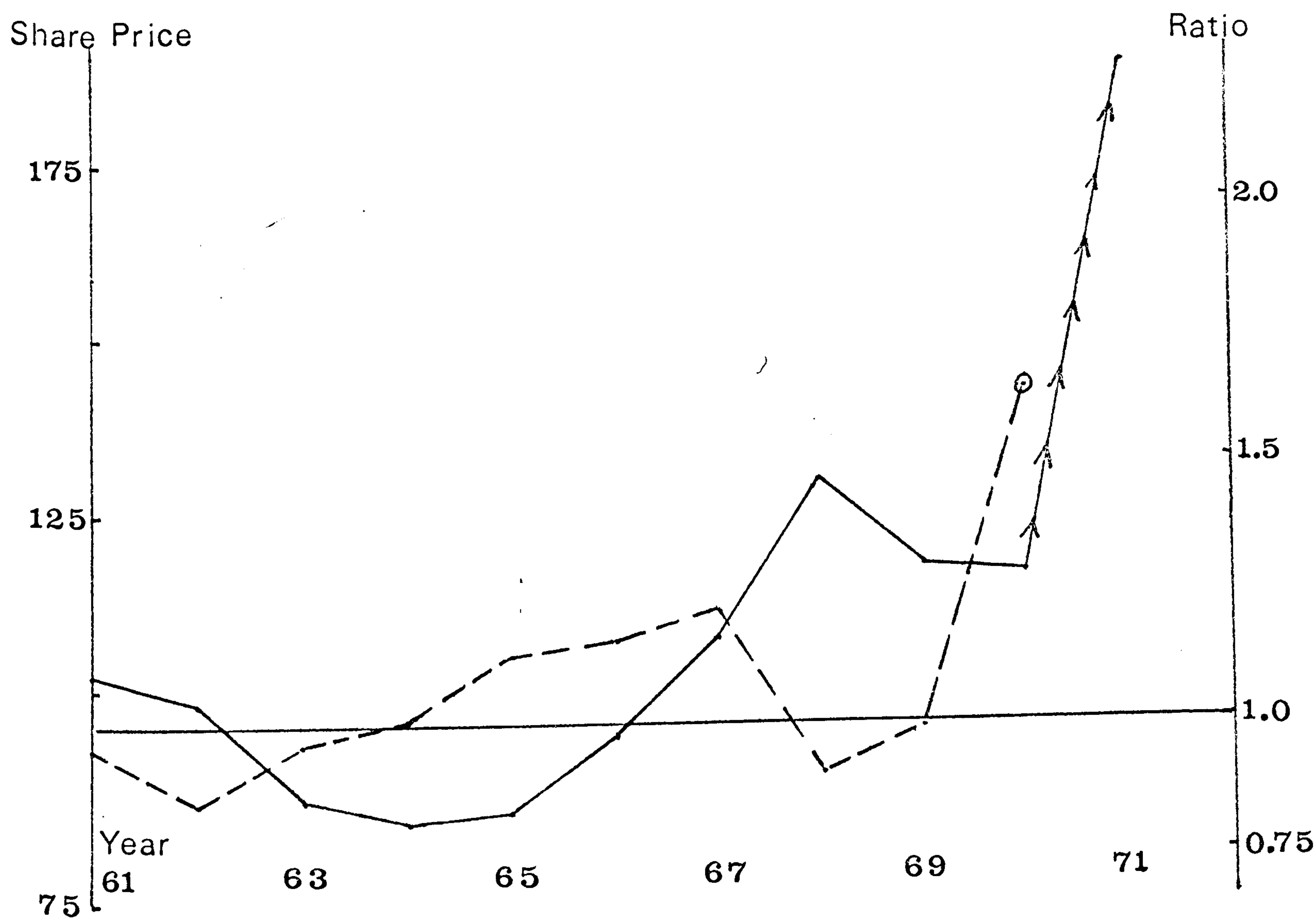
Share 2 H. Samuels



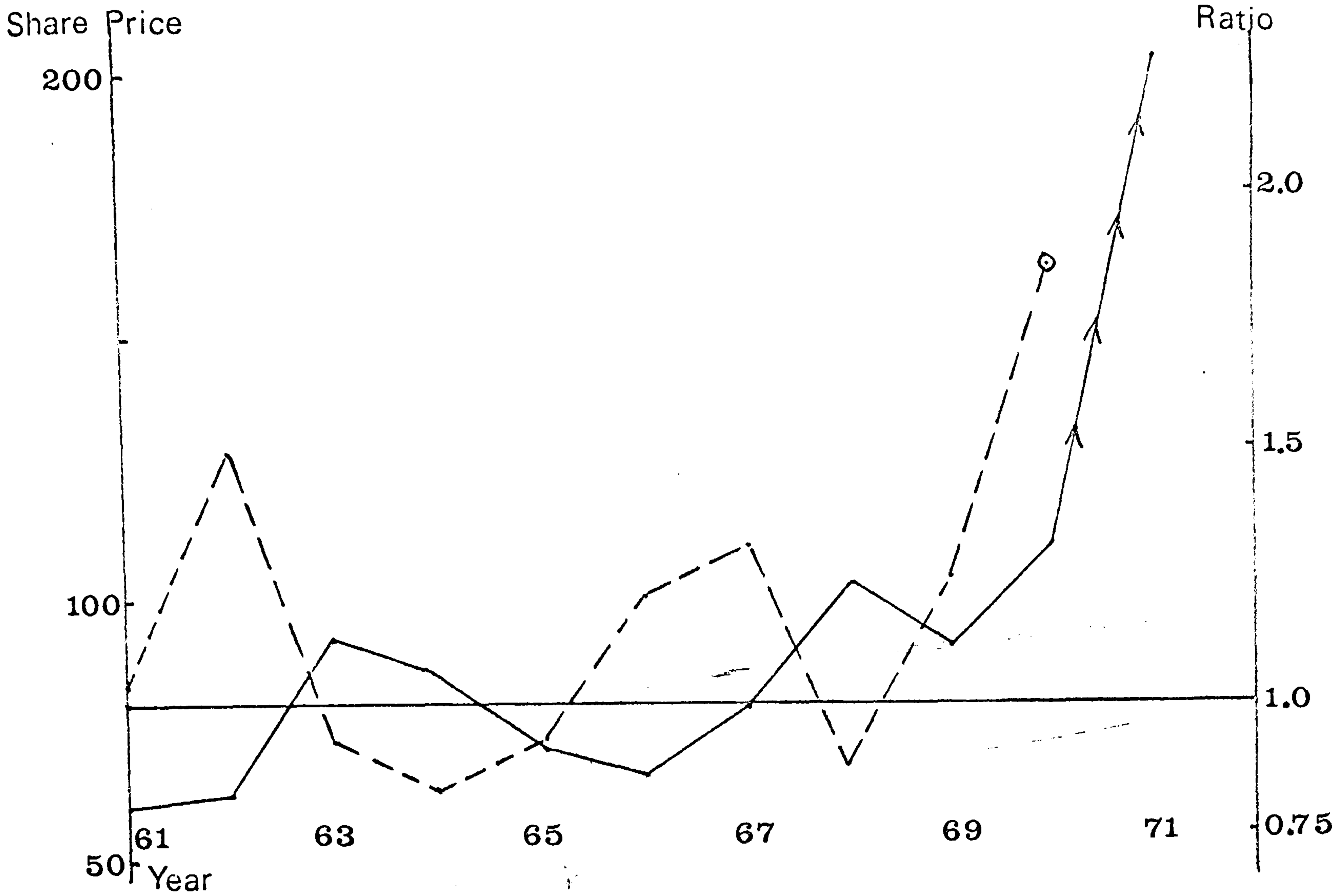
Share 3 Debenhams



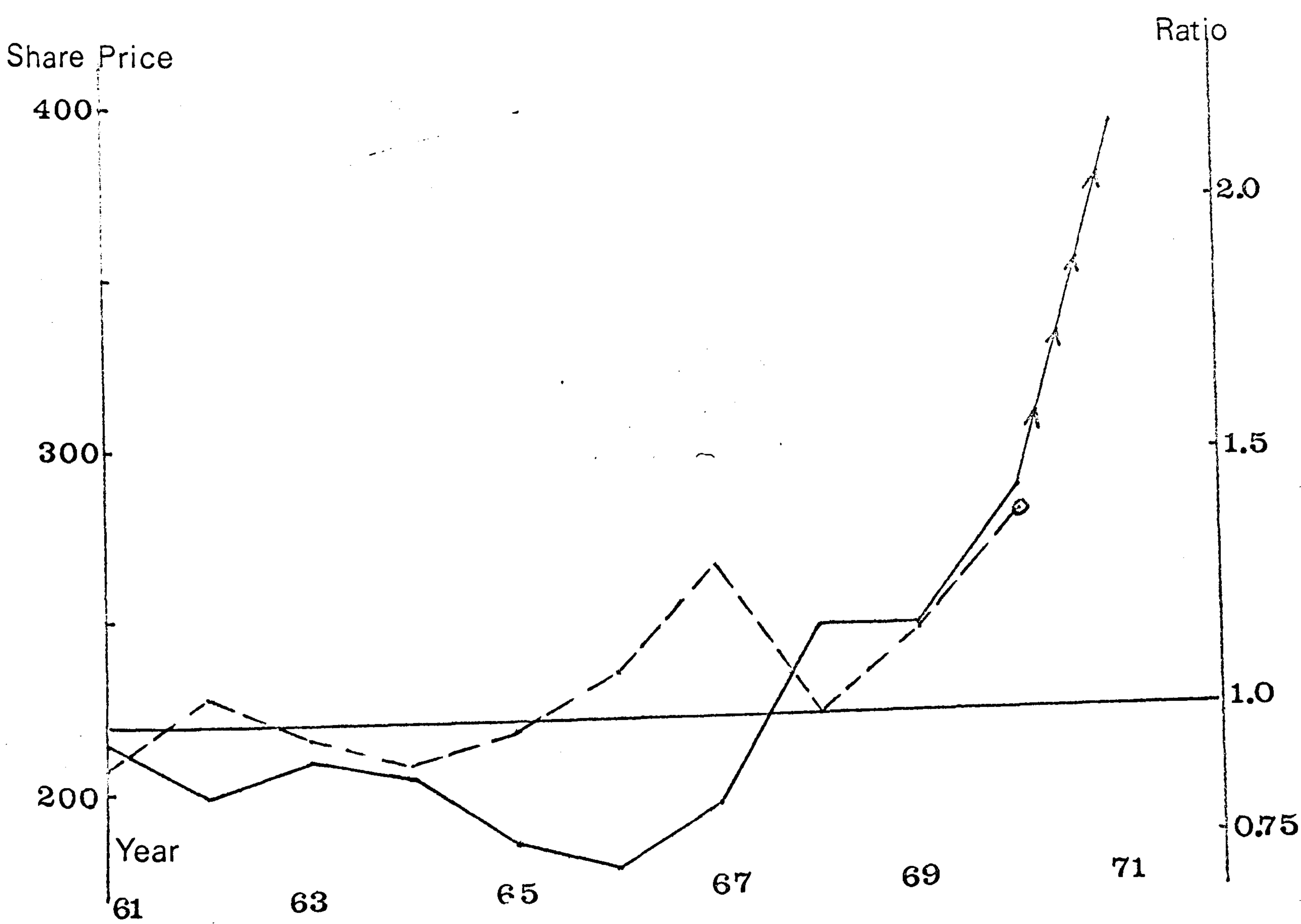
Share 4 Army and Navy



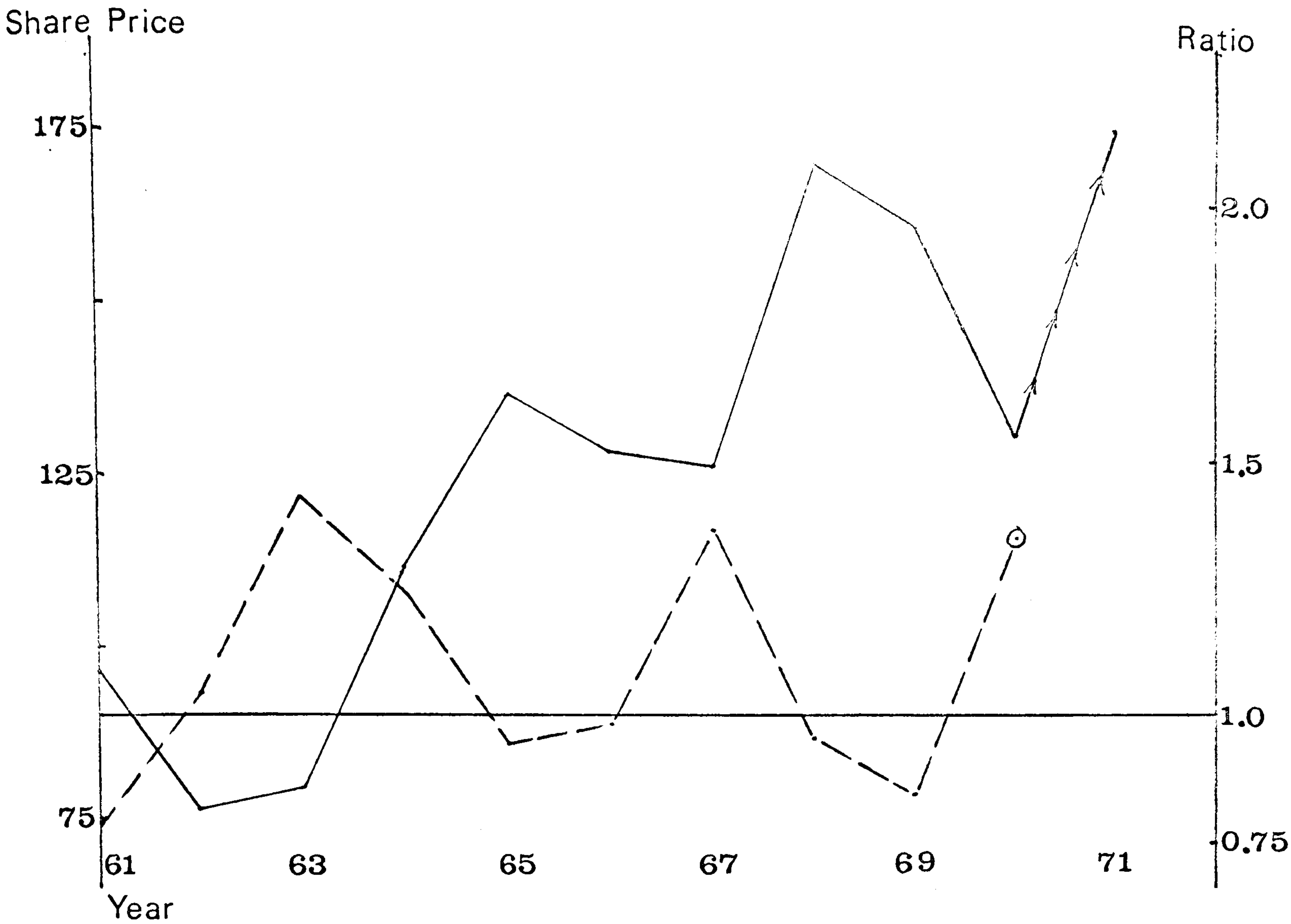
Share 5 Boots



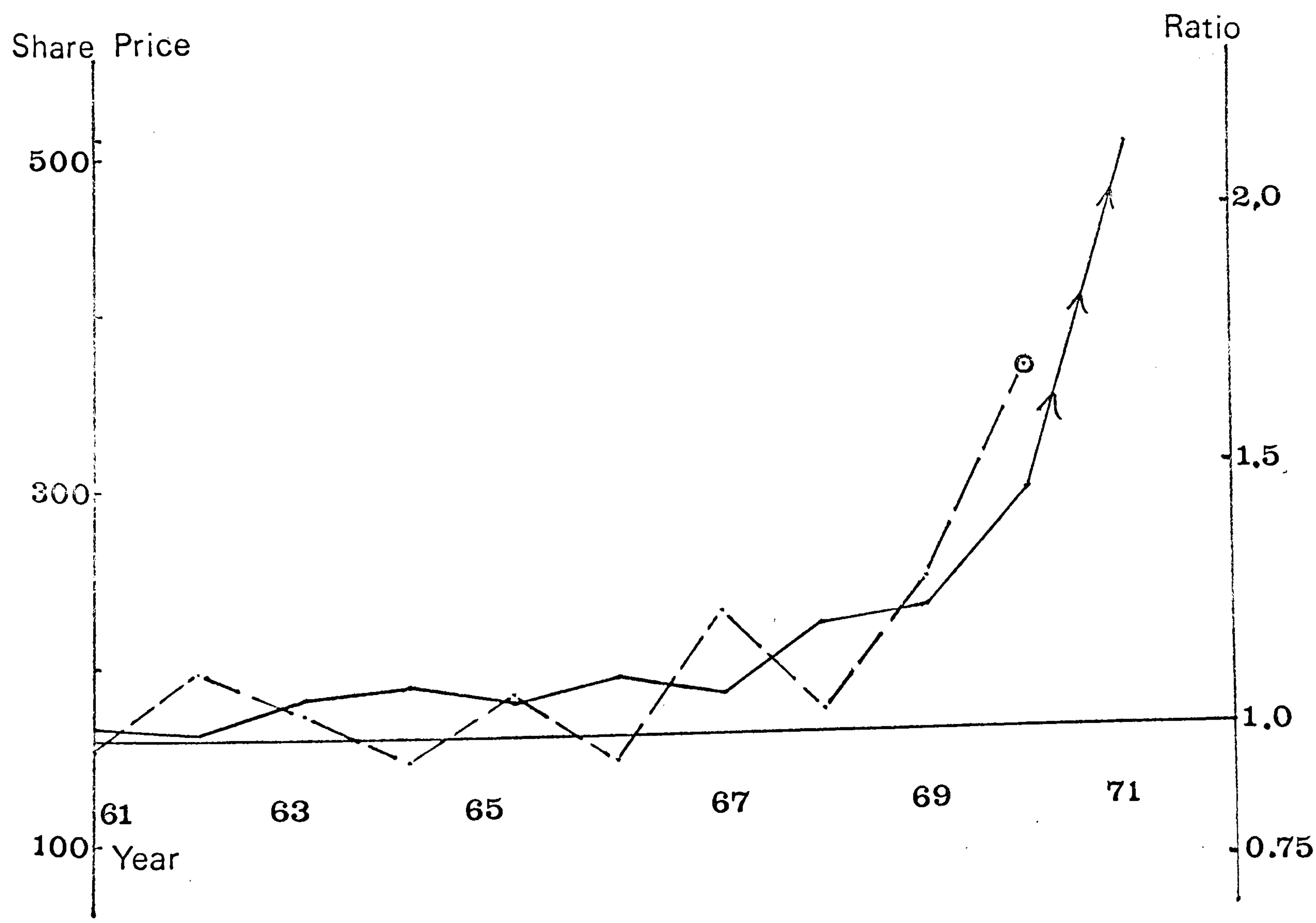
Share 6 Currys



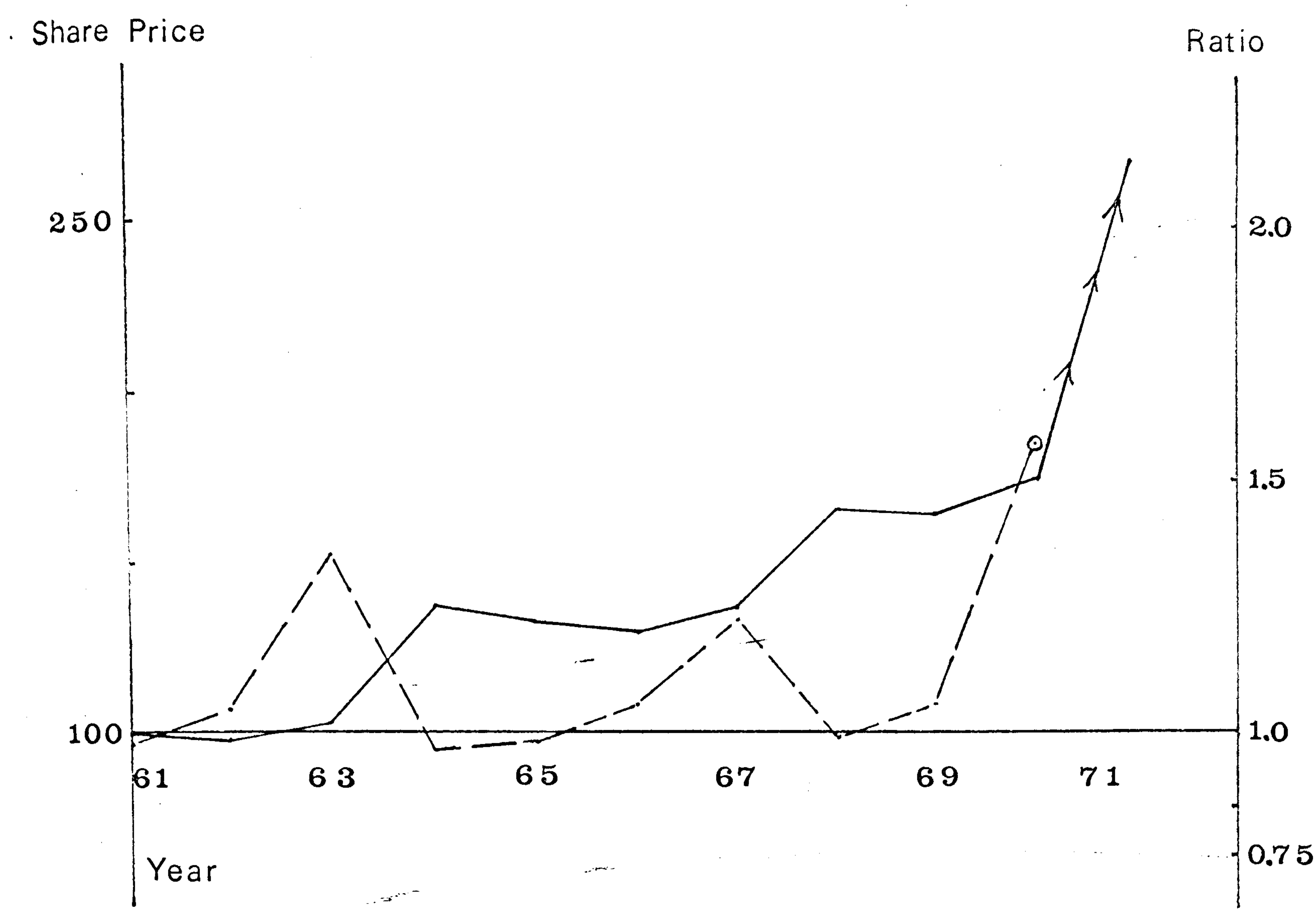
Share 7 Marks and Spencer



Share 8 Dorothy Perkins



Share 9 W.H.Smiths



The Average Share

SHARES 70 AND 71

3	1	1	1	1	3	1	3	1	1	1	0
1	1	1	1	3	3	1	1	1	1	1	1
3	3	3	3	3	3	3	2	3	3	1	3
3	1	1	1	3	1	1	1	3	1	1	3
2	3	3	3	3	3	1	2	1	3	1	1
2	1	1	1	3	1	3	3	1	1	1	3
2	1	1	1	1	3	1	2	1	1	1	1
2	3	3	3	1	1	1	1	3	3	1	1
2	3	3	3	3	1	1	3	1	3	1	3

SHARES 70 AND 72

1	1	1	1	1	3	1	3	3	1	1	0
1	1	1	1	3	3	3	1	3	1	1	1
3	3	3	3	2	3	3	2	1	3	1	3
3	1	1	1	3	3	1	1	1	1	1	3
2	3	3	3	3	3	1	2	3	3	1	1
2	3	1	1	3	1	3	1	1	3	1	3
2	1	1	1	1	3	1	3	3	1	1	1
2	3	3	3	3	3	1	1	1	3	1	1
2	3	3	3	1	1	1	2	1	3	1	3

SHARES 70 AND 73

1	1	1	1	1	1	3	2	3	1	1	0
1	1	1	1	1	3	3	1	1	1	1	1
3	3	3	3	2	3	3	2	1	3	1	1
1	1	1	1	1	1	1	1	3	1	1	3
3	1	1	3	1	1	1	3	1	3	1	1
2	1	1	1	1	1	3	1	1	1	1	1
2	1	1	1	1	1	1	3	1	1	1	1
2	3	3	3	3	1	1	1	3	3	1	1
2	3	3	1	1	1	3	3	1	3	1	3

SHARES 70 AND 74

1	1	1	1	1	1	3	2	1	1	1	0
1	1	1	1	1	3	1	1	3	1	1	1
3	3	3	3	2	1	3	2	1	3	1	3
3	1	1	1	3	1	1	1	3	1	1	3
2	1	1	1	1	1	1	2	1	3	1	1
2	3	1	1	1	3	3	3	3	1	1	1
2	1	3	1	3	1	1	3	3	1	1	1
2	3	3	3	1	1	1	1	1	3	1	3
2	1	3	3	1	1	3	3	1	3	1	3

SHARES 70 AND 75

1	1	1	1	3	3	3	2	1	1	1	0
3	1	1	1	3	3	3	3	3	1	1	1
3	3	3	3	3	3	3	2	3	3	1	1
1	3	3	1	3	1	1	1	1	3	1	3
3	3	3	3	3	3	1	2	1	3	1	3
3	3	1	1	3	3	3	3	3	1	1	3
3	1	1	1	1	1	1	2	3	1	1	1
3	3	3	3	3	3	3	1	3	3	1	1
3	3	3	3	1	1	3	2	1	3	1	1

SHARES 70 AND 76

1	1	1	1	1	3	3	2	3	1	1	0
1	1	1	1	2	3	3	1	3	1	1	1
3	3	3	3	3	3	3	3	3	3	1	1
1	3	3	3	3	3	1	1	3	3	1	3
3	3	3	3	1	3	1	2	1	3	1	3
3	3	1	1	3	3	3	1	1	3	1	3
3	1	1	1	1	3	1	3	3	1	1	1
2	3	1	1	1	1	1	1	1	3	1	1
2	3	3	3	1	1	3	2	1	3	1	3

SHARES 70 AND 77

3	1	1	1	3	1	3	2	3	1	1	0
1	1	1	1	1	3	3	1	3	1	1	1
3	3	3	3	3	3	3	2	3	3	1	1
1	3	3	3	3	1	1	1	3	3	1	3
3	3	3	3	3	1	3	2	1	3	1	1
3	1	1	1	3	1	3	3	1	1	1	3
3	1	1	1	1	3	1	3	3	1	1	1
3	3	3	3	3	1	1	1	3	3	1	1
3	3	3	3	3	1	3	3	1	3	1	3

SHARES 70 AND 78

1	1	1	1	1	1	3	2	1	1	1	0
1	1	1	1	1	3	3	1	3	1	1	1
3	3	3	3	3	3	3	2	3	3	1	1
1	1	1	1	3	3	1	1	3	1	1	3
3	3	1	3	1	3	1	2	3	3	1	1
3	3	3	1	3	3	3	1	3	3	1	3
3	1	3	3	3	3	1	3	3	1	1	1
3	3	3	3	3	2	1	1	3	3	1	3
3	1	1	1	1	1	3	2	1	3	1	3

SHARES 70 AND 79

1	1	1	1	3	3	1	3	3	1	1	0
1	3	3	3	3	3	3	1	3	3	1	1
3	3	3	3	2	3	3	2	3	3	1	3
1	3	1	1	3	1	1	3	3	1	1	3
3	3	3	3	3	3	1	2	1	3	1	1
3	1	1	1	3	1	3	1	1	1	1	3
3	1	1	1	3	1	3	3	1	1	1	1
2	3	3	3	3	1	1	1	1	3	1	1
2	3	3	3	3	1	3	2	1	3	1	3

SHARES 71 AND 72

1	1	1	1	1	3	3	3	3	1	1	0
1	1	1	1	3	1	3	2	3	1	1	1
1	1	3	3	1	3	1	2	1	3	1	1
3	3	3	3	2	3	3	2	1	3	1	3
2	3	1	1	3	3	1	2	3	1	1	3
2	3	3	3	3	3	1	1	3	3	1	3
2	1	1	1	1	3	3	3	3	1	1	3
2	3	3	3	3	3	3	3	1	1	1	1
2	3	3	3	1	1	1	1	1	3	1	3

SHARES 71 AND 73

1	1	1	1	1	1	3	1	3	1	1	0
1	1	1	1	1	1	3	2	3	1	1	1
1	1	3	3	1	3	1	2	1	3	1	1
1	3	1	1	1	1	3	2	3	3	1	1
3	1	1	1	1	1	3	3	1	1	1	3
2	3	3	1	1	1	1	1	3	1	1	1
2	1	1	1	1	1	3	3	1	1	1	1
2	3	3	3	3	3	3	3	1	1	1	1
2	1	1	1	1	3	3	1	3	1	1	1

SHARES 71 AND 74

1	1	1	1	1	1	3	1	1	1	1	0
1	1	1	1	1	1	1	2	3	1	1	1
1	1	1	1	1	1	1	2	1	3	1	3
3	3	3	3	2	1	3	2	3	3	1	1
2	1	1	1	1	1	3	2	3	1	1	3
2	3	3	3	1	3	1	1	3	3	1	1
2	3	3	3	3	1	3	3	3	3	1	1
2	1	1	1	2	1	3	2	1	1	1	3
2	1	1	1	1	1	3	1	1	1	1	1

SHARES 71 AND 75

1	1	3	1	3	1	3	1	1	3	1	0
3	1	1	1	3	1	3	3	3	1	1	1
1	1	1	1	3	3	1	2	1	3	1	1
1	3	3	3	2	3	3	2	1	3	1	3
3	1	1	1	3	3	1	2	3	1	1	3
3	3	3	3	3	3	1	3	3	1	1	3
3	1	1	1	3	1	3	2	3	3	1	3
3	3	3	3	3	3	3	2	1	1	1	1
3	3	3	3	1	1	3	1	1	3	1	1

SHARES 71 AND 76

1	1	1	1	1	3	3	1	3	1	1	0
1	1	1	1	1	3	3	2	3	1	1	3
3	1	3	3	1	3	1	3	1	3	1	1
1	3	3	3	2	3	3	2	3	3	1	3
3	1	1	1	1	1	1	2	3	1	1	3
3	3	3	3	3	3	1	1	3	3	1	1
3	1	1	1	1	3	3	3	3	1	1	1
2	1	1	1	2	3	1	2	1	1	1	1
2	3	3	3	1	3	3	1	3	3	1	3

SHARES 71 AND 77

1	3	3	3	3	1	3	1	3	3	1	0
1	1	1	1	1	1	3	2	3	1	1	1
1	1	1	1	1	1	1	2	1	3	1	1
1	3	3	3	2	1	3	3	3	3	1	3
3	1	1	1	3	1	3	2	3	1	1	3
3	3	3	3	1	1	1	3	3	3	1	1
3	1	1	1	1	3	1	3	3	1	1	1
3	3	3	3	3	3	1	2	1	3	1	1
3	3	3	3	3	3	3	1	3	1	1	3

SHARES 71 AND 78

1	1	1	1	1	1	3	1	1	1	1	0
1	1	1	1	1	1	3	2	3	1	1	1
1	3	3	3	3	3	1	2	1	3	1	1
1	1	1	1	2	3	3	2	3	1	1	3
3	3	1	3	1	3	3	2	3	1	1	3
3	3	3	3	3	3	1	1	3	3	1	1
3	3	3	3	3	3	1	3	3	3	1	3
3	3	3	3	3	3	1	2	1	3	1	3
3	1	1	1	1	3	3	1	3	1	1	3

SHARES 71 AND 79

1	1	3	3	3	3	1	1	3	1	1	0
1	3	3	3	3	1	3	2	3	3	1	3
1	1	1	1	1	3	1	2	1	1	1	1
1	3	3	3	2	1	3	3	3	3	1	3
3	1	1	1	3	3	1	2	2	1	1	1
3	3	3	3	3	1	1	1	1	3	1	3
3	3	3	3	3	1	3	3	1	3	1	3
2	3	1	1	3	3	1	2	1	3	1	3
2	3	3	3	3	1	3	1	1	3	1	3

SHARES 72 AND 76

3	3	3	3	3	1	3	1	1	3	1	0
3	1	1	1	1	3	3	2	3	1	1	3
3	1	3	3	3	3	1	3	3	1	1	1
1	3	3	3	2	3	1	2	3	3	1	3
3	1	1	1	1	1	1	2	1	1	1	3
3	3	3	3	3	3	3	2	3	3	1	1
3	3	3	3	1	1	3	3	1	3	1	1
2	1	1	1	1	1	1	1	3	1	1	1
2	3	3	3	2	3	3	2	3	1	1	3

SHARES 72 AND 73

1	3	1	1	1	1	3	1	1	3	1	0
1	1	1	1	1	1	3	2	1	1	1	1
1	1	3	1	2	1	3	2	1	1	1	1
1	3	1	1	1	1	1	2	3	1	1	1
3	1	1	3	1	1	3	3	1	1	1	3
2	1	3	1	1	1	3	2	3	1	1	1
2	3	3	3	3	1	3	1	1	3	1	1
2	3	3	3	1	1	1	3	3	3	1	1
2	1	1	1	3	3	3	3	3	1	1	1

SHARES 72 AND 77

3	3	3	3	3	1	3	1	1	3	1	0
1	1	1	1	1	1	1	2	3	1	1	3
3	1	1	1	3	1	1	2	3	1	1	1
1	3	3	3	2	1	1	3	3	3	1	3
3	1	1	1	1	1	3	2	1	3	1	3
3	1	1	1	1	1	3	3	3	1	1	1
3	3	2	3	3	3	1	3	1	1	1	1
3	3	3	3	3	1	1	1	3	3	1	1
3	1	3	1	3	3	3	3	3	1	1	1

SHARES 72 AND 74

1	1	3	1	2	1	3	1	1	3	1	0
1	1	1	1	1	1	1	2	1	1	1	3
1	1	1	1	2	1	3	2	1	1	1	3
1	3	3	3	2	1	3	2	3	1	1	1
2	1	1	1	1	1	3	2	1	1	1	1
2	3	3	3	1	3	3	3	3	1	1	1
2	3	3	3	3	1	3	1	1	3	1	1
2	1	1	1	1	1	3	1	3	1	1	3
2	1	1	1	2	1	3	3	1	1	1	1

SHARES 72 AND 78

1	3	1	1	2	1	3	1	1	3	1	0
1	1	1	1	1	3	3	2	1	1	1	1
3	3	3	3	3	3	1	2	3	3	1	1
1	1	1	1	2	1	1	2	3	1	1	3
3	1	1	3	1	3	3	2	3	1	1	1
3	3	3	1	3	3	3	2	3	3	1	1
3	3	3	3	3	3	1	3	1	3	1	3
3	3	3	3	3	1	1	1	3	3	1	3
3	1	1	1	2	3	3	2	3	1	1	1

SHARES 72 AND 75

1	3	3	3	3	1	3	1	1	3	1	0
3	1	1	1	1	1	3	3	3	1	1	3
3	1	1	1	3	1	1	2	3	1	1	1
1	3	3	3	2	1	1	2	3	3	1	1
3	1	1	1	3	3	1	2	1	3	1	3
3	2	3	3	3	3	3	3	3	1	1	3
3	3	3	3	3	1	3	1	1	3	1	3
3	3	3	3	3	3	3	1	3	3	1	1
3	1	1	1	2	3	3	2	1	1	1	1

SHARES 72 AND 79

1	3	3	3	3	3	1	1	3	3	1	0
1	3	3	3	3	1	3	2	1	3	1	3
3	1	1	1	2	1	1	2	3	1	1	3
1	3	1	1	2	1	3	3	3	1	1	1
3	1	3	3	1	1	1	2	1	3	1	1
3	1	1	1	3	1	3	2	1	1	1	3
3	3	3	3	3	1	3	3	1	3	1	1
2	1	1	1	1	1	1	1	3	3	1	3
2	3	3	3	3	1	3	2	1	3	1	1

SHARES 73 AND 74

1	1	3	3	3	3	3	2	1	1	1	0
1	1	3	1	2	3	1	2	3	3	1	3
1	3	1	1	2	1	3	2	2	3	1	3
3	3	3	3	3	3	3	2	1	3	1	3
1	3	3	1	2	3	3	1	3	1	1	1
2	3	3	3	2	3	1	3	3	3	1	1
2	3	3	3	3	1	3	3	3	3	1	1
2	1	1	1	1	1	3	1	1	1	1	3
2	1	3	3	1	1	1	1	1	3	1	3

SHARES 73 AND 75

3	3	3	3	3	3	3	2	1	3	1	0
3	3	3	1	3	2	3	3	3	1	1	3
3	3	1	1	3	3	1	2	3	3	1	3
3	3	3	3	3	3	3	2	1	3	1	3
1	3	3	1	3	3	1	1	3	3	1	3
3	3	1	3	3	3	1	3	3	1	1	3
3	3	3	3	3	1	3	1	3	3	1	3
3	1	1	1	3	3	3	1	3	3	1	3
3	3	3	3	1	1	1	1	1	3	1	1

SHARES 73 AND 76

3	3	3	3	3	3	3	2	1	3	1	0
3	1	1	1	3	3	1	2	3	1	1	3
3	3	1	3	3	3	1	3	3	1	1	3
3	3	3	3	3	3	2	2	3	3	1	3
1	3	3	1	2	3	1	1	3	1	1	3
3	3	3	3	3	3	3	2	3	3	1	3
3	3	1	2	1	3	3	3	3	1	1	1
2	1	1	1	1	3	1	1	1	1	1	3
2	3	3	3	1	3	1	1	3	3	1	3

SHARES 73 AND 77

3	3	3	3	3	3	3	2	1	3	1	0
1	3	3	3	3	3	1	2	3	1	1	3
3	3	1	1	3	1	1	2	3	3	1	3
3	3	3	3	3	3	3	3	3	3	1	3
1	3	3	1	3	3	3	1	3	3	1	3
3	3	1	3	3	3	1	3	1	3	1	3
3	3	1	1	3	3	1	3	3	1	1	3
3	3	1	1	3	3	1	1	1	3	1	3
3	3	3	3	3	1	1	3	1	3	1	3

SHARES 73 AND 78

1	1	1	1	3	3	3	2	1	1	1	0
1	1	3	1	2	3	3	2	3	1	1	1
3	3	3	3	3	3	1	2	3	3	1	3
3	1	1	1	3	3	3	2	1	1	1	3
1	3	3	3	2	3	3	1	3	1	1	1
3	3	3	3	3	3	1	2	3	3	1	3
3	3	3	3	3	3	1	3	3	3	1	3
3	3	2	3	3	3	1	1	3	3	1	3
3	1	1	1	1	3	3	1	3	1	1	3

SHARES 73 AND 79

3	3	3	3	3	3	1	3	3	1	1	0
1	3	3	3	3	3	3	2	3	3	1	3
3	3	1	1	2	3	1	2	3	1	1	3
3	3	3	3	3	3	3	3	1	3	1	3
1	3	3	1	3	3	1	1	3	3	1	1
3	3	1	3	3	3	1	2	1	3	1	3
3	3	3	3	3	3	3	3	3	3	1	3
2	1	1	1	3	1	1	1	1	3	1	3
2	3	3	3	3	1	1	1	1	3	1	3

SHARES 74 AND 75

3	3	3	3	3	3	3	2	2	3	1	0
3	3	3	1	3	1	3	3	3	1	1	1
3	3	3	3	3	3	1	2	3	1	1	1
1	3	3	3	2	3	1	2	1	3	1	3
3	3	3	3	3	3	1	2	3	3	1	3
3	1	1	1	3	3	3	3	3	1	1	3
3	1	1	1	1	3	2	1	1	1	1	3
3	3	3	3	3	3	3	2	3	3	1	1
3	3	3	3	2	3	1	1	3	3	1	1

SHARES 74 AND 76

3	3	3	3	3	3	1	2	3	3	1	0
3	1	1	1	3	3	3	2	3	1	1	3
3	1	3	3	3	3	1	3	3	1	1	1
1	3	3	3	2	3	1	2	3	3	1	3
3	3	3	3	2	3	1	2	3	3	1	3
3	1	1	1	3	1	3	1	1	3	1	3
3	1	1	1	1	3	3	3	1	1	1	3
2	1	1	1	2	3	1	2	3	1	1	1
2	3	3	3	2	3	1	1	3	3	1	3

SHARES 74 AND 77

3	3	3	3	3	3	1	2	3	3	1	0
3	3	3	3	3	1	3	2	3	1	1	3
3	3	3	3	3	3	1	2	3	1	1	1
1	3	3	3	2	3	1	3	3	3	1	3
3	3	3	3	3	3	3	2	3	3	1	3
3	1	1	1	3	1	1	3	1	1	1	3
3	1	1	1	1	3	1	3	3	1	1	3
3	3	3	3	3	3	1	2	3	3	1	1
3	3	3	3	3	3	3	3	3	3	1	3

SHARES 75 AND 77

3	3	3	3	1	1	1	2	3	3	1	0
1	1	3	3	1	3	1	1	1	3	1	3
3	1	3	3	1	1	3	2	3	1	1	3
3	3	3	3	2	1	1	3	3	1	1	3
1	1	1	1	1	1	3	2	1	3	1	1
3	1	1	1	1	1	1	3	1	3	1	1
3	1	1	1	1	3	1	3	3	1	1	1
3	3	1	1	3	1	1	2	1	3	1	3
1	1	3	3	3	3	3	3	3	1	1	3

SHARES 74 AND 78

3	3	1	1	2	3	3	2	1	3	1	0
1	1	1	1	2	3	3	2	3	1	1	1
3	3	3	3	3	3	1	2	3	3	1	1
1	1	1	1	2	3	1	2	1	1	1	3
3	3	3	3	2	3	3	2	3	1	1	3
3	3	3	1	3	3	1	1	3	3	1	3
3	1	3	3	3	3	1	3	3	1	1	3
3	3	3	3	3	3	1	2	3	3	1	3
3	1	1	1	2	3	3	1	3	1	1	3

SHARES 75 AND 78

1	1	1	1	1	1	3	2	1	1	1	0
1	1	1	3	1	3	3	1	1	1	1	1
3	3	3	3	3	3	3	2	3	3	1	3
3	1	1	1	2	3	1	2	3	1	1	3
3	3	1	3	1	3	3	2	3	1	1	1
3	3	3	1	3	1	1	1	1	3	1	1
3	3	3	3	3	3	1	3	3	3	1	3
3	3	3	3	3	1	1	2	3	3	1	3
3	1	1	1	2	3	3	2	3	1	1	3

SHARES 74 AND 79

3	3	3	3	3	3	1	3	3	1	1	0
3	3	3	3	3	3	3	2	3	3	1	3
3	1	1	1	2	3	1	2	3	1	1	1
1	3	1	1	2	3	3	3	3	1	1	3
3	3	3	3	3	3	1	2	1	3	1	1
3	1	1	1	3	1	1	1	1	1	1	3
3	1	1	1	3	3	3	3	1	1	1	3
2	3	3	3	3	3	1	2	3	3	1	1
2	3	3	3	3	3	1	1	3	3	1	3

SHARES 75 AND 79

3	3	3	3	1	3	1	3	3	1	1	0
1	3	3	3	3	3	1	1	1	3	1	3
3	1	1	1	1	2	3	2	3	1	1	3
3	3	1	1	2	1	3	3	3	1	1	3
3	1	3	3	1	1	1	2	1	3	1	1
3	1	1	1	1	1	1	1	1	3	1	1
3	3	3	3	3	3	3	3	1	1	1	1
1	1	1	1	1	1	1	2	1	3	1	3
1	3	3	3	3	1	3	2	1	3	1	3

SHARES 75 AND 76

3	3	1	1	1	3	1	2	3	1	1	0
1	1	1	1	1	3	1	1	1	1	1	3
3	1	3	3	1	3	3	3	3	1	1	1
3	3	3	3	2	3	1	2	3	3	1	3
1	1	1	1	1	1	1	2	1	1	1	3
3	3	3	3	3	1	3	1	1	3	1	1
3	1	1	1	1	3	3	3	3	1	1	1
1	1	1	1	1	1	1	2	1	1	1	1
1	3	3	3	2	3	3	2	3	3	1	3

SHARES 76 AND 77

3	3	3	3	3	1	3	2	3	3	1	0
1	3	3	3	1	1	1	2	3	3	1	1
1	3	1	1	1	1	3	1	1	3	1	3
1	1	1	1	2	1	3	3	1	1	1	1
1	3	3	3	3	1	3	2	1	3	1	1
1	1	1	1	1	1	1	3	1	1	1	3
1	1	1	1	3	3	1	1	3	1	1	3
3	3	3	3	3	1	3	2	3	3	1	3
3	1	1	1	3	1	3	3	1	1	1	1

SHARES 76 AND 78

1	1	1	1	1	1	3	2	1	1	1	0
1	3	3	3	1	1	3	2	1	3	1	1
1	3	3	3	3	1	1	1	3	3	1	3
1	1	1	1	2	1	3	2	1	1	1	3
3	3	1	3	2	3	3	2	3	1	1	1
3	3	3	1	1	3	1	2	3	3	1	3
1	3	3	3	3	3	1	3	3	3	1	3
3	3	3	3	3	3	3	2	3	3	1	3
3	1	1	1	2	3	3	2	3	1	1	1

SHARES 78 AND 79

3	3	3	3	3	3	1	3	3	1	1	0
3	3	3	3	3	1	1	2	1	3	1	3
1	1	1	1	1	1	1	2	1	1	1	3
2	3	3	3	2	1	3	3	3	3	1	1
3	1	3	1	3	1	1	2	1	3	1	1
1	1	1	3	1	1	3	2	1	1	1	3
3	1	1	1	1	1	3	1	1	1	1	1
1	1	1	1	1	1	3	2	1	1	1	1
1	3	3	3	3	1	1	2	1	3	1	3

SHARES 76 AND 79

1	1	3	3	3	3	1	3	3	1	1	0
1	3	3	3	3	1	3	2	1	3	1	3
1	1	1	1	1	1	1	1	1	1	1	3
1	1	1	1	2	1	3	3	1	1	1	1
3	3	3	3	3	3	1	2	1	3	1	1
3	1	1	1	1	1	1	2	1	1	1	3
3	3	3	3	3	1	3	3	1	3	1	3
2	3	3	3	3	1	3	2	1	3	1	3
2	3	3	3	3	1	3	2	1	3	1	1

SHARES 70 AND 45

1	1	1	1	1	3	3	3	3	1	1	0
1	1	1	1	3	3	3	3	3	1	1	1
1	3	3	3	3	3	3	3	3	3	1	1
1	1	1	1	3	1	1	1	3	3	1	3
1	3	3	3	3	3	1	2	1	3	1	1
1	1	1	1	3	3	3	3	1	1	1	3
1	1	1	1	1	3	1	3	3	1	1	1
1	3	3	3	3	1	1	1	1	3	1	3
1	3	3	3	1	1	3	3	1	3	1	3

SHARES 77 AND 78

1	1	1	1	1	1	3	2	1	1	1	0
1	1	1	1	1	3	3	2	1	1	1	1
1	3	3	3	3	3	1	2	3	3	1	1
3	1	1	1	2	3	1	1	1	1	1	3
3	3	1	3	1	3	1	2	3	1	1	1
3	3	3	1	3	3	1	1	3	3	1	1
3	3	3	3	3	1	3	3	1	3	1	3
3	3	3	3	3	3	1	2	3	3	1	3
3	1	1	1	1	3	3	1	3	1	1	1

SHARES 71 AND 45

1	1	1	1	1	3	3	1	3	1	1	0
1	1	1	1	3	1	3	3	3	1	1	1
1	1	1	1	1	3	1	3	1	3	1	1
1	3	3	3	1	3	3	3	3	3	1	3
1	1	1	1	1	3	3	2	3	1	1	3
1	3	3	3	3	3	1	1	3	3	1	1
1	3	3	3	3	3	3	3	3	3	1	3
1	1	3	1	3	3	3	3	1	1	1	3
1	1	3	3	1	3	3	1	3	1	1	3

SHARES 77 AND 79

1	1	1	1	1	3	1	3	3	1	1	0
1	3	3	3	3	3	3	2	1	3	1	3
1	1	1	1	1	3	1	2	3	1	1	3
3	3	1	1	2	3	3	3	1	1	1	1
3	1	3	3	1	3	1	2	1	3	1	1
3	1	1	1	3	3	3	1	1	3	1	3
3	3	3	3	3	1	3	3	1	3	1	3
1	1	1	1	1	1	3	2	1	3	1	3
1	3	3	3	1	1	1	1	1	3	1	3

SHARES 72 AND 45

1	3	3	3	3	1	3	1	1	3	1	0
1	1	1	1	1	1	1	3	1	1	1	3
1	1	1	1	3	1	1	3	3	1	1	1
1	3	3	3	1	1	1	3	3	3	1	1
1	1	1	1	1	1	3	2	1	3	1	3
1	1	3	1	2	3	3	3	3	1	1	1
1	3	3	3	3	3	3	3	1	3	1	1
1	1	1	1	1	1	1	3	3	1	1	3
1	1	1	1	3	3	3	3	3	1	1	1

SHARES 73 AND 45

1	3	3	3	3	3	3	3	2	3	1	0
1	3	3	3	3	3	1	3	3	3	1	3
1	3	1	1	3	3	1	3	3	3	1	3
1	3	3	3	3	3	3	3	1	3	1	3
1	3	3	1	3	3	3	1	3	3	1	1
1	3	1	3	3	3	1	3	3	3	1	3
1	3	3	3	3	3	1	3	3	3	1	3
1	1	1	1	3	3	1	3	1	1	1	3
1	3	3	3	3	3	1	1	1	3	1	3

SHARES 74 AND 45

1	3	3	3	3	3	1	3	3	3	1	0
1	3	3	3	3	2	3	3	3	3	1	3
1	3	3	3	3	3	1	3	3	1	1	1
1	3	2	3	1	3	1	3	1	3	1	3
1	3	3	3	3	3	1	2	3	3	1	3
1	1	1	1	3	1	1	1	1	1	1	3
1	1	1	1	1	3	1	3	3	1	1	3
1	3	3	3	3	3	1	3	1	3	1	3
1	3	3	3	3	3	1	3	3	3	1	3

SHARES 75 AND 45

1	1	1	1	1	3	1	3	3	1	1	0
1	1	3	3	3	3	1	1	1	3	1	3
1	1	3	1	1	2	3	3	3	1	1	3
1	1	1	1	1	1	2	3	3	1	1	3
1	1	1	1	1	1	3	2	1	1	1	1
1	1	3	1	1	1	1	1	1	3	1	1
1	3	3	3	1	3	1	3	3	1	1	1
1	1	1	1	1	1	1	3	1	1	1	3
1	1	1	1	3	3	3	3	3	1	1	3

SHARES 76 AND 45

1	1	1	1	3	3	1	3	3	1	1	0
1	3	3	3	3	1	1	3	1	3	1	1
1	3	1	1	1	1	3	1	1	3	1	3
1	1	1	1	1	1	3	3	1	1	1	1
1	3	3	3	3	3	3	2	3	3	1	1
1	1	1	1	1	1	1	3	3	1	1	3
1	3	3	3	3	3	1	1	3	3	1	3
1	3	3	3	3	1	3	3	1	3	1	3
1	1	1	1	3	3	3	3	1	1	1	1

SHARES 77 AND 45

1	1	1	1	1	3	1	3	3	1	1	0
1	3	3	3	3	3	3	3	1	3	1	1
1	3	1	1	1	3	1	3	1	2	1	1
1	1	1	1	1	3	3	3	1	1	1	1
1	1	1	3	1	3	1	2	3	1	1	1
1	3	3	3	3	3	3	1	3	3	1	3
1	3	3	3	3	1	3	1	1	3	1	3
1	1	1	1	1	1	3	3	1	1	1	3
1	1	1	1	1	3	1	1	3	1	1	1

SHARES 78 AND 45

1	3	3	3	3	3	1	3	3	3	1	0
1	3	3	3	3	1	1	3	3	3	1	3
1	1	1	1	1	1	3	3	1	1	1	3
1	3	3	3	1	1	3	3	3	3	1	1
1	1	3	1	3	1	1	2	1	3	1	3
1	1	1	3	1	1	3	3	1	1	1	3
1	1	1	1	1	1	3	1	3	1	1	1
1	1	1	1	1	1	3	3	1	1	1	1
1	3	3	3	3	1	1	3	1	3	1	3

SHARES 79 AND 45

1	1	1	1	1	1	3	3	1	3	1	0
1	1	1	1	1	1	1	3	3	1	1	1
1	3	3	3	3	2	3	3	1	3	1	1
1	1	3	3	1	3	1	1	1	3	1	1
1	3	1	1	1	2	3	2	3	1	1	3
1	3	3	3	1	3	3	3	3	3	1	1
1	3	1	1	1	3	1	1	3	3	1	3
1	1	3	3	1	3	3	3	1	1	1	3
1	1	1	1	1	3	3	3	3	1	1	1

A WIDER VIEW: ASPECTS OF MODELLING THE INVESTMENT ENVIRONMENT

In this final chapter we return to the starting point and mention several other approaches to the problem and make some comments on the whole field. Investment analysis seems to be unique as a study in that a measureable success can reward the research worker in more ways than one. Not even in technological development, and certainly not in pure scientific research, can one be assured of a financial return. Perhaps it is unnecessary to state this as a reason for the multiplicity of approaches and the amount of published material, but it is not a point to be dismissed easily.

This thesis has described a dichotomy of these multiple approaches into Investment Theories and Theories of Investment. The contention is that the majority of writers have concentrated on Investment Theories, i,e, how to invest well, and have been blinded by the financial rewards. The question remains: Why have these attempts failed. One can, perhaps unkindly, list failures which, as a trained scientist, are obvious. However, there are many problems in science (and especially in the Art/Sciences) which any number of different approaches have failed to solve. This applies especially in brain research (an Art/Science). but also in such diverse studies as earthquake prediction, high-energy particle physics, or the emotional effect of music. These problems <u>are</u> likely to be solved, and there is also no valid reason why the share-price prediction problem is not amenable to solution of some sort.

Two, hopefully new, approaches are outlined below, but first a recapitulation on some of the problems which are needed to put them into perspective. Problem number one is undoubtedly <u>information.</u> There is either too much of it, mostly unorganised; or there is too little of it. The companies themselves, the investment services and the press pour out a wealth of data (until processed it does not deserve the term information) and it is the full-time job of a whole office of people to keep up with the flow. However on the reverse side of the coin, much of the methodology used by big investors is unknown, as are the policies and methods of the jobbing firms - the members of the Stock Exchange. A Theory of Investment really needs both of these items of information to be at all complete. Another major problem has wider significance for sociology where it tries to deal with large groups of people. There is as yet no method of extrapolating and developing the fuzziness of individual behaviour into a group behaviour across a group of individuals.* What this means is that any model of group behaviour (such as investment) tends to be more or less deterministic, and hence is "transparent" to random inputs. To produce random outputs, and the random walk model of share-price behaviour is undisputed here, one must feed random inputs into the model. This is unsatisfactory when attempting to build an explanatory model; for this one needs the randomness of the outputs to result from interactive within the model.

Hence the first suggestion is to produce a mechanism which will produce random outputs from non-random inputs. Luckily this is not completely virgin territory, for the computer scientists have long studied random number generators which are programmable on the deterministic computer. Some of these, relying on the properties of numbers, are very good indeed in that their output is apparently random for long stretches, according to standard statistical tests. This is not to say that these algorithms are anything like the Stock Exchange, but they do demonstrate that the required effect is obtainable. A preliminary study has been carried out using a simple model. The model, although deterministic, incorporates threshold tests, which give the model a discrete nature. The inputs too have a discrete nature. They are assumed distributions of investors (rather naive at present) into discrete classes according to their investment beliefs. The results of running this model were not too encouraging, but there was evidence of the sort of random output mentioned above. This approach was not pursued however, the one described in detail in this thesis being preferred.

This suggestion relates to the Stock Exchange. On the other side of the coin are the great mass of investors who deal with each other through the Stock Exchange.

* This is Hari Seldon's psychohistory in the "Foundation" trilogy by Isaac Asimov. i.e. it is still science fiction.

If a satisfactory model of the structure and behaviour of the investing public is to be constructed, several goals will have to be attained. The classification and estimation of distributions of classes of investors is a prime goal. There are many people who have partial knowledge of the required information here; the problem is to assemble this information in a consistent way. The second goal is to assess the investment policies of the different classes of investor and here we can take heart from Clarkson's work as well as other more recently published work on human thinking and problem-solving. If it is possible to assemble all of this information in a reasonable time, then reasonable predictions as to the reactions of the aggregate mass of investors to general (or even specific) investment situations can be made. We believe that this approach, although at first sight quite daunting, can at least be considered now as a valid one.

Such a model would be difficult to use in individual cases of particular shares. The information as to the classification of the actual share-holders would be impossible to obtain; in other words, even though we may have a list of shareholders there is very little information to be gained from it of the sort we would require. However, we might find more success if we consider not particular shares, but categories of shares; the industry categorisation is one which comes readily to mind but there may be other more useful categories. The policies of types of investor would then be described in terms of categories of shares, and not in terms of actual holdings of individual shares. Admittedly, we are then simplyfying the problem and devaluing the results as a consequence, but it is probably the best way to maximise our use of the available information.

A less ambitious approach in the same area of investor psychology has been mentioned before. We said then that the technicalist process of pattern recognition in price series has not been studied as a human problem, but that the basic philosophy has come under scrutiny through the random walk theory. We undertook an uncompleted study to try to determine whether, in fact, there are patterns, such as the technicalists claim, to be found commonly in share-price series. It would have been included in this thesis but, although the program was written and tested, lack of data-collecting facilities put an early stop to it. The basic postulate in the program is that when human-beings see a regularity in a graph, they will remember the shape of the particular pattern. If the pattern is seen again and again (perhaps in slightly modified form), then the memory is reinforced and in fact the pattern will be actively sought in further graphs. This, we believe, is a rough description of what goes on in the technicalists mind, purely at the level of physical pattern-recognition. The justification of the patterns on external (investment) grounds is something else which probably further reinforces common patterns. This idea of, in effect, simulating the technicalist using very general (non-investment) postulates is similar in spirit to Clarkson's work, and would complement it perfectly.

REFERENCES

Arbib M. A. — Brains, Machines and Mathematics 1965

Ashby W. R. — Introduction to Cybernetics 1956

Beer Stafford — Cybernetics and Management 1956

Braithwaite R. B. — Scientific Explanation 1953

Bruner J. S. Goodnow J. J. and Austin G. A. — A Study of Thinking 1956

Clarkson G. P. E. — Portfolio Selection: A Simulation of Trust Investment 1962

Clarkson G. P. E. and Meltzer. A. H. — Portfolio Selection: A Heuristic Approach Journal of Finance Dec. 1960

Cohen J. B. and Zinbarg E. D. — Investment Analysis and Portfolio Management 1967

Ellinger — Security Price Behaviour Investment Analyst May 1966

Fama E. F. — Random Walks in Stock Market Prices Investment Analyst Dec. 1965

Feigenbaum E. A. & Feldman J. — Computers and Thought 1963

Freeman P. — Some Aspects of Technical Analysis Investment Analyst Oct. 1967

Granger C. W J. — Analysis of Stock Market Price Data Journal of Institute of Mathematics and its Applications Aug. 1972

Granger C. S. J. & Morgenstern O. — The Predictability of Stock Market Prices 1970

Ivakhnenko A. G. & Lapa V. G. — Cybernetics and Forecasting Techniques 1967

Keynes J. M. — A Treatise of Probability 1921

Klir J. & Valach M	Cybernetic Modelling 1967
Lewis C.I & Langford C.U.	Symbolic Logic 1959
Lukasiewicz J.	See Lewis and Langford
Malkiel B.G. & Cragg J.	Expectations and the Structure of Share Prices: An Empirical Study 1969
Minsky M & Papert S.	Perceptrons 1969
Money Which	Sep. 1970
Moodies Investment Service Ltd.	Investment Handbook 1971
Nilsson N.J.	Learning Machines 1965
Popper K.R.	The Logic of Scientific Discovery 1935
Tarski A.	See Lewis and Langford
Trask M.	The Story of Cybernetics 1971
Weaver D & Hall M.G	Evaluation of Ordinary Shares Using a Computer Journal of Institute of Actuaries 1967
West R.R. & Tiniç, S.M.	The Economics of the Stock Market 1971
Zadeh L.A.	Complex Systems and Decision Processes IEEE Transactions on Systems, Man and Cybernetics. Jan. 1973