33

A MATHEMATICAL MODEL OF LEARNING
UNDER
SCHEDULES OF INTERRESPONSE TIME REINFORCEMENT

by
Stephen Ambler
Department of Psychology
University of Stirling

Ph.D. Thesis October 1970 - November 1972

Acknowledgements

[illegible] ever reaches completion without the [illegible] and this one in particular is no except[illegible] [illegible] extend my grateful thanks to my supervisor, [illegible] [illegible] Department of Psychology, Univers[illegible] [illegible] general guidance during the past two years, [illegible] [illegible] Department of Psychology, [illegible] [illegible] transforming my hazy [illegible] [illegible] sophisticated hardware. [illegible] [illegible] University of Stirling) and his [illegible] [illegible] a generous allowance of computer tim[illegible] [illegible] Psychology, [illegible] [illegible] experiments. [illegible] [illegible] while [illegible] [illegible] managed to impose some sem[illegible] [illegible] grammar upon it.

My many thanks.

Acknowledgements

No thesis ever reaches completion without the help of a great many people, and this one in particular is no exception. I should like to extend my grateful thanks to my supervisor, Mr. M.F. Moore (Senior Lecturer, Department of Psychology, University of Stirling) for his general guidance during the past two years, to Angus Annan (Chief Technical Officer, Department of Psychology, University of Stirling) and his staff for transforming my hazy descriptions of experimental equipment into sophisticated hardware, to Mr. Teale (Computer Manager, University of Stirling) and his staff for their cooperation and a generous allowance of computer time, and to Dr. Glyn Thomas (Lecturer, Department of Psychology, University of Stirling) for his advice on operant experiments. Finally, I should like to thank my wife, who bore my ill-humour while I wrote the following pages, and also managed to impose some semblance to English grammar upon it.

My many thanks.

CONTENTS

Summary

After a discussion of the behaviours that are found to occur under various common schedules of reinforcement, a mathematical model of learning is proposed, which covers those schedules where reinforcement can be prescribed as a function of the time between successive responses.

There are several possible conceptualisations of the model. The one considered most often is one that uses the basic notions of Stimulus Sampling Theory. The predictions of the model are tested against the known properties of ratio, interval and DRL schedules of reinforcement. In only one case do the predictions of the model run counter to accepted experimental fact.

The model is also tested against specific data from human subjects on ratio, interval, and DRL schedules. The model fits well to the asymptotic interresponse time distributions for these schedules. However the fit of the model to distributions conditional on the previous response being reinforced, or the previous response not being reinforced, range from good to very bad, suggesting that sequential effects are more complex than implied by the model.

Chapter 1

INTRODUCTION

1.1 The Experimental Background

The statement that responses take place in time expresses a fundamental characteristic of behaviour, (Skinner, 1938, pp 263-264). Responses occur at different rates, in different sequences, and with different temporal patterns, depending on the relations between the responses and other events.

Outside the experimental laboratory, responses and environments are usually highly complex, making it difficult to describe and measure the temporal relationships between events. Some of these problems can be overcome in the artificial simplicity of the laboratory. This thesis will be concerned with a set of simple situations that may be loosely described as "Free Operant Schedules of Reinforcement". To an observer, such situations can be summarised as follows:-

A subject is placed in a simple environment (e.g. an almost empty cubicle) where he may make some response. (Emit an operant - for an analysis of the term 'operant', see Schick, 1971). When this response has been emitted, a stimulus (the reinforcement) is then delivered to the subject by the experimenter, in accordance with some rule the experimenter has devised. (The schedule of reinforcement).

This description is of necessity rather vague and generalised. A complete description of the specific situation used is given in section 3.3, on experimental procedure.

Responses can be characterised by any one of their many propert -ies, e.g. speed, intensity, duration, or location. The time that elapses between between two responses, R_n and R_{n-1}, that are considered to be the same, for the purposes of experiment, is taken (on somewhat arbitrary grounds) to be a property of the later of these two responses, R_n, and is referred to as an Interresponse Time (IRT). Regarded as a property of a response, it is obvious that the IRT can be chosen as the property upon which reinforcement is made contingent. The earliest examples of this can be found in Skinner

(1938), Wilson and Keller (1953), and Sidman (1956). To follow the usage in the literature, the phrase 'reinforcement of an IRT' will be used to refer to the reinforcement of the response that terminates the given IRT, whenever the effects of reinforcement on IRT's are being discussed. However, since responses are always separated by time periods, any schedule of reinforcement, accidentally as it were, reinforces some IRT's, but not others. Does this reinforcement affect, in any systematic way, the patterns of responding on schedules where an IRT contingency is not a specific part of the schedule?

The answer seems to be that it does, and the elucidation of a specific mechanism to account for the precise pattern of results is the aim of this thesis. At a verbal level, Skinner himself (1938), used this effect to account for the different rates of responding on variable ratio (VR) and variable interval (VI) schedules, when the average number of reinforcements obtained per unit time were equalised. (A VR schedule is one where reinforcement is contingent on a variable number of responses being emitted. The only constraint is that over a long period of time, the number of responses per reinforcement must average to some predetermined value. A VI schedule is one where reinforcement is contingent on a variable amount of time having elapsed since the previous reinforcement. The only constraint is that over a long period of time, the number of reinforcements per unit time must average to some predetermined value.) Briefly, Skinner thought that since on VR, the faster a subject responds, the more often he is reinforced, and hence a VR schedule favours short IRT's. (A high rate of responding). On a VI however, the longer a subject waits before responding, the more likely he is to be reinforced, and this favours long IRT's. (A low rate of responding). The weakness of this seems to be the casual assumption of the equivalence in effect of number of reinforcements and probability of reinforcement.

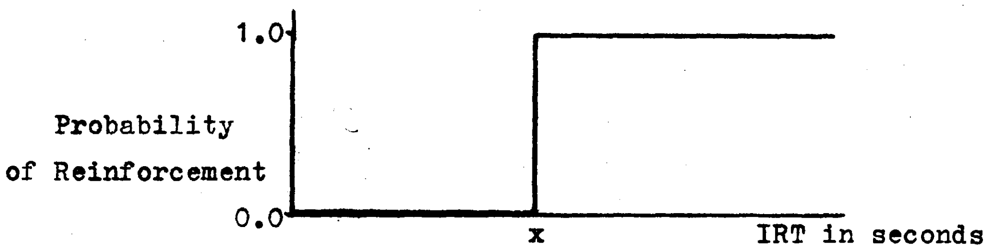
The occurrence of both deliberate and accidental IRT reinforcement has led to the establishment of two separate, but overlapping fields of investigation with respect to IRT's. The results obtained from each section will be considered separately, as follows:-

a) Schedules of IRT reinforcement.
b) Analyses of other schedules of reinforcement, in terms of the IRT distribution they produce.

a) Schedules of IRT reinforcement.

The simplest examples of reinforcement schedules that impose an IRT requirement are usually described as 'differential reinforcement of low rates of responding' (DRL). Here reinforcement is made contingent on the duration of an IRT exceeding some specified value x seconds (DRLx). If the IRT is less than x seconds, no reinforcement is given, and the timing period is begun again. (Figure 1.1.1). The most general schedule on this pattern would be one where the two probability values of the schedule were not 1.0 and 0.0, but p and q.

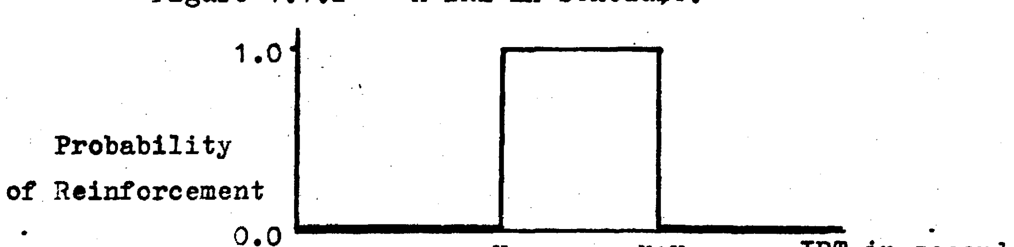
Figure 1.1.1 -- A DRL Schedule.



Such a schedule could be described as a DRLx;p,q. On such a schedule if the IRT exceeds x seconds it is reinforced with probability p. If it is less than x seconds it is reinforced with probability q. Normally it would be expected that p>q.

A variant on this type of schedule is the DRL with limited hold (DRL LH). Here the reinforcement is only made available for a period of y seconds, after the waiting period x has elapsed. IRT's of greater than x+y seconds are not followed by reinforcement. (DRLx LHy). (Figure 1.1.2).

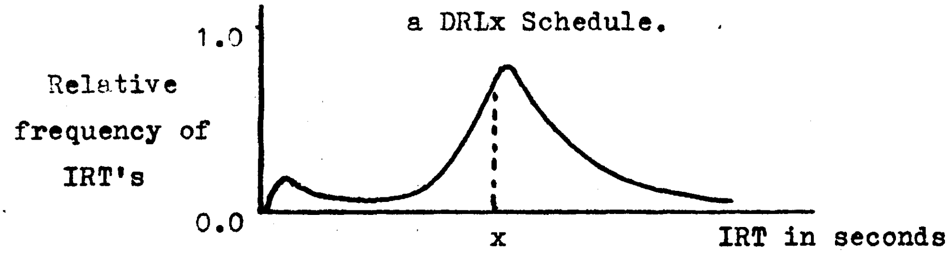
Figure 1.1.2 -- A DRL LH Schedule.



DRL type schedules are often regarded apart from other IRT reinforcement schedules in that the learning to respond under such schedules is customarily treated as the acquisition of a temporal discrimination, and work is often centred around how subjects acquire this discrimination.

The type of result usually obtained for the IRT distribution under training on a DRL schedule is illustrated in Figure 1.1.3, which is hypothetical, in the sense that it does not illustrate the results of any particular experiment. Its chief characteristics are the two peaks, one for very short IRT's, and the other around the cutoff point x.

Figure 1.1.3 -- Example of a typical IRT distribution obtained under a DRLx Schedule.



The existence of a peak around the value zero has not been confirmed by all experimenters, (e.g. Kelleher, Fry and Cook 1959, Weiss 1970) but is of fairly common occurrence and is difficult to eliminate in animal subjects.

Malott and Cumming (1964) suggest that its presence is simply due to response bias and may represent some process that is distinct from the actual conditioning process. E.g. a generalised tendency (Morman 1966) for reinforcement to produce more responses.

Millenson (1966) has argued that the effect is one due to the different susceptability to reinforcement of short IRT's. However, data by Shimp (1967) on the reinforcement of short IRT's do not clearly indicate a susceptability greater than that of long IRT's.

Sidman (1956) has offered the plausible explanation that in some way these short IRT's help set the 'internal clock' that times the long IRT's. He observed that short IRT's often occurred in bursts that followed a non-reinforced IRT that was only slightly less than the critical value required. Such bursts were then often followed by a reinforced response. It is as if such bursts of short IRT's accurately sets the zero of some timing mechanism. Unfortunat -ely, such bursts could equally be regarded as evidence of a failure to reset a timing mechanism. The response fails to reset a timing mechanism. If a response fails to set the mechanism timing the intervals between responses, then it will be read as if the response had not occurred. If such a failure, or series of failures, occurs around the crucial cutoff point, then a series of responses will ensue, as the 'clock' reads 'time to respond' and will continue to do so until reset. These explanations are only valid if the existence of some appropriate internal clock is taken for granted.

There is evidence that the second peak is not, in fact, determi-ned always by the use of an internal clock, but is fixed by what is usually referred to as collateral behaviour.

Collateral behaviour is a term used to describe stereotyped patterns of behaviour that have been observed to occupy the periods between operant responses in animals (Hodos, Ross and Brady 1962, Kelleher, Fry and Cook 1959, Laties, Weiss, Clark and Reynolds 1965) and also in humans (Brunner and Revusky 1961). It is suggested that in the process of learning to respond on DRL schedules, subjects build up an adventitious chain of responding that occupies the waiting period and leads to an accurate performance on the schedule without the need for a specific timing mechanism. Although disruption of collateral behaviour does cause a deterioration in performance on DRL schedules (Laties et al. 1965), there is evidence that collateral behaviour is in some cases not sufficient to account for the continued accuracy of DRL performance. (Laties, Weiss and Weiss 1969, Zuriff 1969).

The difficulty with the use of collateral behaviour as a mediator in timing processes, is that it does not provide, by its mere existence, evidence that it is used as a timing mechanism. In a DRL schedule, a subject is set the task of obtaining reinforcement, and the solution, to the subject (who does not possess the experimenters blinkers) is non-unique. The schedule reinforces timing behaviour, so timing behaviour may occur. On the other hand it also (even if the experimenter did not plan it) reinforces many other behaviours, and these behaviours persist because they are reinforced, not because they mediate timing. Viewed in this light, statements that disrupting collateral behaviour lowers performance on DRL schedules becomes the simple truism that disrupting behaviour disrupts behaviour.

Considerations that cast doubt on the hypothesis that collateral behaviour mediates timing behaviour, (e.g. how does it account for the temporal properties of behaviour under schedules where temporal discrimination is not an explicit part of the schedule), raise the question of why it must be assumed that timing behaviour is mediated. Under the proper set of reinforcement contingencies, properties of a response such as its force, location and duration can be selected and shaped. The IRT of a response can be treated as a conditionable property of a response (Morse 1966, p67) and timing can be regarded as the successful shaping of selected IRT's.

Timing can alternatively be interpreted as temporal discrimination. Under the proper set of circumstances, the behaviour of an organism can be brought under the control of various features of a stimulus. Duration is also a discriminable feature of a stimulus (Stubbs, 1958) and timing can be regarded as the successful discrimination of stimulus duration. The complicated stimulus from which timing is extracted by a subject on a DRL schedule is usually unidentified, and may be unidentifiable. There is no reason to suppose that the stimulus whose duration may be the discriminative stimulus for timing behaviour, must be either an internal stimulus (though this seems the most obvious source, as Anger (1963) suggests) or a chain of responses, as the collateral behaviour hypothesis maintains. Any stimulus - external, internal, or generated by behaviour - might possibly serve as the stimulus from which duration can be abstracted.

Attempts have been made by Schoenfeld, Cumming and Hearst (1956) and Hearst (1958) to define temporal analogues of the VR and VI schedules in the following manner.

A time cycle of two components, t^D and t^Δ is defined. The periods t^D and t^Δ alternate. The first response that occurs in a t^D period is reinforced. All other responses are not reinforced.

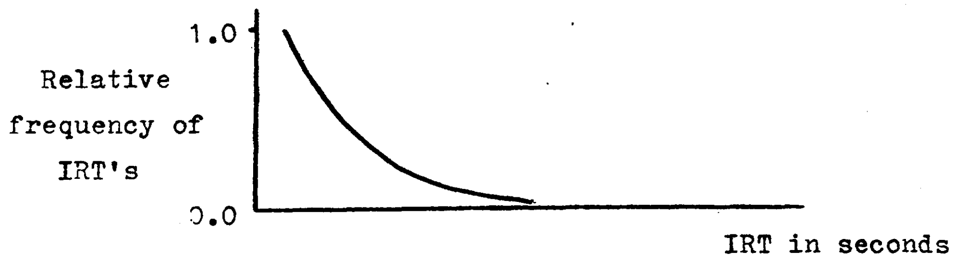
$T = t^D + t^\Delta$ is the cycle length, and $\overline{T} = t^D/T$ is the proportion of a cycle in which reinforcement may occur. If T is short, then short IRT's are likely to be reinforced, (cf. VR) and if T is moderately long, then long IRT's are likely to be reinforced, (cf. VI). Thus as T increases, there is a transition from ratio-like schedules to interval-like schedules.

Schoenfeld et al. appear to have hoped that in this manner they would be able to bring ratio and interval performance into a coherent system. However, work by other experimenters, (Millenson 1958, Clark 1959, Cumming and Schoenfeld 1959, Farmer 1963) have failed to produce such a system. Small T values do produce IRT distributions rather like those obtained from ratio schedules, and larger T values do produce IRT distributions like those obtained from interval schedules, but these cyclic schedules also appear to have peculiar properties of their own, especially when the experimental subject detects the cyclic property. This approach does not seem to have produced any really helpful insights into how the characteristic patterns of VR and VI responding are produced by the schedule.

b) Analysis of Schedules of reinforcement in terms of the IRT distributions they produce.

The first report of a serious attempt at the analysis of responding under VI or VR schedules in terms of IRT's was made by Anger (1956). More recent attempts were those of Morse (1966) and Catania and Reynolds (1968). Other experimenters have contributed some experimental results, notably Ray and McGill (1964), Kintsch (1965), Blough and Blough (1968), and Shimp (1968, 1969). The experimental results can be summarised briefly in the following sketches. (Figures 1.1.4 and 1.1.5). The actual results are affected by the classification system used to record the IRT's and estimate the relative frequencies. Many experimenters have used class intervals of four seconds width, which readily obscures the finer points of the distribution.

Figure 1.1.4 -- Example of a typical IRT distribution obtained under a VR Schedule.



Although it is possible in a gross way to account for the differences between the IRT distributions for VI and VR schedules, (e.g. Skinner 1938) researchers such as Anger (1956), Morse (1966) and Catania and Reynolds (1968), have been concerned with accounting for the specific nature of the IRT distributions, but with notable lack of success.

Figure 1.1.5 -- Example of a typical IRT distribution obtained under a VR Schedule.

Both Anger and Morse, using data from experiments by Anger, conclude that what matters is specifically which IRT's are reinforced. They differ on the choice of which aspects of the IRT distribution are affected by the distribution of reinforced IRT's. Morse chooses the IRT distribution itself, while Anger opts for a conditional function of the IRT distribution. (This he terms the IRT/Ops. function. A definition of this is given in the next paragraph). Since the IRT/Ops distribution is essentially a logarithmic transform of the complement of the cumulative IRT distribution the difference in choice can scarcely be regarded as trivial. (Contrary to Morse's opinion, footnote, Morse 1966, p67). That the data can be construed as supporting both points of view is an indication of their unrelability. Indeed, it is possible to prove that the conclusions drawn by both Anger and Morse must be false.

To do this, some notation and terminology must first be defined.

i) r(t) : the density distribution of IRT's, i.e.,

$$\Pr(t < \mathrm{IRT} \leq t + \delta t) = r(t)\delta t.$$

ii) u(t) : the reinforcement schedule, i.e.,

$$\Pr(\text{reinforcement occurs} \mid t < \mathrm{IRT} \leq t + \delta t) = u(t).$$

iii) IRT/Op.

This is an abbreviation for the interresponse time per opportunity, and is the conditional probability of an IRT in interval (a,b) given that it is longer than a.

It follows that,

$$\begin{aligned}\text{IRT/Op.}\\ \text{for interval } (a,b)\end{aligned} \quad = \quad \frac{\int_a^b r(t)dt}{\int_a^\infty r(t)dt}$$

$$= \quad \frac{R(b) - R(a)}{1 - R(a)}, \qquad (1.1.a)$$

where R(t) is the cumulative probability distribution corresponding to r(t).

iv) Reinfs/Hr.

This is the relative reinforcements per hour for IRT's in a given interval (a,b), and is the number of IRT's whose durations are between a and b that are reinforced per hour.

$$\begin{aligned}\text{Reinfs/Hr.}\\ \text{for interval } (a,b)\end{aligned} \quad = \quad \begin{bmatrix}\text{Reinforcements per IRT for}\\ \text{interval } (a,b)\end{bmatrix} \times \begin{bmatrix}\text{Number of IRT's in interval}\\ (a,b) \text{ per hour}\end{bmatrix}.$$

Now,

$$\begin{aligned}\text{Reinforcements/IRT}\\ \text{for interval } (a,b)\end{aligned} \quad = \quad \frac{\int_a^b u(t)r(t)dt}{\int_a^b r(t)dt}, \qquad (1.1.b)$$

and the average total time required for N IRT's is given by,

$$N \int_0^\infty tr(t)dt \quad = \quad N\mu_T,$$

where μ_T is the mean IRT. Of the N IRT's, a proportion (R(b) - R(a)) will be in interval (a,b), and the number of IRT's in interval (a,b) is thus, for unit time, given by,

$$\frac{N(R(b) - R(a))}{N\mu_T}$$

$$= \frac{(R(b) - R(a))}{\mu_T} . \qquad (1.1.c)$$

Multiplying equation (1.1.c) by equation (1.1.b) gives,

$$\text{Reinfs/Hr. for interval } (a,b) = \frac{R(b) - R(a)}{\mu_T} \cdot \frac{\int_a^b u(t)r(t)dt}{\int_a^b r(t)dt}$$

$$= \frac{1}{\mu_T}\int_a^b u(t)r(t)dt \qquad (1.1.d)$$

With these definitions clear, it is now possible to quote Anger's main conclusion:

"The agreement of the IRT/Op. curves with the Reinfs/Hr curves indicates that relative Reinfs/Hr., not the relative Reinfs/IRT, determines the IRT/Op. curve." Since these two curves are interdepen -dent, Anger concludes that, "Relative stability would result when the IRT/Op curve generates a Reinfs/Hr. curve that produces the same IRT/Op. curve." This conclusion is then tested by changing the reinforcement schedule and noticing its effect on both curves and "... soon the IRT/Op. curve changes until it is in rough agreement with the Reinfs/Hr. curve."

It is possible to argue as to what Anger meant by the terms "determine", and "in rough agreement", but generally they seem to have meant that the graphs of the IRT/Op. and Reinfs/Hr. functions looked very similar to the eye. The simplest mathematical equivalent of this is to assume that the functions are the same to within some linear transformation of the co-ordinates. (I.e. it is possible with at most, changes of scale and a shift of origin to make graphs of the functions look exactly alike). If it is assumed that the only implication is that there is some functional relationship, then the result is trivial. There must be some such relationship if the functions are reasonably well behaved. The result is only non-trivial if some simple relation is specified. However, Anger seems to imply a restriction on the relationship which is greater than that of a linear transformation of the co-ordinates. This is a linear transform -ation of only one of the co-ordinates, as the same time axis is required for both functions. The statement:-

"Analysis of the reinforcements given different interresponse times by the schedules shows that the Reinforcements/Interresponse time are greatest for long interresponse times, but that Reinforcements /Hr. are greatest for short interresponse times. The agreement between the greater Reinforcements/Hr. for shorter interresponse times and ... "

indicates that the time axis is assumed to correspond for all the functions used. Anger's conclusion can thus be restated as, 'at asymptote the functions IRT/Op. and Reinfs/Hr. are the same to within a linear transformation'. I.e. using (1.1.a) and (1.1.d),

$$\frac{1}{\mu_T} \cdot \int_a^b u(t)r(t)dt = \frac{A(R(b) - R(a))}{1 - R(a)} + B , \qquad (1.1.e)$$

where A and B are both constants.

Relation (1.1.e) must hold independently of the values chosen for a and b, as these are simply the result of the experimenter's choice, I.e. (1.1.e) must hold for some A and B, for any choice of a and b. Let (a,b) be the small interval $(t,t+\delta t)$.

(1.1.e) becomes,

$$\frac{u(t)r(t)\delta t}{\mu_T} = \frac{Ar(t)\delta t}{1 - R(t)} + B .$$

Simplifying this gives,

$$u(t)r(t) = \frac{Cr(t)}{1 - R(t)} + D \qquad (1.1.f)$$

where $C = A\mu_T$ and $D = B\mu_T/\delta t$.

Now,

$$0 < \int_0^{\infty} u(t)r(t)dt \leq \int_0^{\infty} r(t)dt = 1,$$

as, $0 \leq u(t) \leq 1$ ($u(t)$ is a probability), $r(t)$ is a probability density function, and assuming both $u(t)$ and $r(t)$ are non-zero for some overlapping set of t-values.

Thus, from equation (1.1.f)

$$0 < \int_0^{\infty} \frac{Cr(t)dt}{1 - R(t)} + \int_0^{\infty} Ddt \leq 1,$$

but,

$$\int_0^{\infty} \frac{Cr(t)dt}{1 - R(t)} + \int_0^{\infty} Ddt = \Big[-\log(1 - R(t)) + Dt\Big]_0^{\infty}$$

$$= \begin{cases} \infty \text{ if } D \neq -C \\ 0 \text{ if } D = -C, \end{cases}$$

neither of which lie in the required range.

Thus equation (1.1.f) cannot be true for all values of t and hence equation (1.1.e) must be false for some values of t. Anger's conclusion is necessarily false.

Morse (1966) proposed a variant on the kind of matching effect suggested by Anger. This time the actual IRT distribution and the distribution of reinforced IRT's are supposed to produce an asymptotic match. "The relative rate of reinforcement of different IRT's will in turn have an effect on the subject so that he tends to produce a distribution of total IRT's approximating the distribution of Reinforced IRT's". In a manner analogous to that used to examine Anger's (1956) statement, that of Morse may be analysed and shown to be only trivially true.

Let $v(t)$ denote the distribution of reinforced IRT's. Then,

$$v(t) = \frac{u(t)r(t)}{\int_0^{\infty} u(z)r(z)dz}.$$

Morse states that, if A and B are constants, then,

$$r(t) = Av(t) + B.$$

Integrating both sides of this equation, between zero and infinity gives,

$$\int_0^{\infty} r(t)dt = \int_0^{\infty} Av(t)dt + \int_0^{\infty} Bdt,$$

or,

$$1 = A + B\left[t\right]_0^{\infty},$$

which implies that A = 1.0 and B = 0.0, if the relation is true. I.e. that,

$$r(t) = v(t)$$

$$= \frac{u(t)r(t)}{\int_0^{\infty} u(z)r(z)dz}.$$

Therefore,

$$1 = \frac{u(t)}{\int_0^{\infty} u(z)r(z)dz},$$

which is only true if u(t) is a constant. Morse's description of the determination of the IRT distribution of reinforced IRT's is true only for a special and trivial case.

Catania (1970, p7) has stated "the fundamental determinant of differentiated responding is rate of reinforcement (reinforcements per unit time) rather than the frequency of reinforcement (reinforce -ments per response). This assumption agrees with the general consensus in the literature." The preceeding analysis of this assumption, applied to the specific versions espoused by Anger and Morse, suggests that this assumption is rather suspect.

Catania and Reynolds (1968) were much more cautious in their predict-ions. They were content to illustrate that the more probable it is that an IRT will be reinforced upon termination, the more likely it is to be terminated. Even their conclusions however, are limited by the assumption that the initiation of an IRT is equally likely to occur during any part of the schedule. There is some evidence that this is true for the initial exposure to certain schedules, (Anger 1965, Mueller 1950, Revusky 1962). With continued exposure to a schedule, the unqualified assumption that the initiation of IRT's is independent of the schedule becomes less and less tenable. (This must be true if the schedule is to affect the distribution of IRT's).

The conclusion of this section is clear, The nature of the IRT distribution produced by the most common schedules of reinforcement is known, but there is no firm theoretical formulation to suggest why certain schedules produce the particular IRT distributions that they do produce. The next section deals with some applications of existing mathematical models of learning to this problem and examines their strengths and weaknesses, to assess whether or not they can be used to provide an appropriate theoretical foundation.

1.2 Mathematical Background

To study the way in which the IRT distribution is controlled by the reinforcement of different IRT's it is necessary to consider how reinforcement schedules can be defined in terms of pure IRT reinforce-ment. Attention has so far being restricted to VR and VI and DRL schedules. How can these be defined solely in terms of IRT's and their reinforcement?

Let $u(t)$ denote a function selected by an experimenter such that $u(t)$ gives the probability that an IRT t (or more strictly an IRT in interval $(t, t+\delta t)$) is reinforced, if it occurs. Can such a $u(t)$ be defined for VR, VI and DRL schedules of reinforcement? The answer is 'yes', and it is done as follows.

i) Ratio Schedules

$$u(t) = p, \quad \forall t. \quad p \text{ is a constant.}$$

(Brandauer, 1958)

This is a special kind of ratio schedule. Ratio schedules are usually generated by selecting a set of numbers, whose mean fixes the ratio. E.g. the set 1,3,6,7,9,4, has a mean of 5. This set of numbers is then randomised and used to determine which responses are reinforced. They could be used in order 3,9,6,4,1,7 and the subject has to produce 3 responses for a reinforcement, then 9 etc. The cycle is repeated when the end is reached. On average, 5 responses have to be emitted for each reinforcement. On average each response has a probability of 0.2 of being reinforced. What the Random Ratio schedule does is make this an exact property rather than an average one. Under the parallel Random Ratio (RR) schedule, each response would have a probability of 0.2 of being reinforced.

ii) Interval Schedules

$$u(t) = 1 - \exp(-\gamma t) \qquad \forall t.$$

(Millenson, 1963)

γ is a constant. $1/\gamma$ gives the mean interreinforcement time for the scheduling (not obtaining) of reinforcements.

The problem in interval schedules has always been how to specify the nature and distribution of the intervals used. Many methods have been tried. In an Arithmetic variable interval schedule, a series of intevals of the form $a-d$, $a-2d$, ... $a-kd$, arranged in random order are used. (a and d are constants, k is some integer. The mean interval is given by $\frac{1}{2}kd$). In Geometric variable interval schedules a series of intervals of the form a, ad, ad^2, ... ad^k, arranged in random order are used. (a and d are constants, k is some integer. The mean interval is given by $(a(1 - d^{k-1}))/((k-1)(1-d))$. Many complex relationships among the intervals are possible. (E.g. Fibonacci series have been used.) All the different methods of prescribing the intervals produce slightly different patterns of behaviour. Recently interest has been aroused in what is often termed the constant probability interval achedule. This is a schedule where the probabili -ty of a reinforcement being scheduled at a time T since the previous reinforcement was scheduled, given that a period of at least T has elapsed since the previous reinforcement, is a constant. I.e.

$$\frac{\text{Probability reinforcement scheduled at T}}{\text{Probability reinforcement scheduled at T or at some later time}} = \text{a constant.}$$

This conditional probability is often referred to in the literature as, rather confusingly, 'the probability of reinforcement'. A better term, and one which illustrates the parallel with the IRT/Op. of 1.1.b(iii) is 'reinforcement per opportunity'.

The function which satisfies this relation is,

$$1 - \exp(-\gamma T) \quad \text{(Millenson, 1963).}$$

I.e. the probability that a reinforcement is scheduled in the interval (0,T) since the previous reinforcement is $1 - \exp(-\gamma T)$. (For an approximation to this see Fleshler and Hoffman, 1962).

The specific choice of intervals chosen to generate the interval schedule is of importance, as Catania and Reynolds (1968) have shown that in variable interval schedules, the rate of responding at a time T since the previous reinforcement seems to be proportional to the reinforcement per opportunity at that time. Thus constant probability schedules can be used to produce very uniform rates of responding. This is then often used as a baseline rate for other studies.

The constant probability schedule is of especial importance in the present context, as Norman (1966) has proved that an interval schedule, with intervals defined in accordance with the constant probability relationship, is also the pure IRT reinforcement schedule

$$u(t) = 1 - \exp(-\gamma t),$$

given at the beginning of this section.

This schedule will in future be referred to as the Random Interval schedule (RI), following Millenson's usage.

iii) Differential Reinforcement of Low Rates of Responding (DRL)

$$u(t) = \begin{cases} p, & t \geq d \\ q, & t < d, \end{cases}$$

where p, q, and d are constants. d is the cutoff value and p>q. In practice, the only values of p and q used are 1.0 and 0.0 respectively. The DRL schedule is naturally an IRT schedule. With the traditional values for p and q, a response is reinforced if its IRT equals or exceeds d, and not reinforced if its IRT is less than d. This schedule is naturally an IRT reinforcement schedule, and so there are no problems in framing an appropriate definition in terms of IRT reinforcement.

In the light of the above definitions of the various u(t)'s, Ratio, Interval, and DRL schedules can all be brought into the same classificatory framework. They are all schedules of interresponse time reinforcement. The only attempts to analyse all schedules as IRT schedules are the experimental works of Malott and Cumming (1964) and the chiefly theoretical paper by Norman (1966).

Malott and Cumming suggest that it is possible to analyse the effect of an arbitrary schedule u(t) by regarding it as a complex one-key concurrent IRT schedule (Shimp, 1968, 1969). The effect of the overall u(t) schedule is described in terms of the effects that the various simple schedules, into which u(t) is decomposed, have upon performance. The basic components seem to be various DRL LH schedules. Malott and Cumming provide a large amount of data on the various DRL LH schedules, but unfortunately give no examples of how these results are to be integrated to predict the outcome of some arbitrary u(t). They further omit to describe how it is decided into what basic elements a given u(t) is divided. It appears that the predictive aspect of the study breaks down under the complexity of the proposed task.

Norman (1966) has a radically different approach. He makes three very basic assumptions about the effects of reinforcement and non-reinforcement on performance. These assumptions are then translated into their mathematical equivalents, and the resulting equations subjected to extensive manipulation in order to discover the properties the model possesses.

Norman's three assumptions are,

i) Reinforcement has a general tendency to increase the rate of responding.

ii) Reinforcement of an IRT t, increases the tendency to emit IRT's of length t.

iii) Non-reinforcenent has a tendency to reduce the rate of responding.

These very simple assumptions - it is hard to think of a weaker set - are termed, in their mathematical form, the Linear Free Responding Model, by Norman. The central problem of the paper is that the very weakness and generality of the assumptions leads to very great difficulty in manipulating the equations to produce any predictions. Norman surmounts these problems to some extent, by making assumptions about the smallness of the parameters involved and produces some approximate results. The final conclusions about the relationship between the mean rates of responding and the mean rates of reinforcement scarcely seem worthy of the preceeding mathematics. Unfortunately the model is too difficult to handle with ease.

Let r(t) denote the asymptotic IRT distribution. (I.e. the probability of an IRT whose duration lies between t and $t+\delta t$ is approximately $r(t)\delta t$.) The problem is, very simply stated,

Given a reinforcement function u(t), what is the asymptotic IRT distribution, r(t), that a subject will eventually produce?

Norman's model - the linear free responding model - has been rejected as a solution to this problem, as it makes the problem too intractable. Can an alternative be found? Norman essentially considered that the IRT was itself the response reinforced, and that as such the response could take any value between zero and infinity. Thus the response is essentially a response on a continuum. This suggests that other models for responding on a continuum may be applicable to the present problem. These are:-

a) Linear models for Responses Measured on a Continuous Scale. (Anderson 1964).

This model deals only with the mean response on a given trial. It is thus a model for group effects. The function r(t) is an individual function. This model is thus unsuitable for application to the problem here.

b) Linear Model for a Continuum of Responses.
(Suppes 1959, Suppes and Rouanet 1961, Suppes and Zinnes 1961, Suppes, Rouanet, Levine and Frankmann 1964).

Formally, this model can be regarded as a simple version of the Linear free responding model of Norman. Essentially it is the Norman model without assumptions (i) and (iii). Briefly the model can be described as follows. (For a complete formal specification, see the paper by Suppes, 1959).

An experiment consists of a sequence of responses (x_n) and reinforcements (y_n), denoted by s_n, upto and including the nth trial. (The nth response followed by the nth reinforcement.) I.e.

$$s_n = (y_n, x_n, y_{n-1}, x_{n-1}, \ldots\ y_1, x_1).$$

S_n is a history of responses x and reinforcements y. On any trial i, the subject made response x_i and a reinforcement y_i followed, indicating that y_i would have been the correct response. The basic axiom of the model is,

$$j_{n+1}(x|s_n) = (1-\theta)j_n(x|s_{n-1}) + \theta k(x;y_n), \qquad (1.2.a)$$

where $j_{n+1}(x|s_n)$ is the density distribution of the response x on trial n+1, given a history s_n, θ is a constant between zero and one, and k(x;y) is a density distribution on x with a mode at y, the point of reinforcement. k(x;y) is known as the smearing distribution, because it smears the effect of a reinforcement at y over the portion of the x-continuum around the value y. The basic axiom simply states that the effective density distribution of x on trial n+1 is the weighed average of the density distribution on trial n, with a tendency to repeat responses around the reinforced response y_n.

The model assumes that every trial is followed by a reinforcement, so to keep matters simple, consider the application of the model to the case of continuous reinforcement. When the continuum x is the time t, this is a random ratio schedule with p = 1.0. I.e,

$$u(t) = 1.0 \qquad \forall t.$$

The conditional probability density distribution, on trial n, is thus,

$$f_n(y|t)\delta t \;=\; \Pr(\text{Reinforcement occurs at } y \mid \text{response is in interval } (t, t+\delta t))$$

$$= u(t)\delta(t-y)\delta t$$

$$= \delta(t-y)\delta t, \qquad \forall n,$$

where $\delta(t)$ is the Dirac δ-function.

The mean response distribution on trial n+1 is given by,

$$r_{n+1}(t) \;=\; \int j_{n+1}(t, s_n)ds_n,$$

i. e. $j_{n+1}(t,s_n)$ is summed over all possible past histories s_n. ($j(\ldots)$ is used as an omnibus function. It denotes the density distribution of whatever appears inbetween the brackets.)

Now,

$$\int j_{n+1}(t,s_n)ds_n \;=\; \int j_{n+1}(t,y_n,t_n,s_{n-1})dy_n dt_n ds_{n-1}$$

$$= \int j_{n+1}(t|y_n,t_n,s_{n-1})j_n(y_n,t_n,s_{n-1})dy_n dt_n ds_{n-1}$$

$$= (1-\theta)\int j_n(t|s_{n-1})j_n(y_n,t_n,s_{n-1})dy_n dt_n ds_{n-1}$$

$$+ \;\theta\int k(t;y_n)j_n(y_n,t_n,s_{n-1})dy_n dt_n ds_{n-1}$$

(from basic axiom)

$$= (1-\theta)\int j_n(t|s_{n-1})j(s_{n-1})ds_{n-1}$$

$$+ \;\theta\int k(t:y_n)f_n(y_n|t_n)j_n(t_n,s_{n-1})dy_n dt_n ds_{n-1}$$

$$= (1-\theta)r_n(t)$$

$$+ \;\theta\int k(t:y_n)\delta(t_n-y_n)j_n(t_n,s_{n-1})dy_n dt_n ds_{n-1}$$

$$= (1-\Theta)r_n(t) + \Theta\int k(t;y)r_n(y)dy.$$

At asymptote this becomes,

$$r(t) = (1-\Theta)r(t) + \Theta\int k(t;y)r(y)dy,$$

therefore,

$$\underline{r(t) = \int k(t;y)r(y)dy.} \qquad (1.2.b)$$

This equation is a homogeneous Fredholm equation of the second kind. Solution of this equation, in the traditional manner, by series expansion of $k(t;y)$, yields the obvious and trivial solution,

$$r(t) = \text{a constant}, \qquad \forall t,$$

as $k(t;y)$ is a probability density function. When the further constraint, that $r(t)$ be a probability density function is added, it can be seen that no such $r(t)$ exists with the required properties. The difficulty stems basically from the following property of the model, viz. that all responses t are on some trial reinforced and hence ultimately all responses are equally likely. As the response continuum is the interval $(0,\infty)$, the actual probability associated with a particular response is arbitrarily close to zero. Thus instead of the high rate of responding actually found under continuous reinforcement, the model predicts a near zero rate of responding. The model does not seem directly applicable to the description of behaviour under schedules of interresponse time reinforcement.

c) Stimulus Sampling Theories for a Continuum of Responses. (Suppes 1959, Suppes and Frankmann 1961, Suppes and Zinnes 1966).

For simplicity consideration will be limited to the one-element model, and as far as possible the notation will be the same as that used in section (b).

The stimulus element has associated with it a smearing distribution $k(x;y)$ which is a density distribution with a mode at y, and a variance independent of y. If a response t_n is made and is followed by a successful reinforcement y_n, on trial n, then the mode of the smearing distribution shifts from its previous value to y_n. Note that,

$$r(t \mid \text{mode of smearing distribution at } y) = k(t;y).$$

Let response t_n occur on trial n, followed by reinforcement y_n. Let $g_n(z)$ denote the density distribution of the probability that the mode of the smearing distribution is at z on trial n. The model states that, if conditioning is not effective, (with probability $1-\Theta$), then,

$$g_{n+1}(z \mid t_n) = g_n(z),$$

and if reinforcement is effective, (with probability Θ),

$$g_{n+1}(z \mid t_n) = f_n(z \mid t_n),$$

where $f_n(z \mid t_n)$ has the same meaning and form as in section 1.2.(b).

Combining these equations gives,

$$g_{n+1}(z \mid t_n) = (1-\Theta)g_n(z) + \Theta f_n(z \mid t_n). \qquad (1.2.c)$$

Now $r_{n+1}(t)$ satisfies,

$$\begin{aligned} r_{n+1}(t) &= \int r_{n+1}(t \mid z, t_n) g_{n+1}(z \mid t_n) r_n(t_n) dt_n dz \\ &= \int k(t;z) g_{n+1}(z \mid t_n) r_n(t_n) dt_n dz \\ &= (1-\Theta)\int k(t;z) g_n(z) r_n(t_n) dt_n dz \\ &\quad + \Theta \int k(t;z) f_n(z \mid t_n) r_n(t_n) dt_n dz, \end{aligned}$$

(using equation 1.2.c)

$$\begin{aligned} &= (1-\Theta) r_n(t) \\ &\quad + \Theta \int k(t;z) r_n(z) dz, \end{aligned}$$

when $f_n(z|t_n) = \delta(z-t_n)$, as in section 1.2.(b). (I.e. it is again assumed that the schedule of reinforcement is the continuous reinforcement schedule, or RR 1.0).

At asymptote,

$$r(t) = (1-\theta)r(t) + \theta\int k(t;z)r(z)dz,$$

therefore,

$$\underline{r(t) = \int k(t;z)r(z)dz.} \qquad (1.2.d)$$

This result is of the same form as that found in section 1.2.(b). Again the result is meaningless in the present context. The difficulty is similar to that found with the linear model. All responses have the same probability of reinforcement, and hence are finally all equally likely to be the mode of the smearing distribution.

d) Stimulus Sampling Theory for Continuous-Time Processes: Extension to a continuum of responses.
(Suppes and Donio 1967).

This model is a model for tasks such as monitoring a particular tone, or tones. The response 'listening for frequency f' can be regarded as responding on a continuum, as the frequencies f form a continuum. The model is concerned with predicting the response made at a given time, There are conceptual difficulties in applying this model to schedules of interresponse time reinforcement, as time is both the response, and the period in which the response takes place. (E.g. it is not easy to speak of an IRT of duration t being made at time T, since an IRT is not actually specified until it terminates. However, it is possible to think that each time the subject sets out to wait until a period t has elapsed since the previous response. This waiting could be underway at some time T.)

Let T be some baseline time, running say, from the beginning of the experiment. Let t denote the IRT. For simplicity again restrict the model to a single element. Let g(z,T) denote the density distribution of the mode z, of the smearing distribution k(t;z), at time T. This parallels the function $g_n(z)$ of section 1.2.(c). The continuous parameter T replaces the discrete parameter n. r(t,T) parallels $r_n(t)$ and denotes the density distribution of the probability that an IRT of duration t is in the process of passing at time t.

Obviously,

$$r(t) = \lim_{T\to\infty} r(t,T),$$

and,

$$r(t,T) = \int k(t,z)g(z,T)dz.$$

Consider a small interval $(T, T+\delta T)$. In this interval a reinforcement will occur, provided a response terminating an IRT occurs. The probability that this has been of duration t is, $r(t,T)\delta T$. Thus if reinforcement is effective, with probability Θ,

$$g(z, T+\delta T\ t) = f(z,T|t)$$

$$= f(z|t).$$

($f(z|t)$ has the same meaning as before. It is the conditional reinforcement function. This function does not depend on T).

If reinforcement is not effective, with probability $(1-\Theta)$, then,

$$g(z,T+\delta T|t) = g(z,T|t).$$

Finally, if no response occurs, with probability $(1-r(t,T)\delta T)$,

$$g(z,T+\delta T|t) = g(z,T|t).$$

Combining these equations gives the mean result,

$$g(z,T+\delta T|t) = g(z,T|t) + \Theta r(t,T)(f(z|t)-g(z,T|t))\delta T,$$

which as δT tends to zero, becomes the partial differential equation,

$$\frac{\partial g}{\partial T}(z,T|t) = +\Theta r(t,T)(f(z|t)-g(z,T|t)).$$

With an inital boundary condition denoted by $g(z,0|t)$ this equation has the solution,

$$g(z,T|t) = f(z|t) + (f(z|t)-g(z,0|t)) .\exp(-\Theta\int_0^T r(t,y)dy).$$

At asymptote,

$$\lim_{T\to\infty} g(z,T|t) = g(z|t)$$

$$= f(z|t),$$

(from above).

Hence,

$$r(t) = \lim_{T\to\infty} r(t,T)$$

$$= \lim_{T\to\infty} \int k(t,z)g(z,T)dz$$

$$= \int k(t,z)g(z)dz$$

$$= \int k(t,z)r(z)dz,$$

because,

$$g(z) = \int g(z|t)r(t)dt$$

$$= \int f(z|t)r(t)dt$$

but,

$$f(z|t) = \delta(z-t)$$

(the schedule is RR 1.0).

Therefore,

$$g(z) = r(z).$$

Thus the result is again,

$$\underline{r(t) = \int k(t;z)r(z)dz.}$$

This model must also be rejected, as again, like (b), and (c) it predicts results which do not in any way correspond to the experimental evidence available.

In conclusion, all the models outlined above, in various amounts of detail, face (with the exception of Norman's model) one central problem when applied to IRT schedules of reinforcement. They predict that at asymptote, responses that are equally likely to be reinforced are equally likely to occur. The simplest example of an IRT reinforcement schedule, RR 1.0 has been used to illustrate this. Here every response is equally likely to be reinforced. It is well known however, that on such schedules, very short IRT's vastly outnumber long IRT's.

As well as having this common failing, the models above all make a common assumption, that the IRT is itself the response. In the next chapter a model is developed, which , while related to the stimulus sampling models outlined above, does not make this assumption. Finally, in the ensuing chapters, this model is put to test against experimental data.

Chapter 2

THE MODEL

2.1 The Stimulus Process

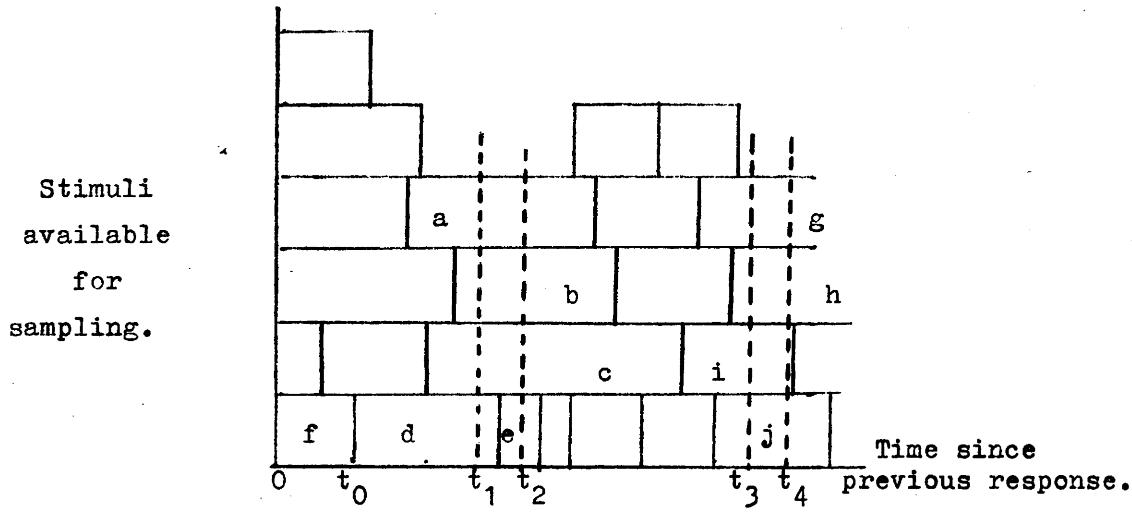
"Animals have available some events, either internal or in their behaviour, that change in a consistent way with time after the last response, reinforcement, etc. These events function like external stimuli, at least to the extent that differences in responding can be conditioned to these organism differences."

The above quotation is taken from Anger, (1963). In a very basic way, all that follows is rooted in the assumption that this statement is essentially true.

Assume that any response generates, or is associated with a pattern of stimuli X that change over time. If t denotes the time since the last response occurred, the pattern found at time t will be called X(t). The changing pattern X(t) can be conceptualised in many ways -- a stimulus trace, the reading of some internal clock, the position in a chain of collateral responses -- the particular conceptualisation is not important. What is required however, is that changes in X(t) are consistent, in the sense that, after two responses, R_1 and R_2, which are considered to be the same, then the patterns X(t) are the same for each value of t, for each response. (It is possible to weaken this assumption, without seriously affect -ing the results that follow. It does however introduce a somewhat needless complexity. The weakening is done by partitioning X(t) into two parts, a consistent part, and a randomly varying part.) Perhaps the most general way to envisage the situation is to consider that X(t) consists of a set of stimuli. Changes in X can then be described as the appearance, or disappearance of stimuli from this set. The following diagram may help clarify the ideas about the stimuli X(t). (Figure 2.1.1).

A stimulus is available as a component of a pattern for a period of time. E.g. stimulus f (figure 2.1.1) is available during interval $(0,t_0)$. At any time t, since the last response, a characteris -tic set X(t) of all the then available stimuli exists. These sets are not all unique. E.g. $X(t_3) \equiv X(t_4)$. The set actually detected by a subject at any time t will be called $x(t)$, and is a subset of the total available set $X(t)$. E.g. (figure 2.1.1)

Figure 2.1.1 -- A diagramatic interpretation of the Stimulus Patterns available after each Response.



Stimuli crossed by a vertical dotted line indicate the particular stimuli available at that particular time to form $X(t)$.

$$X(t_1) \quad = \quad (a,b,c,d), \text{ while}$$

$$x(t_1) \quad = \quad (a,b,c), \text{ or possibly}$$

$$x(t_1) \quad = \quad (a,b,d), \text{ etc.}$$

Differences in sets $x(t)$ can be used to discriminate among values of t, but not perfectly reliably. E.g. if,

$$x(t_1) \quad = \quad (a,b,c,d)$$

and

$$x(t_2) \quad = \quad (a,b,c,e),$$

then t_1 is distinguished from t_2, but if,

$$x(t_1) \quad = \quad (a,b,c)$$

and

$$x(t_2) \quad = \quad (a,b,c)$$

then t_1 will not be distinguished from t_2. The overall pattern $X(t)$ is the same after every response, but the associated pattern $x(t)$, that is the one detected, can vary within limits. Variations in $x(t)$ describe trial by trial variations in the accuracy of temporal discrimination, while the actual nature of $X(t)$ fixes an upper limit to the accuracy of temporal discrimination.

2.2 The Response Process

Given that there exists this continuum of patterns X, indexed by t, how is the occurrence of a response, at a time t since the previous response, determined. Let each pattern $x(t)$ have associated with it a probability of responding, ϕ. This value of ϕ, $\phi(x(t))$, can be regarded as the proportion of the stimulus elements in the pattern $x(t)$ that are conditioned to the appropriate response. The average value of $\phi(x(t))$ at a given t will be denoted by $\phi(t)$, i.e.,

$$\phi(t) \quad = \quad \Pr(\text{response} \mid \text{some pattern } x(t)).$$

Although $\phi(t)$ is a response probability, it does not directly determine the occurrence of a response, as it is really a conditional probability. To reduce confusion, $\phi(t)$ will usually be referred to as the response strength function. It is to be assumed that the set of patterns $X(t)$ does not in any way affect responding unless the subject is actually sampling from the stimulus set $X(t)$. This assumption is very important, as it enables the model being developed to avoid the pitfalls of the models discussed in section 1.2.

It is a truism that organisms do not spend all their time on a single activity. It will be assumed that there are many different sources of stimuli, and that the organism samples from these different sources, or continua, in accordance with some rule. The occurrence of a response depends both on the appropriate continuum being sampled and the local response probability for the pattern found upon sampling. It will be assumed, for simplicity, though perhaps unrealistically, that the response considered always has a zero probability of occurrence for continua other than $X(t)$. Let $\psi(t)$ denote the probability that the continuum X is sampled at time t since the previous response, i.e.,

$$\psi(t)\delta t \quad = \quad Pr(\text{sampling occurs in interval } (t, t+\delta t)).$$

Notice that $\phi(t)$ can now be redefined as,

$$\phi(t) \quad = \quad Pr\,(\text{response occurs} \mid \text{sampling occurs in interval } (t,\ t+\delta t)).$$

With $\phi(t)$ and $\psi(t)$ defined, it is important to realise that the response considered is essentially regarded as an instantaneous process. In the next paragraph a prediction for the distribution of times between successive responses is developed. A crucial difference between this model and other models of interresponse times may be stated here. The interresponse time is not of itself the response. The existence of interresponse times is the result of,

i) Responses themselves take time to occur. (This aspect is being temporarily neglected here.)

ii) Organisms do not spend all their time executing a single response. (Though sometimes they may try hard to!). Behaviour is variable, and between any two responses, regarded as being identical, other responses occur.

Let $r(t)$ denote the density distribution of IRT's. I.e.,

$$Pr(t < IRT \leq t+\delta t) \quad = \quad r(t)\delta t,$$

and let $R(t)$ be the corresponding cumulative distribution. Then if a response occurs in period $(t,\ t+\delta t)$,

$$R(t+\delta t) \;-\; R(t) \quad = \quad (1 - R(t))\psi(t)\delta t\phi(t),$$

as a response must not have occurred in the interval $(0,t)$, with a probability $(1 - R(t))$, a sampling must have occurred in period $(t, t+\delta t)$, with a probability $\psi(t)\delta t$, and given a sampling occurred, the response actually occurred with a probability $\phi(t)$. As t tends to zero, this equation becomes,

$$\frac{dR}{dt} \quad = \quad (1 - R(t))\psi(t)\phi(t), \qquad \bullet$$

which has solution,

$$R(t) = \frac{1 - \exp(-\int_0^t \psi(y)\phi(y)dy)}{1 - \exp(-\int_0^\infty \psi(y)\phi(y)dy)}$$

If it is assumed that, as is quite likely, the integral given in the exponent in the denominator does not converge, then the denominator is simply one, and,

$$r(t) = \psi(t)\phi(t)\exp(-\int_0^t \psi(y)\phi(y)dy) \qquad (2.2.a)$$

If $\phi(t) = k$, say, a constant, and $\psi(t) = \rho$, as a simple example, it is possible to see easily the effect that the introduction of sampling has on the model. It makes,

$$r(t) = \rho k \exp(-\rho k t).$$

By allowing the possibility of not responding, even if k were 1.0, it transforms a constant response probability into an exponential distribution of interresponse times.

2.3 The Learning Process

Having established how a response probability $\phi(t)$ can be translated into details about the time between responses, it becomes appropriate to ask how the values of $\phi(t)$ are determined. $\phi(t)$ is determined by the conditioning history of the subject in the following manner.

The experimental situation can be described as consisting of a series of trials, $i = 1,2,3,\ldots$, the trials being demarcated by the subjects behaviour. Responses and reinforcements are assumed to be instantaneous events. A particular trial, q_i, can be described by an ordered pair,

$$q_i = \langle k_i, t_i \rangle,$$

where the suffix i indicates that this is the ith trial, i.e the interval t_i between the (i-1)th and the ith response. If the response that terminated the trial after time t_i is reinforced, k_i is one; if otherwise it is zero. Thus k_i tells whether or not the ith response was reinforced.

Let,

$$s_i = \langle q_{i-1}, q_{i-2}, \ldots, q_1 \rangle,$$

i.e. the history upto the ith trial. Then $\phi_i(t|s_i)$ gives the response strength function after a particular history of responding and reinforcing, described by s_i.

Initially it will be assumed that non-reinforcement has no effect. Thus, if a response is not reinforced on trial i,

$$\phi_{i+1}(t|t_i, s_i) = \phi_i(t|s_i). \qquad (2.3.a,i)$$

If reinforcement occurs on trial i, but is not effective,

$$\phi_{i+1}(t|t_i, s_i) = \phi_i(t|s_i), \qquad (2.3.a,ii)$$

and finally if reinforcement occurs and is effective,

$$\phi_{i+1}(t|t_i, s_i) = \phi_i(t|s_i) + (1 - \phi_i(t|s_i))w(t;t_i). \qquad (2.3.a,iii)$$

This rather complex looking equation needs explanation of its origin. Remember that at time t, there is a pattern of stimuli X(t) available for sampling. X(t) is regarded as a set of stimuli. Let a proportion of these stimuli, (which proportion gives the response probability) $p_i(t)$ be conditioned on trial i. Let the sample x(t) taken from X(t) contain a proportion σ of the total stimulus elements of X(t), Then, if θ is the probability that a sampled element is conditioned, given the response is reinforced,

$$p_{i+1}(t) = p_i(t) + ((1 - p_i(t))\sigma)\theta,$$

on average, for $(1 - p_i(t))\sigma$ represents the proportion of unconditioned elements in the sample x(t), and θ is their probability of becoming conditioned.

The exactly analogous equation is, using the present notation,

$$\phi_{i+1}(t_i|t_i,s_i) \quad = \quad \phi_i(t_i|s_i) + (1 - \phi_i(t_i|s_i))w,$$

where $w = \sigma$.

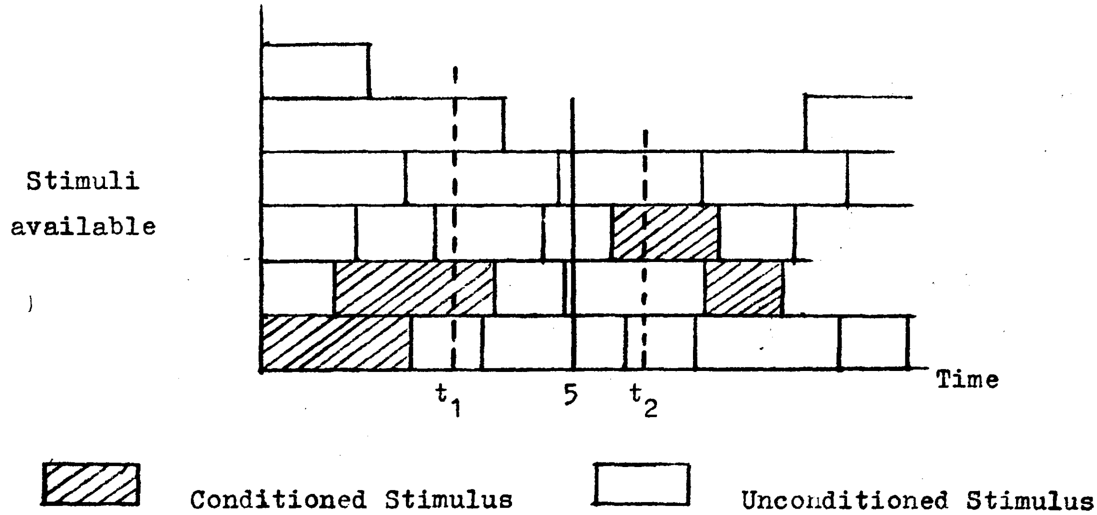
However, this is an inadequate model of the conditioning process as it restricts the effect of reinforcement only to the point $t = t_i$. Consider in practice the pattern $X(t_i+\delta t_i)$, a pattern adjacent to $X(t_i)$. This will contain some stimuli common to $X(t_i)$. (See figure 2.1.1). Thus reinforcement in the presence of $x(t_i)$ affects not only the future response probability at t_i, but also at $t_i+\delta t_i$. Thus the equation must be modified to spread out the effect of reinforcement, i.e. it becomes,

$$\phi_{i+1}(t|t_i,s_i) \quad = \quad \phi_i(t|s_i) + (1 - \phi_i(t|s_i))w(t;t_i).$$

$w(t;t_i)$ is termed the spread function. It takes values between 0.0 and 1.0, and is assumed to have a maximum at t_i, the point of reinforcement.

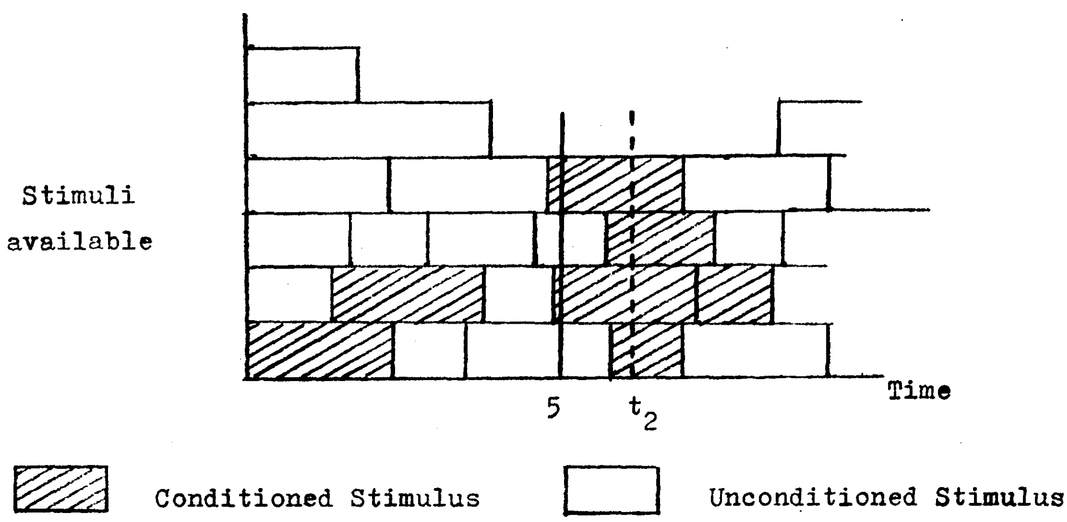
The following diagrammatic explanation may help clarify these ideas. Assume that the subject is on a DRL5 schedule, and that figure 2.3.1 represents the stimulus configuration and its state of conditioning at some time early in exposure to this schedule.

Figure 2.3.1 -- Representation of an early state of Conditioning.



Suppose that a sampling of stimuli takes place at t_1 and all the elements available are sampled. There is one element conditioned so that figure 2.3.1 shows a response probability of 1/5. Assume no response is made, but that a further sampling is made at t_2, where the response probability can be seen to be 1/4 (figure 2.3.1). Assume this time that a response is made. As $t_2 > 5$, a reinforcement will occur, and it will be assumed that reinforcement is effective for all the elements available at t_2. The state of conditioning becomes as in figure 2.3.2.

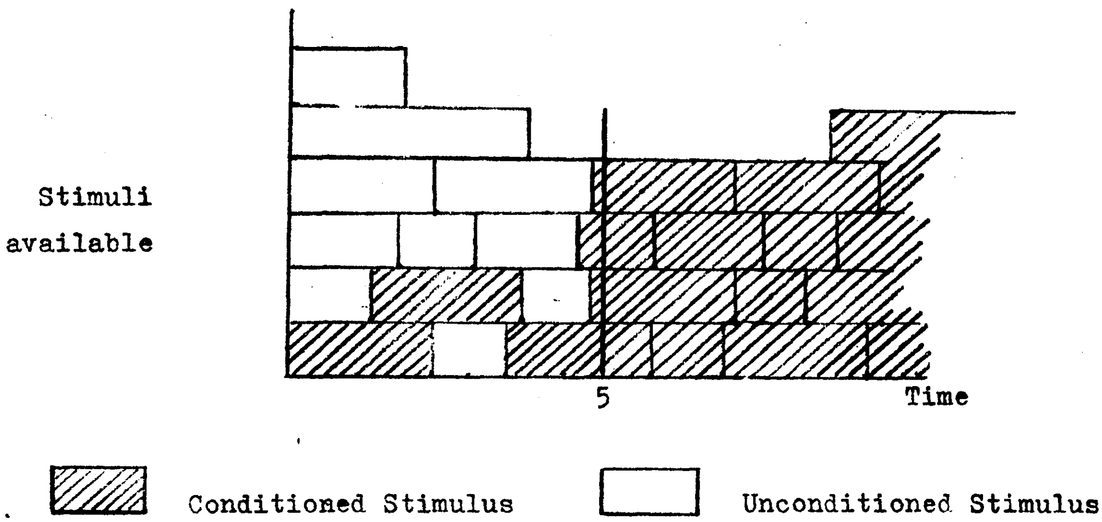
Figure 2.3.2 -- Representation of state of Conditioning after Reinforcement at t_2.



Notice that in figure 2.3.2, not only does reinforcement affect the stimulus complex at t_2, but also at adjacent values of t, as the conditioned stimuli occupy finite periods of time. Thus the effect of conditioning is spread out. This spreading effect is the effect described by the function $w(t:t_i)$ in the learning equation.

It is obvious that after a lot of responding and reinforcing under a DRL5 schedule that the state of conditioning should be as illustrated in figure 2.3.3. In this figure, all the stimuli that are available when $t > 5$ have become conditioned. In this situation, if a sample is taken for any t greater than 5, then a response will be made. Note that although responses with IRT's less than 5 seconds are not, and will not be reinforced, such responses will still be emitted. This is because they represent responses to stimuli that

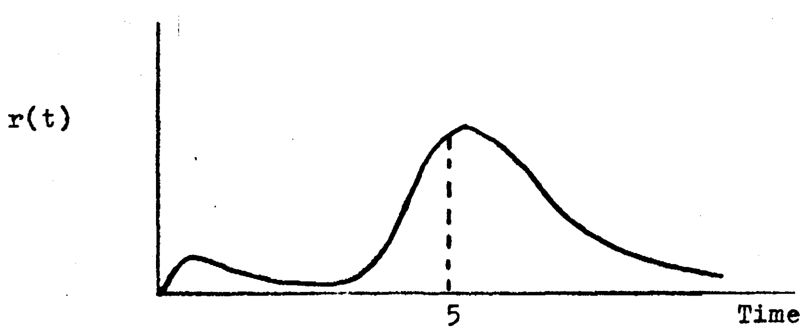
Figure 2.3.3 -- Representation of the final state of Conditioning.



were conditioned before the start of training, or they represent responses to stimuli that overlap the cutoff value 5, and which have thus been reinforced as parts of samples taken for IRT's greater than 5 seconds.

Equation 2.2.a says roughly that the IRT distribution is an exponential transform of $\phi(t)$, the response strength function. Using figure 2.3.3, $\phi(t)$ can be found for any t. (It is the proportion of stimuli conditioned). The exponential transform is graphed roughly in figure 2.3.4. This result is, qualitatively at least, like the known experimental results, which were illustrated

Figure 2.3.4 -- IRT distribution corresponding to the state of Conditioning illustrated in Figure 2.3.3.



in figure 1.1.3.

Before the previous equations for the trial by trial transforma -tions of $\emptyset(t)$ (equations 2.3.a,i,ii,iii) can be combined to give a mean equation, two further functions need to be defined. These are,

i) $c_i(t)$.

This is the probability that reinforcement will be effective on trial i, following an IRT of duration t.

ii) $u_i(t)$.

This is the reinforcement schedule,

$$u_i(t) \quad = \quad \Pr(\text{Reinforcement occurs on trial } i \mid t < IRT \leq t + \delta t).$$

It is the dependence of $u_i(t)$ upon t that makes the schedules considered interresponse time reinforcement schedules, though under the model explicated here it is not actually the interresponse times that are reinforced, Examples of IRT schedules are,

Random Ratio

$$u_i(t) \quad = \quad p, \qquad \forall i,t.$$

Random Interval

$$u_i(t) \quad = \quad 1 - \exp(-\gamma t), \qquad \forall i,t.$$

Differential Reinforcement of Low Rates of Responding

$$u_i(t) \quad = \quad \begin{cases} p & \text{if } t \geq d \\ q & \text{if } t < d, \end{cases} \qquad \forall i,t.$$

p,q,γ and d are constants in the above equations.

With this notation,

$$\emptyset_{i+1}(t \mid t_i, s_i) \quad = \quad \emptyset_i(t \mid s_i) \quad \text{with probability,}$$

$(1 - u_i(t_i))$ -- Non-reinforcement occurs.

$$\phi_{i+1}(t|t_i,s_i) = \phi_i(t|s_i) \quad \text{with probability,}$$

$(1 - c_i(t_i))u_i(t_i)$ -- Reinforcement occurs, but is not effective.

$$\phi_{i+1}(t|t_i,s_i) = \phi_i(t|s_i) + (1 - \phi_i(t|s_i))w(t;t_i)$$

with probability, $c_i(t_i)u_i(t_i)$ -- Reinforcement occurs and is effective.

Combining these equations gives,

$$\underline{\phi_{i+1}(t|t_i,s_i) = \phi_i(t|s_i) + (1 - \phi_i(t|s_i))u_i(t_i)c_i(t_i)w(t;t_i).}$$

If $j(s_i)$ gives the distribution of histories s_i, then taking expectations over s_i gives,

$$\int\phi_{i+1}(t|t_i,s_i)j(s_i)ds_i = \int\phi_i(t|s_i)j(s_i)ds_i + \int(1 - \phi(t|s_i))u_i(t_i)c_i(t_i)w(t;t_i)j(s_i)ds_i.$$

Therefore,

$$\underline{\phi_{i+1}(t|t_i) = \phi_i(t) - (1 - \phi_i(t))c_i(t_i)u_i(t_i)w(t;t_i)}$$

If $r_i(t)$ is the distribution of interresponse times t on trial i, taking expectations over t_i gives,

$$\int_0^\infty\phi_{i+1}(t|t_i)r_i(t_i)dt_i = \int_0^\infty\phi_i(t)r_i(t_i)dt_i + \int_0^\infty(1-\phi_i(t))u_i(t_i)c_i(t_i)w(t;t_i)r_i(t_i)dt_i.$$

Therefore,

$$\underline{\phi_{i+1}(t) = \phi_i(t) + (1 - \phi_i(t))z_i(t),} \qquad (2.3.a)$$

where,

$$z_i(t) = \int_0^\infty u_i(t_i)c_i(t_i)r_i(t_i)w(t;t_i)dt_i.$$

If this equation is written in the form,

$$\phi_{i+1}(t) = (1 - z_i(t))\phi_i(t) + z_i(t),$$

it can be compared with equations 1.2.a and 1.2.c of chapter 1. The important distinction lies in the replacement of the $(1-\Theta)$ of these equations, by the $(1 - z_i(t))$ of the above. The replacement of the constant parameter Θ by the function $z_i(t)$ represents a change from the fixed set of stimuli, and the fixed conditioning process, to one where the stimuli, and the conditioning process, change with time.

If asymptotic convergence for $\phi_i(t)$ is assumed, then as i tends to infinity, 2.3.a becomes, dropping the suffixes,

$$\phi(t) = \phi(t) + (1 - \phi(t))z(t),$$

so that,

$$0 = z(t)(1 - \phi(t)),$$

and either,

$$z(t) = 0,$$

or,

$$\phi(t) = 1.$$

z(t) will not in general be zero, though it may be so, if for some values of t, the component functions have ranges where they are zero, and these ranges overlap to occupy the whole continuum. Essentially z(t) will be zero for those regions where reinforcement is not effective. These regions are not those where u(t) is zero, (no reinforcement given to these IRT's) but regions where reinforce -ment has no influence. Some regions, although associated with non -reinforcement, have the effects of reinforcement spread into them by the action of the w(t;t') function. In general, z(t) will be zero whenever,

$$\int_0^\infty u(t')w(t;t')dt' = 0,$$

for all other regions, $\phi(t)$ will be one. This conclusion produces the slightly unrealistic prediction that all u(t), such that u(t)>0.0 for all t, produce the same asymptotic response strength function, and hence the same type of IRT distribution. This contradicts the available evidence somewhat, e.g. Random Ratio and Random Interval schedules of reinforcement produce rather different types of IRT distribution.

This suggests that the model requires modification, and an obvious and interesting modification would be to allow non -reinforcement to have an effect similar, but in the opposite direction to reinforcement. All the functions in the following development have the same meanings as indicated in the previous sections. The superfix 1 is used to show when they refer to reinforcement, and the superfix 0 to show when they refer to non-reinforcement. The equations for the effect of reinforcement become,

$$\phi_{i+1}(t|t_i,s_i) = \phi_i(t|s_i),$$

with probability $(1 - c_i^1(t_i))u_i^1(t_i)$ -- Reinforcement occurs and is not effective.

$$\phi_{i+1}(t|t_i,s_i) = \phi_i(t|s_i) - (\alpha^1 - \phi_i(t|s_i))w^1(t;t_i),$$

with probability $c_i^1(t_i)u_i^1(t_i)$ -- Reinforcement occurs and is effective.

The analagous equations for non-reinforcement are,

$$\phi_{i+1}(t|t_i,s_i) = \phi_i(t|s_i),$$

with probability $(1 - c_i^0(t_i))u_i^0(t_i)$ -- Non-reinforcement occurs and is not effective.

$$\phi_{i+1}(t|t_i,s_i) = \phi_i(t|s_i) - (\alpha^0 - \phi_i(t|s_i))w^0(t;t_i),$$

with probability $c_i^0(t_i)u_i^0(t_i)$ -- Non-reinforcement occurs and is effective.

Notice that in these equations,

$$u_i^0(t_i) = 1 - u_i^1(t_i).$$

A further modification (in excess of the introduction of a non-reinforcement effect) will have been noticed in the previous equations. This is the introduction of α^1 and α^0. The α^1 replaces a previous "1" and α^0 an expected "0". The interpretation of these two terms is simple. As will be proved later, they form bounds on $\phi(t)$, for,

$$\alpha^0 \leqslant \phi(t) \leqslant \alpha^1.$$

The psychological interpretation of α^0 and α^1 can be as follows. Every stimulus set X(t) can be divided into two parts, in the proportions, a consistent proportion, $(\alpha^1 - \alpha^0)$, and a random fluctuation proportion, $(1 - \alpha^1 + \alpha^0)$. (E.g. see Estes and Burke 1953, Estes 1956). Only the proportion $(\alpha^1 - \alpha^0)$ is generally manipulable, in practice, by the conditioning and deconditioning process, Of the remainder, a proportion $\alpha^0/(1 - \alpha^1 + \alpha^0)$ is always conditioned to the response, (e.g. this could be the 'operant level') and a proportion $\alpha^1/(1 - \alpha^1 - \alpha^0)$ are effectively never conditioned. The proportion $(1 - \alpha^1 - \alpha^0)$ is assumed to vary randomly from trial to trial, so that even if some of these stimuli are conditioned on any one trial, they are unlikely to be present on the next trial to affect the response strength,

These two parameters are useful at a verbal level to account for temporary, but not very short-term, fluctuations in behaviour. E.g.,

i) Spontaneous recovery:-

Intensive non-reinforcement may drive $\phi(t)$ below the normal levels of α^0, by forcing out of the X(t) patterns, all conditioned stimuli. Cessation of training may allow a substantial change in the fluctuation proportions, and some of these stimuli may be conditioned to the response. Responding may thus start again, or move to levels higher than that when extinction was discontinued.

ii) Fatigue:-

Fatigue may have the effect of flooding sets X(t) with non-conditioned stimuli, that vary from trial to trial. Thus α^1 will fall, but as the organism recovers, the $\phi(t)$ values can re-assert themselves, and responding begins again.

There are however other ways of explaining the effects of fatigue.

It could be suggested that fatigue affects $\Psi(t)$, to reduce the rate of sampling, and hence the rate of responding.

iii) Satiation:-

Parts of the sets $X(t)$ may be deprivation associated stimuli. As reinforcement is obtained, these stimuli may disappear, thus lowering the effective value of α^1. The behaviour will then become less probable and the rate of responding decrease.

Let,

$$k_i = \begin{cases} 1 & \text{if reinforcement occurs on trial } i \\ 0 & \text{if reinforcement does not occur on trial } i, \end{cases}$$

and,

$$l_i = \begin{cases} 1 & \text{if conditioning/deconditioning is effective on trial } i \\ 0 & \text{if conditioning/deconditioning is not effective on trial } i. \end{cases}$$

Let,

$$c_i^{k_i 0}(t) = 1 - c_i^{k_i}(t),$$

and

$$c_i^{k_i 1}(t) = c_i^{k_i}(t).$$

With this notation, the equations describing the various possible transformations of $\phi_{i+1}(t|t_i,s_i)$ can be summarised by the single equation,

$$\phi_{i+1}(t|t_i,s_i) = \phi_i(t|s_i) - (\alpha^{k_i} - \phi_i(t\ s_i))l_i w^{k_i}(t;t_i),$$

with probability, $u_i^{k_i}(t_i)c_i^{k_i l_i}(t_i)$.

Taking expectations over l_i, k_i, s_i, and t_i gives,

$$\phi_{i+1}(t) = \phi_i(t) + \sum_{k_i}(\alpha^{k_i}-\phi_i(t))\int_0^\infty w_i^{k_i}(t;t_i)u_i^{k_i}(t_i)c_i^{k_i}(t_i)r_i(t_i)dt_i.$$

Assuming the asymptotic convergence of $\phi_i(t)$, then,

$$\sum^{k}(\alpha^k - \phi(t))\int_0^{\infty} w^k(t;y)u^k(y)c^k(y)r(y)dy \quad = \quad 0.0$$

and hence,

$$\phi(t) \quad = \quad \frac{\int_0^{\infty}\sum^{k}\alpha^k w^k(t;y)u^k(y)c^k(y)r(y)dy}{\int_0^{\infty}\sum^{k} w(t;y)u^k(y)c^k(y)r(y)dy} \qquad (2.3.b)$$

To obtain a feel for the kind of function $\phi(t)$ is, for a given u(t), put,

$$\theta \quad = \quad \frac{c^0(t)}{c^1(t)} \qquad \forall t,$$

and

$$w^k(t;y) \quad = \quad \begin{cases} 1 \text{ if } t = y \\ 0 \text{ if } t \neq y. \end{cases}$$

(This value of $w^k(t;y)$ is obviously unrealistic in terms of the psychological background of the model as it makes the spread effect zero. This does not matter for the moment however, as it can be regarded simply as an approximation used only to give an idea of the nature of $\phi(t)$.)

Introduction of these values gives,

$$\phi(t) \quad = \quad \frac{\alpha^0\theta(1-u(t)) + \alpha^1 u(t)}{\theta(1-u(t)) + u(t)} \ .$$

If $\theta = 0.0$, (i.e. non-reinforcement is not effective) then,

$$\phi(t) \quad = \quad \alpha^1.$$

(Unless $u(t) = 0.0$ also, when $\phi(t) = \alpha^0$)

If $\Theta = \infty$, (i.e. reinforcement is not effective) then,

$$\phi(t) = \alpha^0.$$

Otherwise, $\phi(t)$ lies between these two values. If these two values are in fact zero and one, then,

$$\phi(t) = \frac{u(t)}{\Theta(1-u(t)) + u(t)}$$

$$= \frac{u(t)}{\Theta + (1-\Theta)u(t)} .$$

For the special value $\Theta = 1.0$, this gives $\phi(t) = u(t)$. It is only under these very restrictive conditions that any kind of matching will occur. Since, (equation 2.2.a)

$$r(t) = \psi(t)\phi(t)\exp(-\int_0^t \psi(y)\phi(y)dy),$$

this suggests that it is highly unlikely that any simple matching law will relate the forms of r(t) and u(t).

Equations (2.2.a) and (2.3.b) are the fundamental results derived from the model. These two equations give implicit determinations of the functions $\phi(t)$ and r(t). Only for very simple u(t) do these equations have explicit solutions for $\phi(t)$ and r(t). Before even these cases can be considered however, it is necessary to investigate in greater detail the spread function w(t;t') and the sampling function $\psi(t)$.

2.4 The Spread Function

From the heuristic description of the conditioning process given in figures 2.3.1 to 2.3.3, it is obvious that the spread function is highly variable in form, depending on how the various stimuli available are assumed to interact with one another. It would seem likely that w(t;t') could be related to the limiting form for the generalisation gradient for temporal discrimination. It is still a matter for controversy as to whether or not generalisation gradients

(Generalisation gradients give the probability of a stimulus s eliciting a response that a subject has been trained to produce to stimulus s', when s and s' lie on some common continuum. The response to s' is said to generalise to s.) are innate or learned. (Terrace 1966). However it would not be unreasonable to consider that w(t;t') might be the function that described the sharpest possible stimulus control available at t'. If this were so, then experimental investigation of temporal discrimination could elucidate the spread function.

In the absence of any adequate data to specify the spread function, it would appear best to choose some function that is either extremely tractable, or which satisfies some assumed theoretical constraints. Several possibilities are open. For example, an interesting and relatively simple choice might be,

$$w(t;t') = \begin{cases} p, & (1-k)t' \leq t \leq (1+k)t' \\ 0, & \text{otherwise,} \end{cases}$$

where p and k are constants.

The spread effect then occupies a region kt' in width on either side of t'. The spread effect is directly proportional to t'. This is equivalent to saying that the greater the time since the last response, the greater the persistence of a present stimulus. Since discrimination is dependent on changes in X(t), this is equivalent to saying that discrimination becomes more difficult as time passes. At time t since the last response, a period of kt must elapse, on average, before X(t) changes, and hence before a discrimination of time passage is made. This is similar to proposing that temporal discriminations follow a Weber's Law with parameter k.

Another possibility would be to choose,

$$w(t;t') = \frac{p.\exp\left[\dfrac{-(t-t')}{1/\sqrt{2\pi}+kt'}\right]^2}{1+kt'\sqrt{2\pi}}$$

where k and p are again positive constants. (The constant p, in both these examples gives the ratio of the number of elements in the sampled set x(t), to the maximum number of elements available, which is the number of elements in X(t)). The spread effect is here normally distributed around a mean of t' and with a standard deviation of $(1 + kt'\sqrt{2\pi})/\sqrt{2\pi}$. (The $1/\sqrt{2\pi}$ factor is simply to make $w(0;0) = p$. This is the maximum value $w(t;t')$ can take for any pair of values (t,t').) Making the standard deviation of the spread effect proportional to t' is again equivalent to making a Weber's type law hold for temporal discriminations. Also however, as t' increases, the maximum value of $w(t;t')$ falls in accordance with $p/(1 + kt'\sqrt{2\pi})$. Thus reinforcements for small t values have a proportionally greater effect than those for large t values. (This is an effect that has been suggested by Millenson, 1963, to account for the persistence of short IRT's on some DRL schedules.)

It is not necessary that $w(t;t')$ should be symmetric about t', though this is usually convenient. The simplest symmetric function is,

$$w(t;t') = \begin{cases} p & t'-a \leq t \leq t'+a \\ 0 & \text{otherwise,} \end{cases}$$

where p and a are constants.

This is merely a rectangular function of fixed width 2a. It is chosen for its mathematical simplicity and tractability. In terms of figure 2.3.1, all the stimuli are assumed to be roughly equivalent in duration, and any pattern x(t) contains a large number of stimuli.

In all the applications of the model, the following further simplifications and notations will be used.

$$w^1(t;t') = w^0(t;t').$$

The spread effect is the same for both conditioning and deconditioning. This is the natural assumption to make, as they are presumed to refer to the same sets of stimuli.

$$c^1(t) = c^1,$$

and

$$c^0(t) = c^0.$$

The effectiveness of conditioning and deconditioning is a constant.

independent of the value of t.

Let,

$$pc^1 = \theta_1,$$

and

$$pc^0 = \theta_0.$$

Equation (2.3.b) now becomes,

$$\phi(t) = \frac{\int_{t-a}^{t+a} \sum^{k} \alpha^{k}\theta_k u^k(y)r(y)dy}{\int_{t-a}^{t+a} \sum^{k} \theta_k u^k(y)r(y)dy}. \qquad (2.3.c)$$

This form for $\phi(t)$ will be used in all further developments of the model.

2.5 The Sampling Function

So far, very little has been said about the sampling function $\psi(t)$, except to assume that it is independent of the previous response history, and of the trial number. I.e.

$$\psi_i(t|s_i) = \psi(t) \quad \forall i, s_i,$$

and thus $\psi(t)$ is not manipulated by the learning process. This assum -ption is made on the following grounds.

If the sampling process is manipulated by learning, sampling must be regarded as the response component of some process. This response (sampling from X(t)) is a response to certain stimuli. These stimuli are themselves presumably sampled by some higher (higher in the sense of more distant from the observed responses) mechanism. However, could not this sampling also be manipulated by learning? This leads to an infinite hierarchy of stimuli and sampling, which is undesirable. It appears better to assume that sampling is not affected by learning.

The chief difficulty in accepting this argument is that it is not intuitively unreasonable, on other grounds, to let $\psi(t)$ be affected by a learning process. In some sense, $\psi(t)\delta t$ could be said to give the probability of attending to the appropriate stimulus dimension in the interval $(t, t+\delta t)$. Thinking of the sampling process in terms of an attentional process readily suggests that any parameters of $\psi(t)$ might easily be manipulated by factors such as learning, previous exposure to IRT schedules, etc. This attentional idea of sampling seems closely related to the notion of an observing response, and observing responses can be conditioned. (Wyckoff, 1952).

The final choice of $\psi(t)$ rests chiefly on its simplicity. In many ways, t is not the natural variable to index ψ with. When dealing with responses, IRT's are easier to manipulate than other variables which could be used to describe the distribution of responses in time. (E.g. it would be possible to use time since the start of the experiment, and investigate this distribution, or to use time since reinforcement, etc.) Similarly, it is preferable for the sampling process to use, if possible, intersampling times, rather than time since the last response, to specify ψ.

Let T denote an intersampling time, and let $\psi^S(T)$ denote the density distribution of intersampling times. The basic assumption is that,

$$\psi^S(T) \quad = \quad \rho.\exp(-\rho T),$$

i.e. that intersampling times follow a poisson process with parameter ρ.

Let the probability that the kth sampling takes place in interval $(t, t+\delta t)$ be denoted by $\psi^S_k(t)\delta t$. Then,

$$\psi(t) \quad = \quad \sum_{k=1}^{\infty}\psi^S_k(t).$$

However, $\psi^S_k(t)$ is the convolution of $\psi^S(T)$ with itself k times, and thus,

$$\psi^S_k(t) \quad = \quad \frac{\rho(\rho t)^{k-1}}{(k-1)!}\mathrm{Exp}(-\rho t).$$

Therefore,

$$\psi(t) \quad = \quad \sum_{k=1}^{\infty}\frac{\rho(\rho t)^{k-1}}{(k-1)!}\mathrm{Exp}(-\rho t)$$

$$= \rho \sum_{k=0}^{\infty} \frac{(\rho t)^k \mathrm{Exp}(-\rho t)}{k!}$$

$$= \rho$$

Thus the assumption that intersampling times are exponentially distributed (i.e. random in time) leads to the very simple expression for (t). $\psi(t)$ is a constant for all t. This assumption will be utilised in the following applications of the model, and ρ will always refer to the parameter of the sampling distribution.

With this value of $\psi(t)$ the expression for the IRT density distribution becomes,

$$r(t) = \rho\phi(t)\mathrm{Exp}\left(-\int_0^t \rho\phi(y)dy\right) \qquad (2.5.a)$$

2.6 Some Asymptotic IRT Distributions

It is perhaps convenient to collect together here the basic equations for the asymptotic IRT distribution. Their present simple form, accepting the conclusions of the previous two sections, is,

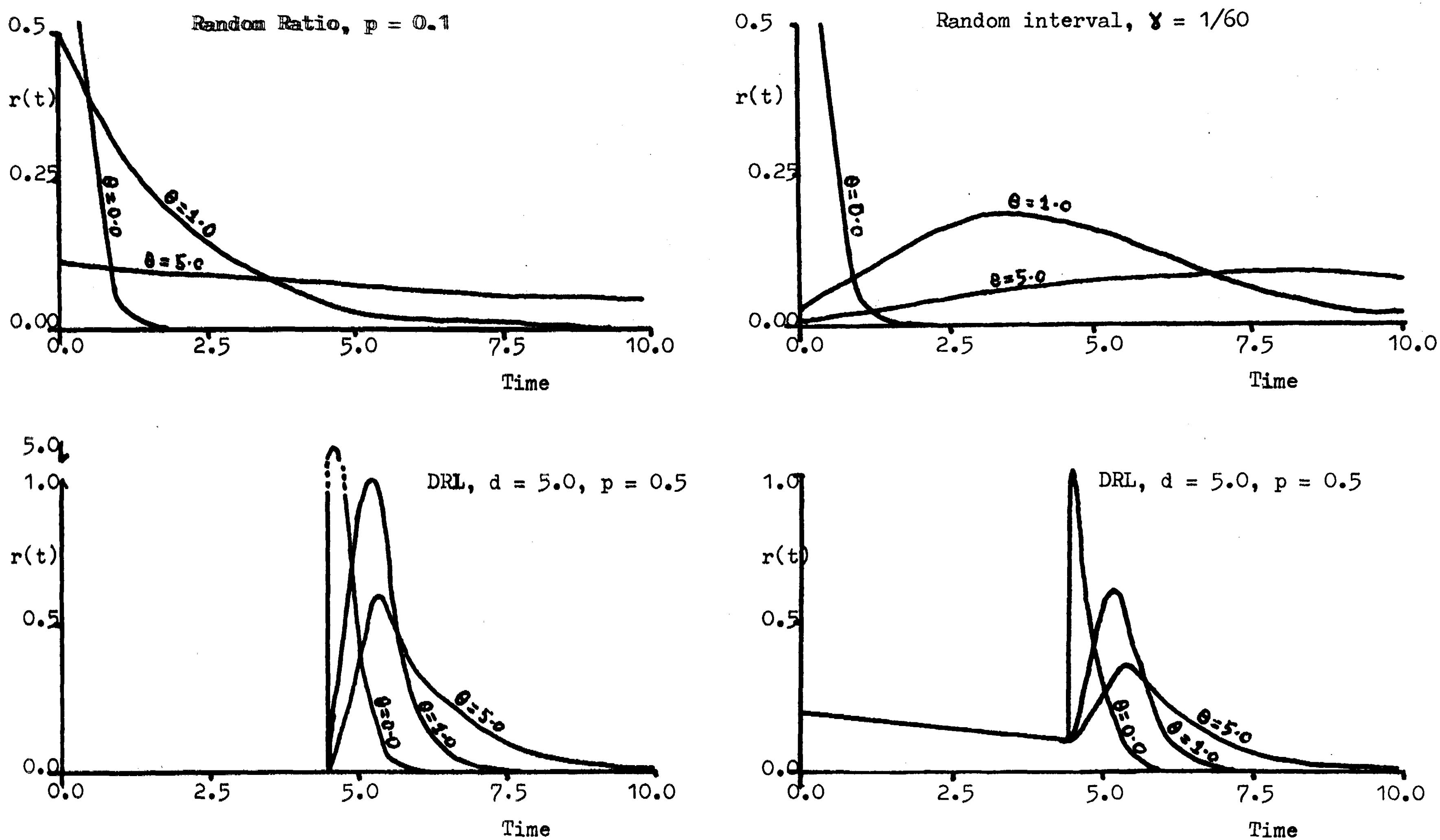
$$\phi(t) = \frac{\alpha^0 \Theta \int_{t-a}^{t+a} (1-u(y))r(y)dy + \alpha^1 \int_{t-a}^{t+a} u(y)r(y)dy}{\Theta \int_{t-a}^{t+a} (1-u(y))r(y)dy + \int_{t-a}^{t+a} u(y)r(y)dy}, \qquad (2.6.a)$$

where $\Theta = \Theta_0/\Theta_1$, and represents the relative effectiveness of reinforcement and non-reinforcement, and,

$$r(t) = \rho\phi(t)\mathrm{Exp}\left(-\int_0^t \rho\phi(y)dy\right). \qquad (2.6.b)$$

The difficulty with these equations is that r(t) is defined in terms of $\phi(t)$ and $\phi(t)$ is defined in terms of r(t). This pair of equations is not, in general, soluble for r(t), for an arbitrary reinforcement schedule u(t), though solutions do exist for certain u(t). An iterative method was devised, for finding r(t), given u(t),

Figure 2.6.1 — Some theoretical IRT distributions, as predicted by the model for various reinforcement schedules.



$\Psi(t)$, $w(t;t')$ and α^0, α^1, and Θ. This was set up as an algol programme, SAGENAR, which is, together with some comments as to its mode of operation, given in Appendix I.

Table 2.6.i -- Model and Schedule parameters used to calculate the curves of Figure 2.6.1. Values of Θ were variable (V) and are given on corresponding curves.

Schedule	Schedule Parameters	Model Parameters				
		α^0	α^1	Θ	a	ρ
RR p	p = 0.1	0.0	1.0	V	0.5	5.0
RI γ	γ = 1/60	0.0	1.0	V	0.5	5.0
DRL d;p,q	d=5.0 p=0.5 q=0.0	0.0	1.0	V	0.5	5.0
DRL d;p,q	d=5.0 p=0.5 q=0.0	0.02	1.0	V	0.5	5.0

Figure 2.6.1 gives some IRT distributions calculated for various reinforcement schedules from equations 2.6.a and 2.6.b. Table 2.6.i gives the details of the parameter values used. These graphs can be used to give an idea as to the general properties of the model. It can be seen that as Θ increases, (non-reinforcement becomes more effective, or reinforcement less effective) that the overall rate of responding falls, illustrated by the general flattening of the curves.

A comparison of ratio and interval predictions shows that interval schedules favour longer IRT's than do ratio schedules. Ratio schedules always have their mode at zero, while interval schedules have a mode at progressively larger values of t as Θ increases.

The predictions for the DRL schedules show a striking peak in the region of the cutoff point. The interesting point to notice is that if the effect of reinforcement is greater than that for non-reinforcement, then the peak tends to lie to the left of the cutoff value. A large number of the IRT's emitted are just less than the required value for reinforcement, and performance is relatively inefficient. If however the effect of non-reinforcement is greater than the effect of reinforcement, then the peak tends to lie to the right of the cutoff value, most IRT's are reinforced and performance is relatively efficient. One of the characteristic differences between pigeon and rat performance on DRL schedules is that pigeons tend to produce an IRT distribution whose peak is to the left of the cutoff point, while rats tend to produce a distribution whose peak is to the right of the cutoff point. Interpreted through the framework of the present model, this indicates that rats are relatively more sensitive to the effects of non-reinforcement than are pigeons, and suggests for example that superstitious conditioning (Herrnstein 1966) should be more frequent in occurrence for pigeons, where the effects of the occasional spurious reinforcement will only slowly extinguish through non-reinforcement, than for rats.

The attempt to account for the peak near zero in the IRT distribution from a DRL schedule (by setting $\alpha^0 \neq 0.0$) does not appear too successful, as the gradient of this region is too small. On the whole, the model seems promising, however.

2.7 Some Conditional Statistics

Generally,

$$r(t|s_i) = \rho\phi(t|s_i)\mathrm{Exp}\left(-\int_0^t \rho\phi(y|s_i)dy\right),$$

so that finding conditional statistics for the IRT distribution reduces to the problem of finding conditional statistics for the response strength function. Conditional statistics of particular interest are those distributions of IRT's which follow reinforcement or non-reinforcement. The following section proves an interesting property of the means of these distributions.

Let,

$$k = \begin{cases} 1 & \text{if reinforcement occurs,} \\ 0 & \text{if non-reinforcement occurs.} \end{cases}$$

Then,

$$\phi(t|k) = \phi(t) + (\alpha^k - \phi(t))\theta_k \int_{t-a}^{t+a} r(y)dy.$$

Now if reinforcement has no effect, i.e. $\theta_1 = 0.0$, then from equation (2.6.a),

$$\phi(t) = \alpha^0.$$

Similarly, if non-reinforcement has no effect, i.e. $\theta_0 = 0.0$, then from equation (2.6.a),

$$\phi(t) = \alpha^1.$$

Thus,

$$\alpha^0 \leq \phi(t) \leq \alpha^1,$$

and hence,

$$\alpha^0 - \phi(t) \leq 0 \leq \alpha^1 - \phi(t).$$

As,

$$\theta \geq 0.0,$$

and,

$$\int_{t-a}^{t+a} r(y)dy \geq 0.0,$$

then the three previous results taken together imply that,

$$\phi(t|k) \leq \phi(t) \leq \phi(t|1) \qquad (2.7.a)$$

At asymptote, the mean IRT is given by,

$$\begin{aligned}\int_0^\infty tr(t)dt &= \int_0^\infty t\,\rho\theta(t)\mathrm{Exp}(-\int_0^t \rho\phi(y)dy)dt \\ &= \Big[-t\mathrm{Exp}(-\int_0^t \rho\phi(y)dy)\Big]_0^\infty \\ &\quad + \int_0^\infty \mathrm{Exp}(-\int_0^t \rho\phi(y)dy)dt.\end{aligned}$$

The first term on the RHS is zero, giving,

$$\text{Mean IRT} = \int_0^\infty \mathrm{Exp}(-\int_0^t \rho\phi(y)dy)dt.$$

By analogy,

$$\text{Mean (IRT}|k) = \int_0^\infty \mathrm{Exp}(-\int_0^t \rho\phi(y|k)dy)dt \qquad (2.7.b)$$

From 2.7.a, as ρ is always positive,

$$\rho\phi(t|0) \leq \rho\phi(t) \leq \rho\phi(t|1).$$

Integrating this inequality over the interval (0,t) gives, as the functions are always positive,

$$\int_0^t \rho\phi(y|0)dy \leq \int_0^t \rho\phi(y)dy \leq \int_0^t \rho\phi(y|1)dt.$$

Therefore,

$$\mathrm{Exp}(-\int_0^t \rho\phi(y|0)dy) \geq \mathrm{Exp}(-\int_0^t \rho\phi(y)dy) \geq \mathrm{Exp}(-\int_0^t \rho\phi(y|1)dy),$$

and as these are always positive valued,

$$\int_0^\infty \mathrm{Exp}(-\int_0^t \rho\phi(y|0)dy)dt \geq \int_0^\infty \mathrm{Exp}(-\int_0^t \rho\phi(y)dy)dt \geq \int_0^\infty \mathrm{Exp}(-\int_0^t \rho\phi(y|1)dy)dt.$$

Hence,

$$\mathrm{Mean}(IRT|0) \geq \mathrm{Mean}\ IRT \geq \mathrm{Mean}(IRT|1).$$

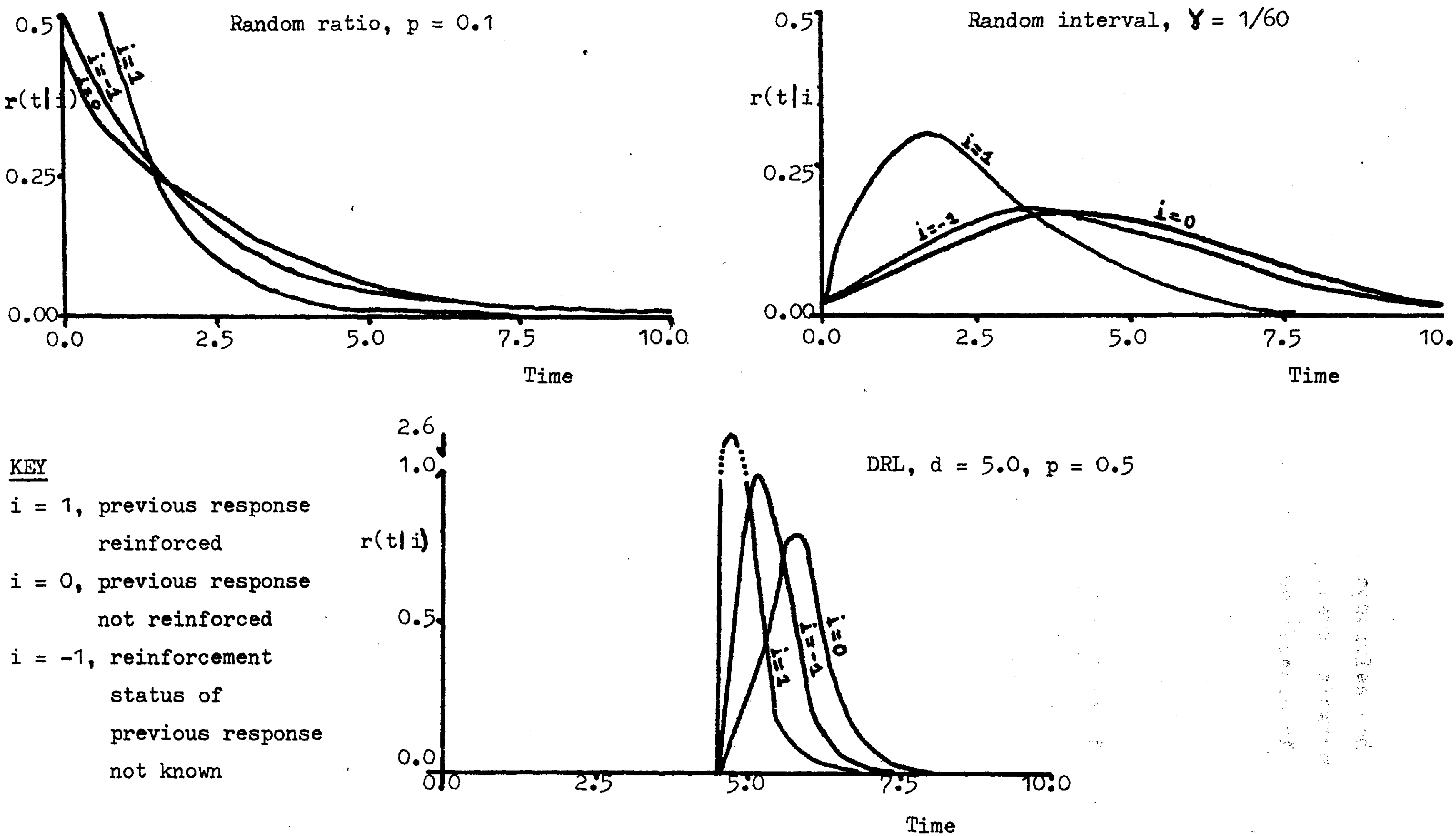
Thus, on average, responses which follow reinforcement tend to have shorter IRT's than responses which follow non-reinforcement. Thus a general effect of reinforcement is to speed up responding,

Table 2.7.i -- Model and Schedule parameters used to calculate the curves of Figure 2.7.1.

Schedule	Schedule Parameters	Model Parameters					
		α^0	α^1	θ_0	θ_1	a	ρ
RR p	p = 0.1	0.0	1.0	0.5	0.5	0.5	5.0
RI	= 1/60	0.0	1.0	0.5	0.5	0.5	5.0
DRL d;p,q	d=5.0 p=0.5 q=0.0	0.0	1.0	0.5	0.5	0.5	5.0

while a general effect of non- reinforcement is to slow down responding. These two effects have often been noticed in operant conditioning procedures. Indeed, Norman (1966) used these specific characteristics as the starting point for his linear model for free responding. In the present model these effects arise naturally from the properties of the learning process.

Figure 2.7.1 -- Some theoretical IRT distributions, as predicted by the model for various reinforcement schedules. These distributions are conditional on some preceeding event, specified by the key.



KEY

i = 1, previous response reinforced

i = 0, previous response not reinforced

i = -1, reinforcement status of previous response not known

A graphical illustration of this effect is given in figure 2.7.1. (These distributions were calculated by an algol programme SACONDAR (see Appendix I) a modification of the SAGENAR programme.) All the schedules (Table 2.7.i gives details of the schedules and the parameter values used.) show quite clearly the general slowing down effects of non-reinforcement, and the speeding up effects of reinforcement.

2.8 Other Measures of the IRT Distribution

Anger (1956) introduced the IRT/Op. function as a measure of the IRT distribution, and defined it as the conditional probability of an IRT in interval (a,b), given that it is longer than a.

It follows that,

$$\text{IRT/Op. for interval } (a,b) = \frac{\int_a^b r(t)dt}{1 - \int_0^a r(t)dt}$$

The use of the interval (a,b) was simply a practical artifact to estimate r((a-b)/2) and so the function is really of the form,

$$I(t) = \frac{r(t)}{1 - R(t)}.$$

Care must be taken when dealing with I(t) in this form, as it obscures the origin of the function somewhat. I(t), considered as a function of t is not a probability density function, for t occupies not only the role of the random variable, but also the role of the condition applied. Anger thought that this function might yield a clearer understanding of the nature of interresponse time reinforcement, as he felt that it was difficult to speak of the effects of reinforcement unless the IRT's considered had had the opportunity to occur.

In terms of ρ and $\phi(t)$,

$$I(t) = \frac{\rho\phi(t)\text{Exp}(-\int_0^t \rho\phi(y)dy)}{1 - (1 - \text{Exp}(-\int_0^t \rho\phi(y)dy))}$$

$$= \underline{\rho\phi(t)} .$$

At the end of section 2.3, conditions were given under which $\phi(t)$ could equal $u(t)$. This suggests that if it is desired to find aspects of the IRT distribution which match the reinforcement function, then the function $I(t)$ is the one to look at.

The relationship between $I(t)$ and $\phi(t)$ is deceptively simple, and suggests that the function $I(t)$ would be much easier to work with than $r(t)$. In practice however, serious problems arise in the estimation of $I(t)$.

Consider a sample of N IRT's, which for estimation purposes are grouped into intervals of one second width. Let n_k be the number of IRT's of duration between (k-1) and k seconds. Then,

$$\begin{array}{l} \text{IRT/Op.} \\ \text{for interval } (k-1,k) \end{array} = \hat{I}_k = \frac{n_k}{N - \sum_{i=1}^{k-1} n_i} .$$

The calculation of I_k in this manner is like calculating a binomial parameter on a sample size of N_k, where,

$$N_k = N - \sum_{i=1}^{k-1} n_i .$$

The variance of this estimate of I_k is given by,

$$\mathrm{var}(\hat{I}_k) = \frac{\hat{I}_k(1 - \hat{I}_k)}{N_k} .$$

(Kendall and Stuart 1952, Vol 2, p11.)

Now,

$$\frac{\hat{I}_k(1 - \hat{I}_k)}{N_k} = \frac{n_k(N - n_k)}{N(N - \sum_{i=1}^{k-1} n_i)} .$$

Thus as the value of k increases, the estimate of I_k becomes progressively more unreliable. (Its variance increases.) E.g. if $I(t)$ is in fact a constant, then the variance grows geometrically as k increases. In practice, if the time range used, (the longest IRT) is not long, compared with the interval width used to group the IRT's for estimation purposes, then the increase in variance of the estimates of I(t) is a serious problem. For this reason, the I(t) function will not be used in chapters 3-5 to describe the IRT distribution.

Weiss (1970) introduced what he termed the 'dwell function' to characterise the IRT distribution. The dwell function is the proportion of the total time occupied by IRT's of duration t. The function D(t) is thus, simply,

$$D(t) = tr(t).$$

The use of the dwell function is simply one of many possible ways of giving increased weight to long IRT's. It does not however seem to possess any especial properties which might make it of particular interest.

2.9 Conclusion

The model developed in this chapter seems to possess appropriate properties for a model of learning under schedules of interresponse time reinforcement. The following three chapters deal in detail with the predictions of the model for specific IRT schedules and with experimental tests of these predictions. The first schedule to be examined is the especially simple case of the Random Ratio schedule.

Chapter 3

THE RANDOM RATIO SCHEDULE

3.1 Previous Experimental Findings

Ratio schedules have not been very extensively studied for their IRT distribution properties. The general observation is that, under ratio schedules, IRT's are very short (Kintsch 1965, Ray and McGill 1964). The response rate, (reciprocal of the mean IRT) appears to be very insensitive to quite large changes in the probability of reinforcement (Sidley and Schoenfeld 1964). Williams, (1968) found little evidence of sequential effects among the IRT's, suggesting that IRT's were independent of one another. However his data was restricted, in that all the IRT's were very short.

The next section will show that all these results are predict -able from the model developed in chapter 2.

3.2 Predictions from the Model

The random ratio schedule is one where,

$$u(t) = p, \quad \forall t.$$

p is the probability that a response will be reinforced.

If this u(t) function is inserted in equation 2.6.a, it gives,

$$\phi(t) = \frac{\alpha^0 \theta \int_{t-a}^{t+a} (1-p) r(y) dy + \alpha^1 \int_{t-a}^{t+a} p r(y) dy}{\theta \int_{t-a}^{t+a} (1-p) r(y) dy + \int_{t-a}^{t+a} p r(y) dy}$$

$$= \frac{\alpha^0 \theta (1-p) + \alpha^1 p}{\theta (1-p) + p}.$$

This result is of particular importance, as it is independent of the parameter a. Indeed, it is independent of any choice for w(t;x), the spread function, and hence provides a valuable test for the model. This test is independent of any specific assumptions made about the stimulus distribution.

Let,

$$k = \frac{\alpha^0(1-p)\theta + \alpha^1 p}{(1-p)\theta + p}.$$

If θ is close to zero, i.e. non-reinforcement has little effect, then,

$$k \simeq \alpha^1.$$

Thus, for small θ values, $\phi(t)$ is insensitive to the value of p. Insensitivity to p values was noted by Sidley and Schoenfeld, 1964.

Further, if, as suggested in section 2.5,

$$\psi(t) = \rho,$$

then,

$$\underline{r(t) = \rho k \mathrm{Exp}(-\rho k t).} \tag{3.2.a}$$

The IRT distribution is thus exponential, and shows a preponder -ance of very short IRT's. (As reported by Kintsch (1965), and Ray and McGill (1964)).

Sequential statistics are fairly simple to derive, as the expression for the asymptotic IRT distribution (3.2.a) is so simple in form.

Let the notation,

$$r(t|1) = r(t|\text{previous response was reinforced})$$

$$r(t|0) = r(t|\text{previous response was not reinforced}),$$

be introduced. Let $\phi(t|1)$, $\phi(t|0)$, etc. have similar interpretations. Then,

$$\phi(t|1) = \phi(t) + (\alpha^1 - \phi(t))\theta_1\int_{t-a}^{t+a} r(y)dy$$

$$= \emptyset(t) + (\alpha^1 - \emptyset(t))\Theta_1 \int_{t-a}^{t+a} \rho k \mathrm{Exp}(-\rho k y)dy$$

$$= k + (\alpha^1 - k)\Theta_1 \Big[-\mathrm{Exp}(-\rho k y)\Big]_{t-a}^{t+a}$$

$$= k + (\alpha^1 - k)\Theta_1 (\mathrm{Exp}(-\rho k(t-a)) - \mathrm{Exp}(-\rho k(t+a)))$$

$$= k + (\alpha^1 - k)\Theta_1 \mathrm{Exp}(-\rho k t)(\mathrm{Exp}(\rho k a) - \mathrm{Exp}(-\rho k a))$$

$$= k + (\alpha^1 - k)\Theta_1 2\mathrm{sh}(\rho k a)\mathrm{Exp}(-\rho k t).$$

Put,

$$k_1 = 2(\alpha^1 - k)\Theta_1 \mathrm{sh}(\rho k a),$$

this makes the above equation take the form,

$$\emptyset(t|1) = k + k_1 \mathrm{Exp}(-\rho k t).$$

Now,

$$\int_0^t \rho(k + k_1 \mathrm{Exp}(-\rho k y))dy = \rho\Big[ky + k_1 \frac{\mathrm{Exp}(-\rho k y)}{\rho k}\Big]_0^t$$

$$= \rho\Big[kt - k_1 \frac{\mathrm{Exp}(-\rho k t)}{\rho k} + \frac{k_1}{\rho k}\Big]$$

$$= \rho k t + \frac{k_1}{k}(1 - \mathrm{Exp}(-\rho k t)).$$

Thus,

$$\underline{r(t|1) = (\rho k + \rho k_1 \mathrm{Exp}(-\rho k t))\mathrm{Exp}(-(\rho k t + (k_1/k)(1 - \mathrm{Exp}(-\rho k t))),}$$

and by analogy,

$$\underline{r(t|0) = (\rho k + \rho k_0 \mathrm{Exp}(-\rho k t))\mathrm{Exp}(-(\rho k t + (k_0/k)(1 - \mathrm{Exp}(-\rho k t))),}$$

where, $k_0 = 2(\alpha^0 - k)\Theta_0 \mathrm{sh}(\rho k a)$.

If the two parameters, k_1 and k_0, are inspected, the conditions under which sequential effects are small can be found.

$$k_1 = 2(\alpha^1 - k)\Theta_1 sh(\rho ka).$$

$$k_0 = 2(\alpha^0 - k)\Theta_0 sh(\rho ka).$$

There are no sequential effects, if,

i) $a \simeq 0.0$. I.e the spread effect is zero. This seems unlikely in view of the psychological interpretation of the spread function.

ii) $\Theta_0 \simeq 0.0$. If Θ is close to zero, non-reinforcement has very little effect. When Θ_0 is approximately zero, then k_0 is likely to be close to zero also. However, $\Theta \simeq 0.0$ also implies that, $k \simeq \alpha^1$ and hence that $k_1 \simeq 0.0$. Thus sequential effects after reinforcement are small, when the effects of non-reinforcement are small.

Since many unreinforced responses are emitted on ratio schedules, with apparently little effect on the rate of responding, the assumption that Θ_0 is close to zero is not unreasonable. This single assumption thus leads directly to the predictions that rates of responding are insensitive to reinforcement probability (Sidley and Schoenfeld, 1964) and that sequential effects are likely to be small (Williams, 1968).

It is interesting to note that the parameters k_1 and k_0 are very simply related, for,

$$\frac{k_1}{k_0} = \frac{2(\alpha^1 - k)\Theta_1 sh(\rho ka)}{2(\alpha^0 - k)\Theta_0 sh(\rho ka)}$$

$$= \frac{(\alpha^1\Theta(1-p) + \alpha^1 p - \alpha^0\Theta(1-p) - \alpha^1 p)}{(\alpha^0\Theta(1-p) + \alpha^0 p - \alpha^0\Theta(1-p) - \alpha^1 p)\Theta}$$

$$= \frac{\Theta(1-p)(\alpha^1 - \alpha^0)}{p(\alpha^0 - \alpha^1)\Theta}$$

$$= \frac{-(1-p)}{p}. \qquad (2.3.b)$$

I.e. the ratio of k_1 to k_0 is equal to minus the inverse of the ratio of the probability of reinforcement to non-reinforcement.

Equation (3.2.b) is a simple version of a more complex relation between the results of reinforcement and non-reinforcement.

Define,

$$v(t) = \frac{\int_{t-a}^{t+a} (1-u(y))r(y)dy}{\int_{t-a}^{t+a} u(y)r(y)dy} .$$

Then, for any u(t),

$$\phi(t) = \frac{\alpha^0 \theta v(t) + \alpha^1}{\theta v(t) + 1} .$$

Generalising the constants k_1 and k_0 to functions $k_1(t)$ and $k_0(t)$, defined by,

$$k_1(t) = (\alpha^1 - \phi(t))\theta_1 \int_{t-a}^{t+a} r(y)dy,$$

($k_1(t)$ thus describes the effect of a reinforcement on $\phi(t)$) and,

$$k_2(t) = (\alpha^0 - \phi(t))\theta_0 \int_{t-a}^{t+a} r(y)dy.$$

Then,

$$\frac{k_1(t)}{k_0(t)} = \frac{(\alpha^1 - \phi(t))}{(\alpha^0 - \phi(t))\theta}$$

$$= \frac{\alpha^1 \theta v(t) + \alpha^1 - \alpha^0 \theta v(t) - \alpha^1}{(\alpha^0 \theta v(t) + \alpha^0 - \alpha^0 \theta v(t) - \alpha^1)\theta}$$

$$= \frac{v(t)(\alpha^1 - \alpha^0)}{(\alpha^0 - \alpha^1)}$$

$$= -v(t).$$

Thus,

$$\frac{k_1(t)}{k_0(t)} = \frac{-\int_{t-a}^{t+a} (1-u(y))r(y)dy}{\int_{t-a}^{t+a} u(y)r(y)dy}$$

This relation states that for any value of t, the ratio of the effects of reinforcement and non-reinforcement on the value of $\phi(t)$ at that point is given by minus the inverse of the ratio of the probabilities of reinforcement and non-reinforcement influencing the value of $\phi(t)$ at that point t. The relative effects of reinforcement and non-reinforcement are thus closely tied to the actual probabilities of occurrence of reinforcement and non-reinforcement.

The next section deals with an experiment set up to obtain IRT distributions from an actual random ratio schedule. Data from this experiment is then used to obtain values for parameters such as k, k_1, and k_0. The values found are given in section 3.4. The parameters, whose estimates are denoted by putting a hat ($\wedge$) over the appropriate symbol, are usually estimated by computing least square fits to certain sets of data. Unless specifically stated otherwise it is to be assumed that a parameter is estimated by a least squares method. The estimations are usually done by search procedures. It is possible in simple cases to derive exactly expressions for the estimates of some parameters. (E.g. the maximum likelihood estimate for ρk is given by the reciprocal of the mean IRT). However, to maintain uniformity of method across all reinforcement schedules, least square fits were used as the basic method throughout.

3.3 The Experiment

In order to test the predictions of the model more closely, some experiments with Random Ratio reinforcement were undertaken. Two values of p were used, these being 0.5, and 0.1. Four subjects were run under each schedule.

The subjects were all students at the University of Stirling. In most experiments on Operant behaviour, it has been customary to use animal subjects, chiefly rats and pigeons. When human subjects have been used, the commonest form of response chosen as the operant response has been the observing response. E.g. subjects have to monitor a meter, to check its deflection (Laties and Weiss, 1960, 1963). Pressing a button illuminates the meter, and if the needle is seen to be deflected on illumination, this is regarded as a reinforcement. The most extensive investigations of human operant

Figure 3.3.1 -- The subject's response box. (Actual size).

behaviour have been undertaken by Weiner (1962, 1964a, 1969). Most of Weiner's experiments have been on fixed interval schedules, (1964b). However, one experiment (1964c) did consider a fixed ratio schedule. The performances obtained by Weiner, who also used an observing response, are closely similar to those found in work with animal subjects, suggesting that the use of human subjects is not unreason -able, or likely to produce any highly unexpected results.

Apart from the possible intrinsic interest of using human subjects, in contrast to animal ones, the choice was influenced by practical considerations, Reinforcement can be made relatively instantaneous for human subjects, (flashing lights were used). Animals, in contrast, are usually food reinforced, and the consumption of the food takes a relatively long time. Since it is times that are to be recorded, this 'eating time' would confound the results and is better eliminated if possible.

From the subject's point of view, the experimental situation closely resembled that of the ubiquitous Skinner Box, used in animal experimentation. Subjects were seated at a table in a small sound-protected room, approximately 6 ft. by 8 ft. by 7 ft. The table was equipped with a small box, on which were mounted a push-button, a red pilot light, a green pilot light and a digital counter. (See figure 3.2.1). The following instructions were given to the subject.

"You are asked to try and score points by pressing and releasing this button in some fashion. The counter will record your score. If you make a correct response the green light will come on to inform you of this, and the counter will add one. If you make an incorrect response then the red light will come on. You cannot score points by holding the button down. When you have pressed the button you must release it and not make the next response until the light has gone out.

Please pay close attention to the lights and the counter,so that you can score as many points as possible.

You will be allowed five minutes to settle in to the experiment. There will then be a short pause. The rest of the session will follow without a break.

If you have any questions, please ask them now."

Questions were answered in as vague a manner possible, to avoid giving the subject any idea of what was expected. Essentially, the only constraints placed on the subjects were injunctions to score points somehow, and not to respond while any of the display lights were lit. These lights were on for 100 msec. This instruction was inserted because IRT's shorter than this occasionally caused the recording apparatus to jam. This does not seem to have imposed any real constraint on the subjects' behaviour, as it was substantially lower than most of the IRT's emitted by any of the subjects.

Subjects came for 10 sessions in all. They came daily, at any convenient time of the day until they had completed all their sessions. (Subjects did not however come at weekends). They were paid at the rate of 30p per session, but had to complete all the sessions to receive payment. Each session ran for 45 minutes, excluding the five minute warm-up session.

At the start of each session, the subject was told his score from the previous session. This was done to try and maintain the subject's interest, by giving him a comparison against which to work. At the end of each session, subjects were also asked to say what they thought determined whether or not they obtained a reinforcement.

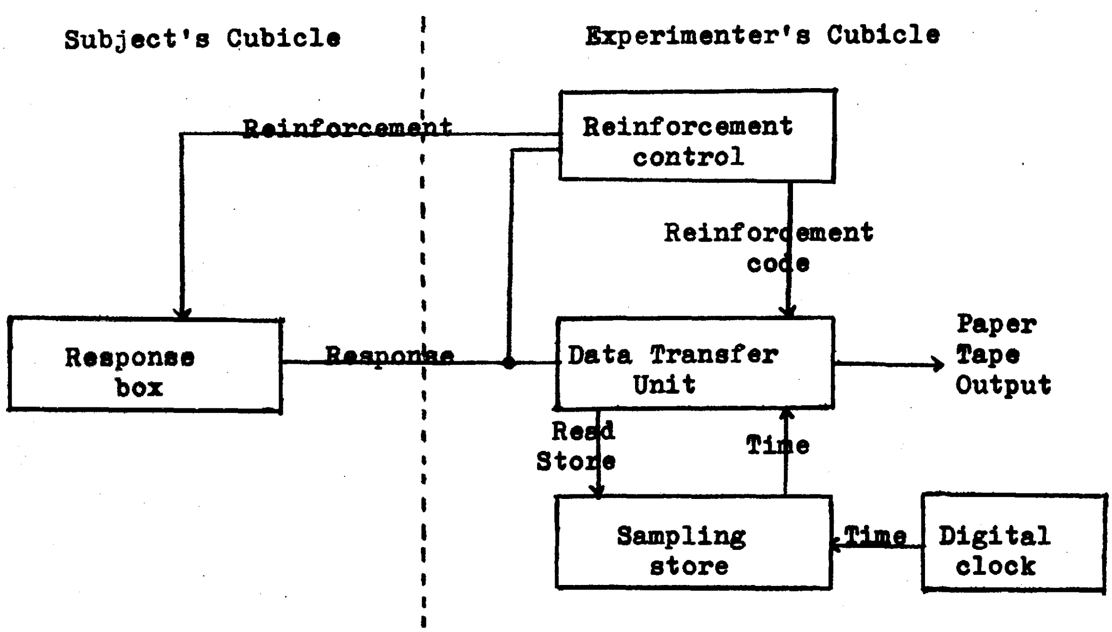


Figure 3.3.2 -- Block diagram of the layout of the Experimental apparatus.

The control and recording equipment were placed in a cubicle adjacent to the experimental cubicle. Figure 3.3.2 gives a block diagram layout of the system. A digital clock reads continuously into the sampling store. The subjects response triggers (a) the reinforcement, which is shown to him as a coloured light and which is also output to the data transfer unit, and (b) causes the data transfer unit to read the time off from the sampling store. The time, and a code for reinforcement/non-reinforcement are output onto paper tape by the data transfer unit. The use of the sampling store as a buffer between the data transfer unit and the digital clock prevents the 'read' pulse from the data transfer unit from resetting the digital clock and has the following advantage. No allowance needs to be made for the reset time of the clock and the record time of the data transfer unit when calculating the IRT's. The time readings are cumulative and IRT's are given by the difference between two consecutive readings. The clock ran in milliseconds. An example is:-

0 - 001352 - 200
0 - 008473 - 100
0 - 013289 - 100
0 - 017984 - 200
0 - 022839 - 100

Leading zero. This occurs every third number. It is used to check that the data is aligned. If any set of three numbers does not begin with zero, the analysis programme reads on until this is so. This misalignment is caused by occasional punching errors from the data transfer unit. The number of such errors was in fact small.

Code for reinforcement (100), or non-reinforcement (200).

Times in milliseconds. Zero time is the start of the experimental session. Differences between successive times give IRT in milliseconds.

Figure 3.3.3 -- An example of the data output format.

3.4 The Results

Figure 3.4.1 gives the session by session IRT distribution for subjects 1-4. (p = 0.5). The term session 0 is used to denote the warm-up period of the very first session. In this experiment every other response was reinforced, on average. Due to a fault in the data transfer unit, this experiment had to be abandoned after the sixth session, and data from that session scrapped. The IRT's for each session were classified into intervals of one second width (0-1, 1-2, etc.) and the description Pr(IRT) is used as an abbreviation for the probability of an IRT lying in one of these intervals.

All the subjects 1-4 show the same kind of session by session changes in the IRT distribution. Initially the mean IRT is in the region of 1.5-2.0 seconds, but rapidly shortens to around 0.5 seconds, and all subjects have over 95% of their IRT's less than one second long by the fifth session. (The rate of responding is around 100 per minute, which is very high.)

In view of the high rate of responding, it was decided to use the fifth session data as asymptotic data for parameter estimation. Figure 3.4.2 shows the data from the fifth session, for subjects 1-4, compared with the predicted values. The predicted values are calculated from the least squares fit of the model to the data. Table 3.4.(i) gives the value of the exponential parameter ρk and the least squares error (LSE). These figures confirm the visual impression obtained from figure 3.4.2. The fits are very close.

Subject	$\hat{\rho k}$	LSE.10^6	No. Data Points
1	4.99	6	4
2	3.33	601	4
3	2.98	1670	4
4	4.21	172	4

Table 3.4.(i) -- Estimates of the exponential parameter ρk and a list of the least square errors (LSE) between the obtained and predicted values of the asymptotic IRT distribution for subjects 1-4. The last column gives the number of data points used in the parameter estimation.

Figure 3.4.1 -- Random ratio schedule. $p = 0.5$. The session by session IRT distributions obtained from subjects 1-4. The IRT's were grouped into intervals of one second width.

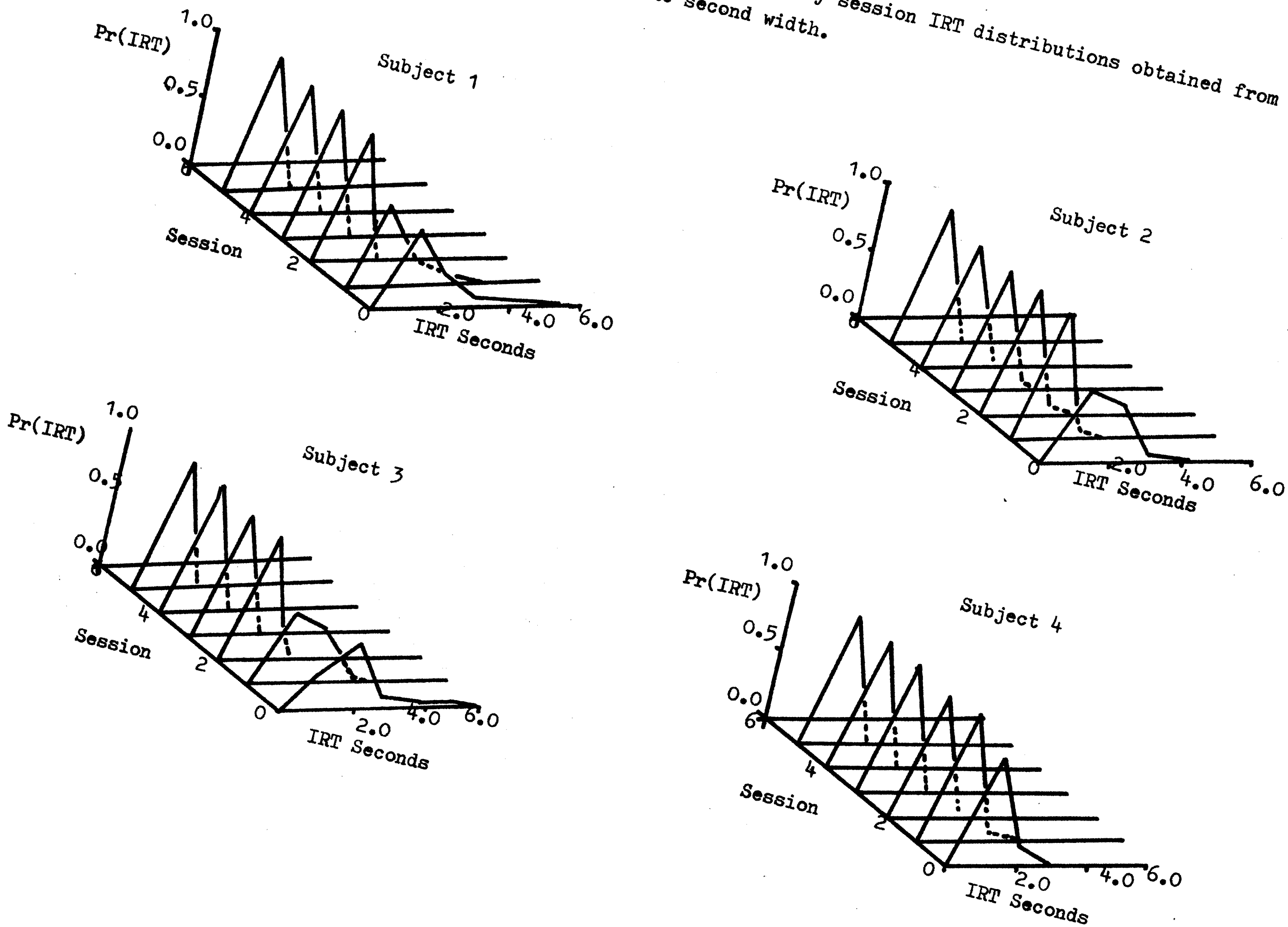


Figure 3.4.2 -- Random ratio schedule. p = 0.5. A comparison of the data and predictions for the asymptotic IRT distribution for subjects 1-4. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one second width.

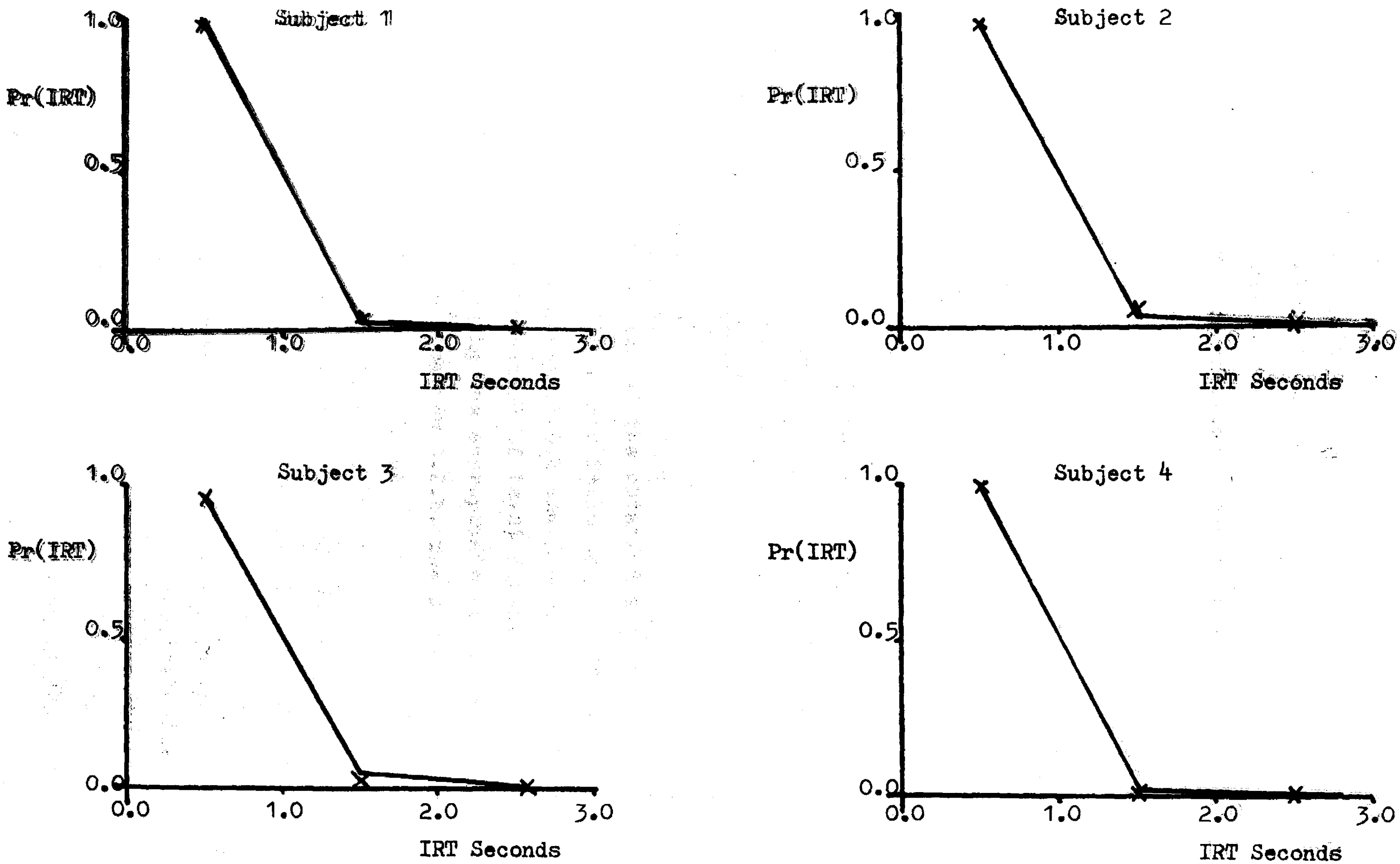


Figure 3.4.3 gives data for the asymptotic distribution of IRT's, given that the preceeding response was reinforced. The least square fit is also shown, and again is very close. Table 3.4.(ii) gives the parameter estimates (k_1/k) and the least squares error. Generally, the fits to this sequential statistic are better than that to the asymptotic IRT distribution. The values of (k_1/k) are all positive and apparently (see section 3.5) illustrate the general speeding up effect that reinforcement has on responding.

Subject	$\widehat{k_1/k}$	$LSE.10^6$	No. Data Points
1	0.14	3	4
2	1.98	1	4
3	1.33	801	4
4	0.29	37	4

Table 3.4.(ii) -- Estimates of k_1/k for subjects 1-4. The LSE column gives the least square error between the obtained and predicted values of the IRT distribution conditional on the previous response being reinforced. The last column gives the number of data points used in the parameter estimation.

Figure 3.4.4 gives a comparison of the obtained and predicted results for the distribution of IRT's that followed a non-reinforced response. Unlike the previous figures (3.4.2 and 3.4.3), the predicted values were not calculated by the method of least squares. The estimate of (k_0/k) was made by using the appropriate estimate of (k_1/k) and the relation given in equation 3.2.b. The parameter (k_1/k) seems to be very sensitive to quite small changes in the IRT distribution. E.g. compare subjects 1 and 2, in figure 3.4.3, where the results look very similar, to the parameter values in table 3.4.(ii)). This sensitivity seems to be reflected in the poor fits obtained, for subjects 2 and 3, to the IRT distribution given the previous response was not reinforced, which is illustrated in figure 3.4.4. Table 3.4.(iii) gives the parameter values calculated from equation 3.2.b, together with the square error between the data and the predicted values .

Better fits can actually be obtained by estimating (k_0/k) from the data. However it seems more interesting to use equation 3.2.b.

Figure 3.4.3 -- Random ratio schedule. p = 0.5. A comparison of the data and predictions for the asymptotic IRT distribution conditional on the previous response being reinforced for subjects 1-4. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one second width.

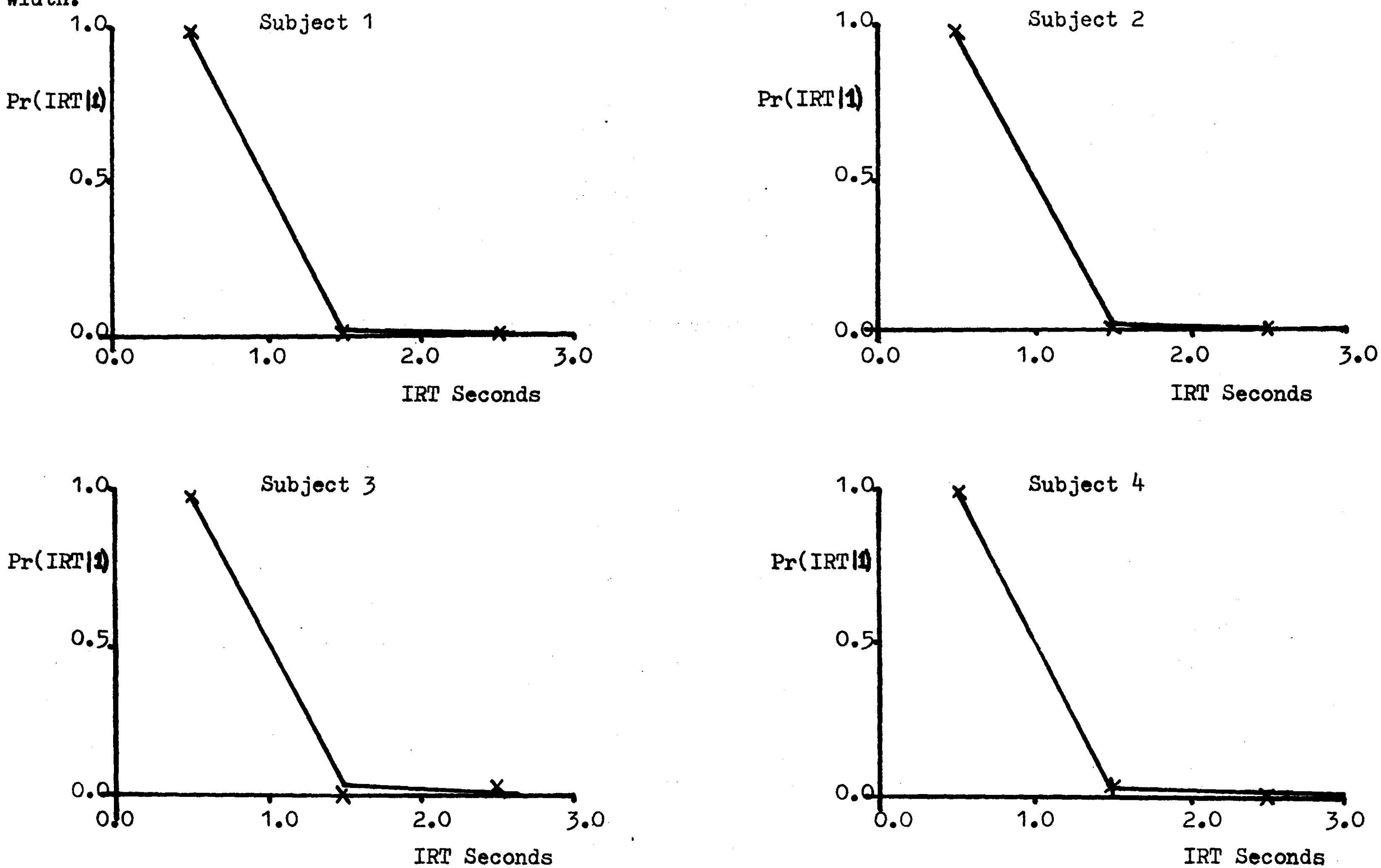
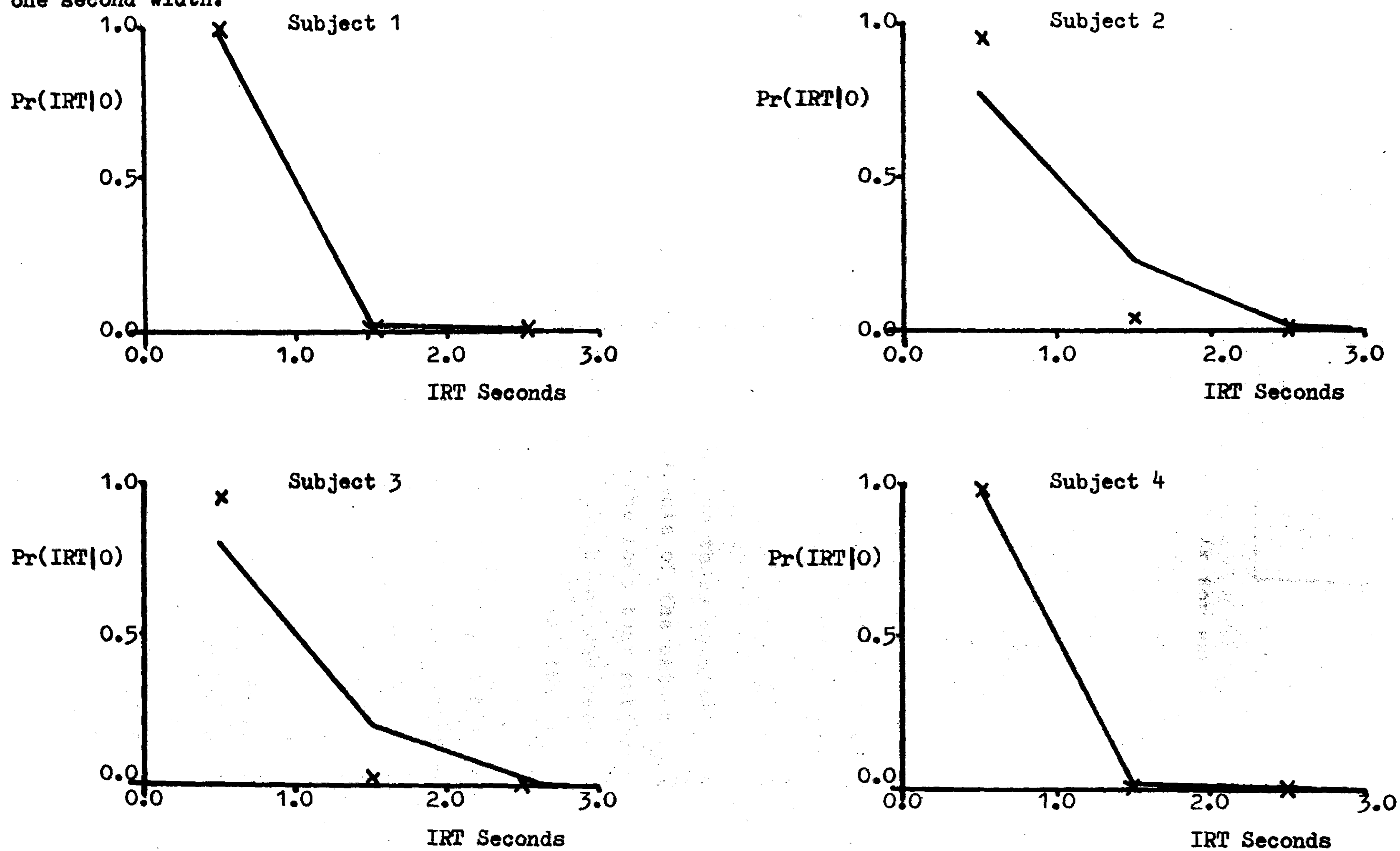


Figure 3.4.4 -- Random ratio schedule. p = 0.5. A comparison of data and predictions for the asymptotic IRT Distribution conditional on the previous response not being reinforced for subjects 1-4. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one second width.



Subject	$\widehat{k_0/k}$	SE.10^6	No. Data Points
1	-0.14	14	4
2	-1.98	81260	6
3	-1.33	47223	5
4	-0.29	546	4

Table 3.4.(iii) -- Estimates of the parameter k_0/k for subjects 1-4. SE column gives the square error between the obtained and predicted values of the IRT distribution conditional on the previous response not being reinforced. The last column gives the number of points used in the calculation of SE. The values of k_0/k given here are not estimated from the conditional IRT distribution, but from $\widehat{k_1/k}$.

The relatively poor fit obtained overall, suggests that sequential effects are more complex than outlined in the model, though it is not clear to what extent a different choice of the spread function w(t;x) influences these conditional results.

The results for subjects 5-8 are similar to those obtained for subjects 1-4. In fact the fits to various aspects of the data are, on the whole, rather better. This may be due to the fact that subjects 5-8 received the full 10 sessions of training and hence the results from session 10 are much closer to the asymptotic values, than those obtained from subjects 1-4.

Figure 3.4.5 gives the session by session results for subjects 5-8, where $p = 0.1$. I.e. on average every 10th response was reinforced. As for subjects 1-4 (figure 3.4.1), these results show a gradual shift from a fairly low rate of responding, to a very high rate of responding, by the 10th session. As might be expected with the less frequent reinforcement, the trend towards shorter IRT's appears to be much slower for these subjects, than for subjects 1-4. (Subject 7 seems to be an exception. For this subject the rate of responding was very high from the beginning. Given the simple nature of the task, it might have been expected that a decrement in performance would appear, across sessions, due to boredom. However a very high rate was maintained throughout.) Subject 6 shows clearly the pattern of session by session changes in the IRT distribution. The subject begins by first eliminating the very long IRT's, giving a distribution

Figure 3.4.5 -- Random ratio schedule. p = 0.1. The session by session IRT distributions obtained from subjects 5-8. The IRT's were grouped into intervals of one second width.

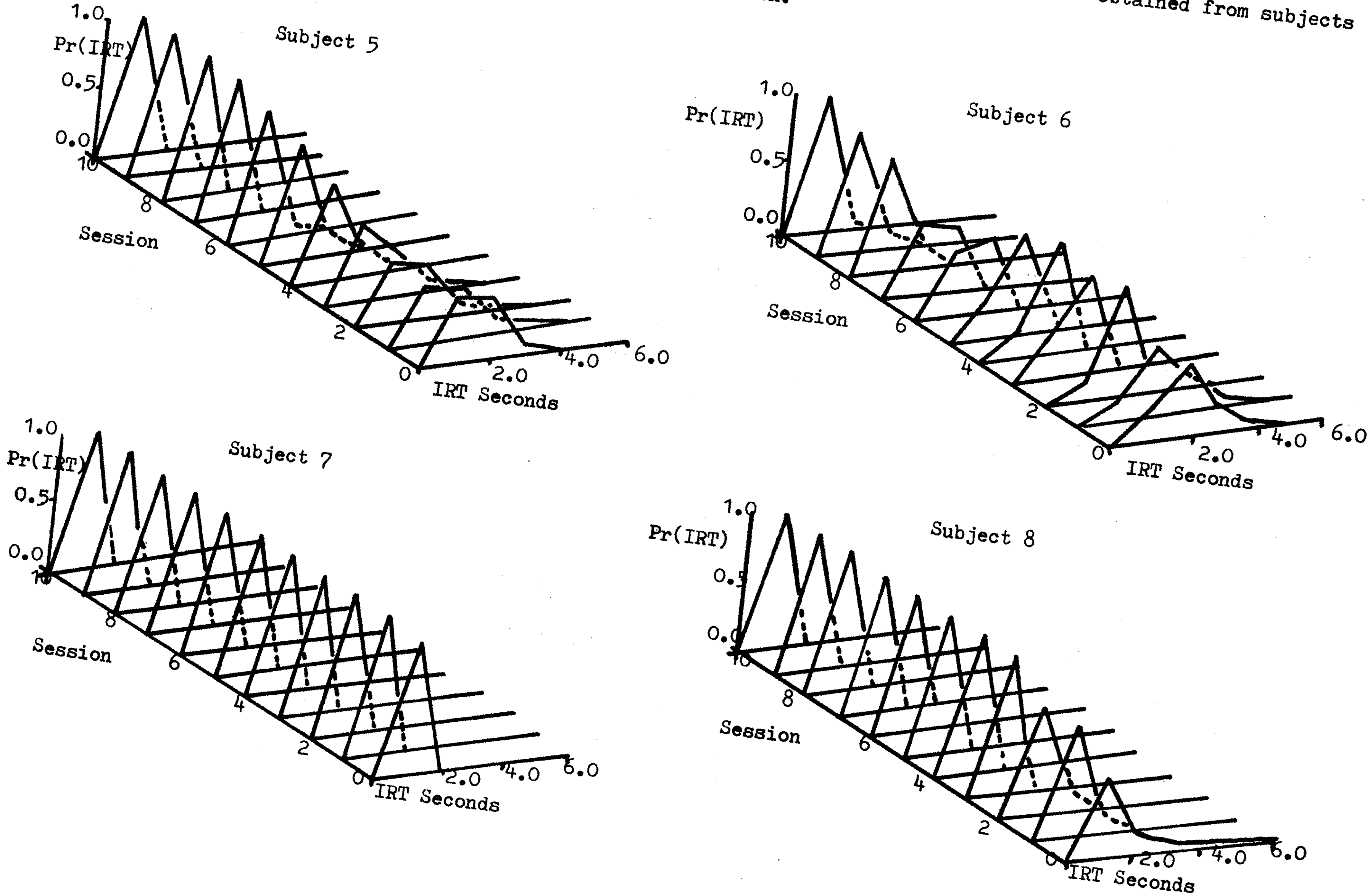
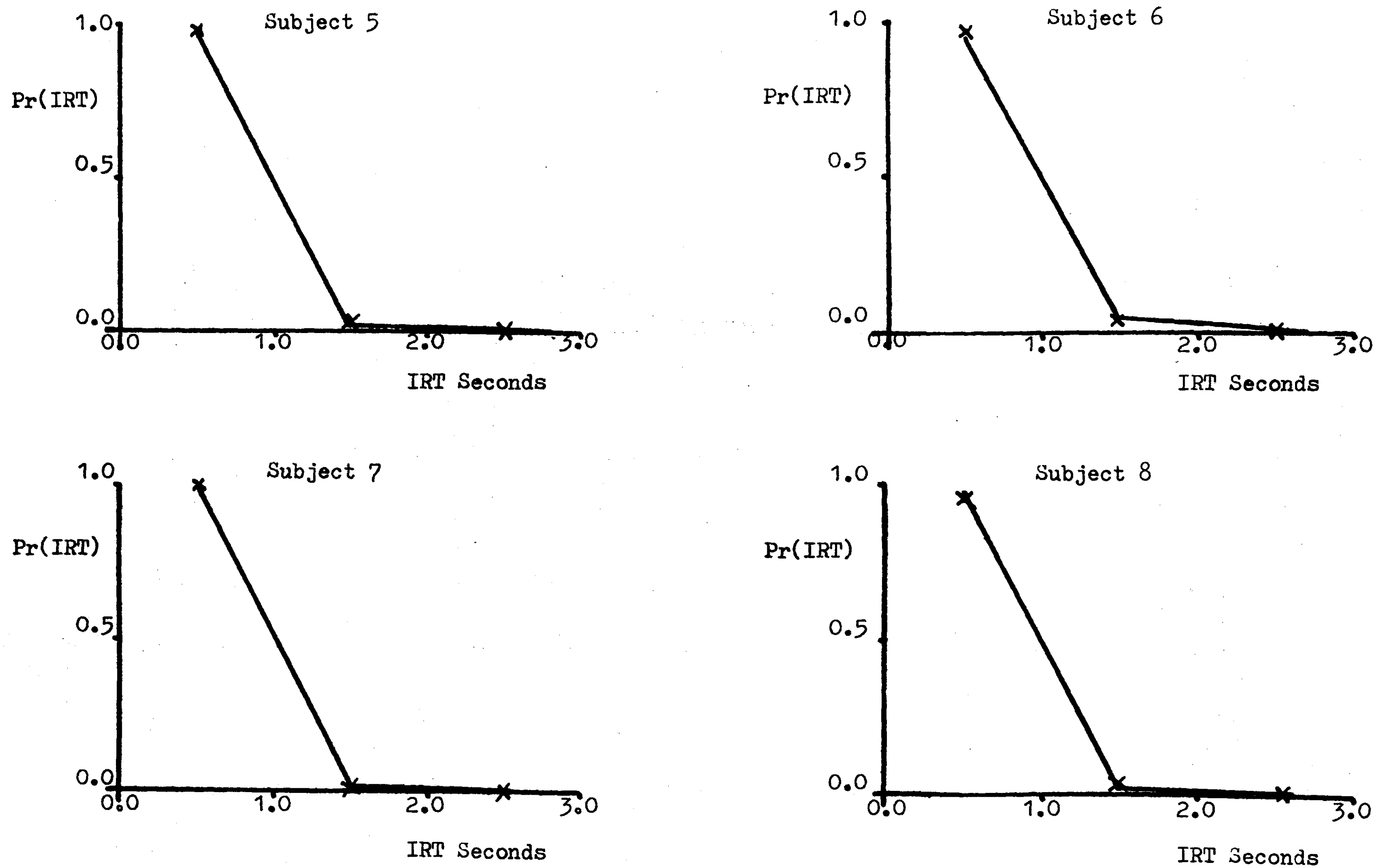


Figure 3.4.6 -- Random ratio schedule. p = 0.1. A comparison of the data and predictions for the asymptotic IRT distribution for subjects 5-8. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one second width.



with a sharp cutoff around 2 seconds. There is then a general trend towards shorter IRT's, illustrated by the increasing number that fall in the interval 0.0-1.0 seconds, and the decreasing number in the interval 1.0-2.0 seconds. Sessions 6 and 7, for subject 6 show clearly the changeover of the mode of the IRT distribution from the 1.0-2.0 second interval, to the 0.0-1.0 interval. This is followed by a further shortening of the IRT's, when the subjects push their rates into the regions of the highest that it is physically possible to make.

Figure 3.4.6 shows the data from the 10th session for subjects 5-8, compared with the least squares fit of the model to this data. As for subjects 1-4, the fits are very good. Table 3.4.(iv) gives the estimates of ρk for this data, as well as the least square error. The values of $\hat{\rho k}$ for subjects 5-8 are of the same order as those for subjects 1-4, the chief difference between the two groups being the high value of subject 7, compared with the rather low value for subject 3.

Subject	$\hat{\rho k}$	$LSE.10^6$	No. Data Points
5	4.96	1	4
6	3.12	8	4
7	5.30	0	4
8	4.35	21	4

Table 3.4.(iv) -- Estimates of the exponential parameter ρk and a list of the least square errors (LSE) between the obtained and predicted values of the asymptotic IRT distribution for subjects 5-8. The last column gives the number of data points used in the parameter estimation.

Figure 3.4.7 gives the IRT distribution conditional on the previous response being reinforced. These fits to sequential statistics are again very good, with the rather striking exception of subject 6. This subject shows a strong trend in the opposite direction to that predicted by the model. After a reinforcement his IRT's are in general longer, rather than shorter. Whether or not this effect would disappear with further training is difficult to say. This subject did in fact comment that when he obtained a

Figure 3.4.7 -- Random ratio schedule. p = 0.1. A comparison of the data and predictions for the asymptotic IRT distribution conditional on the previous response being reinforced for subjects 5-8. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one second width.

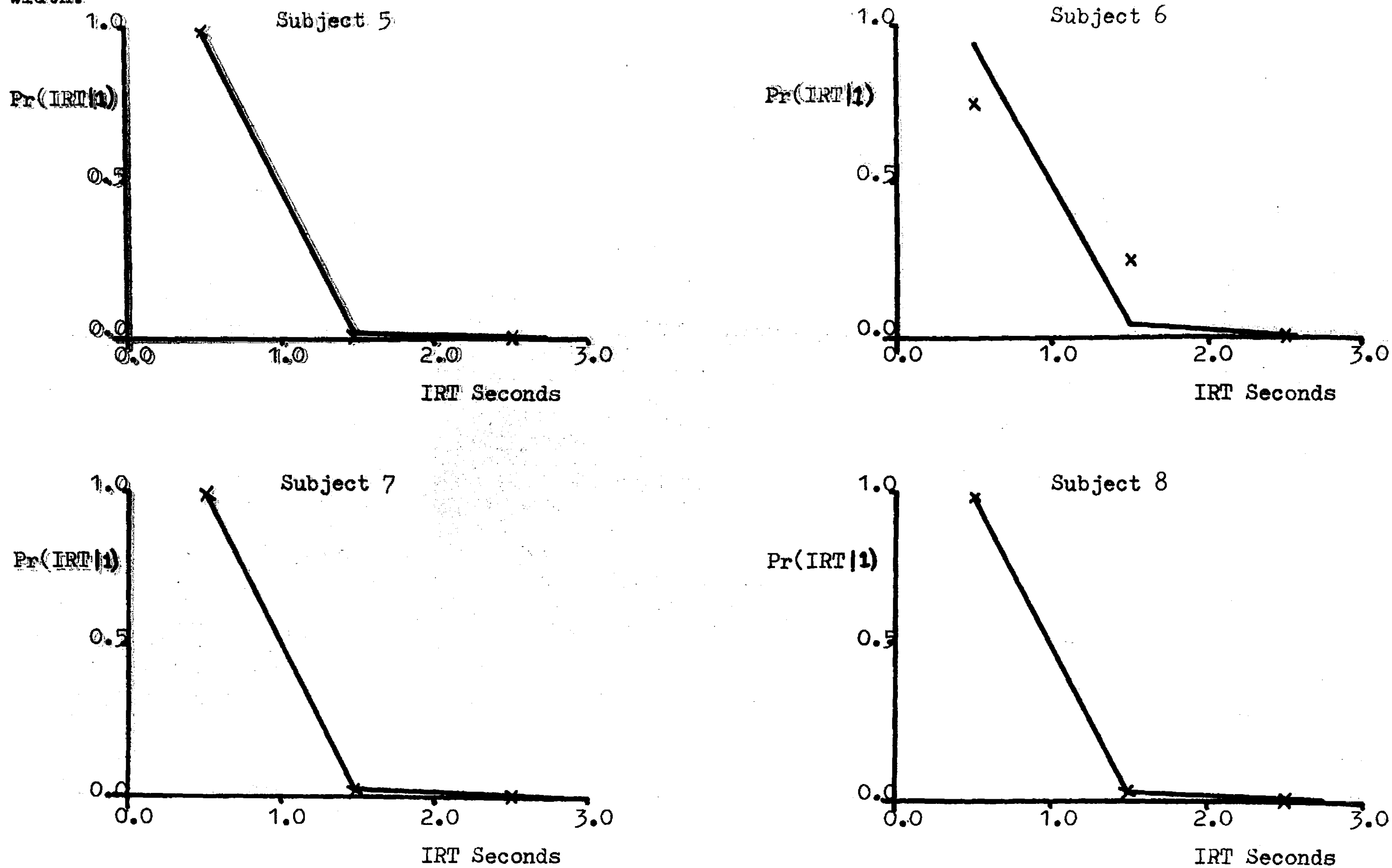
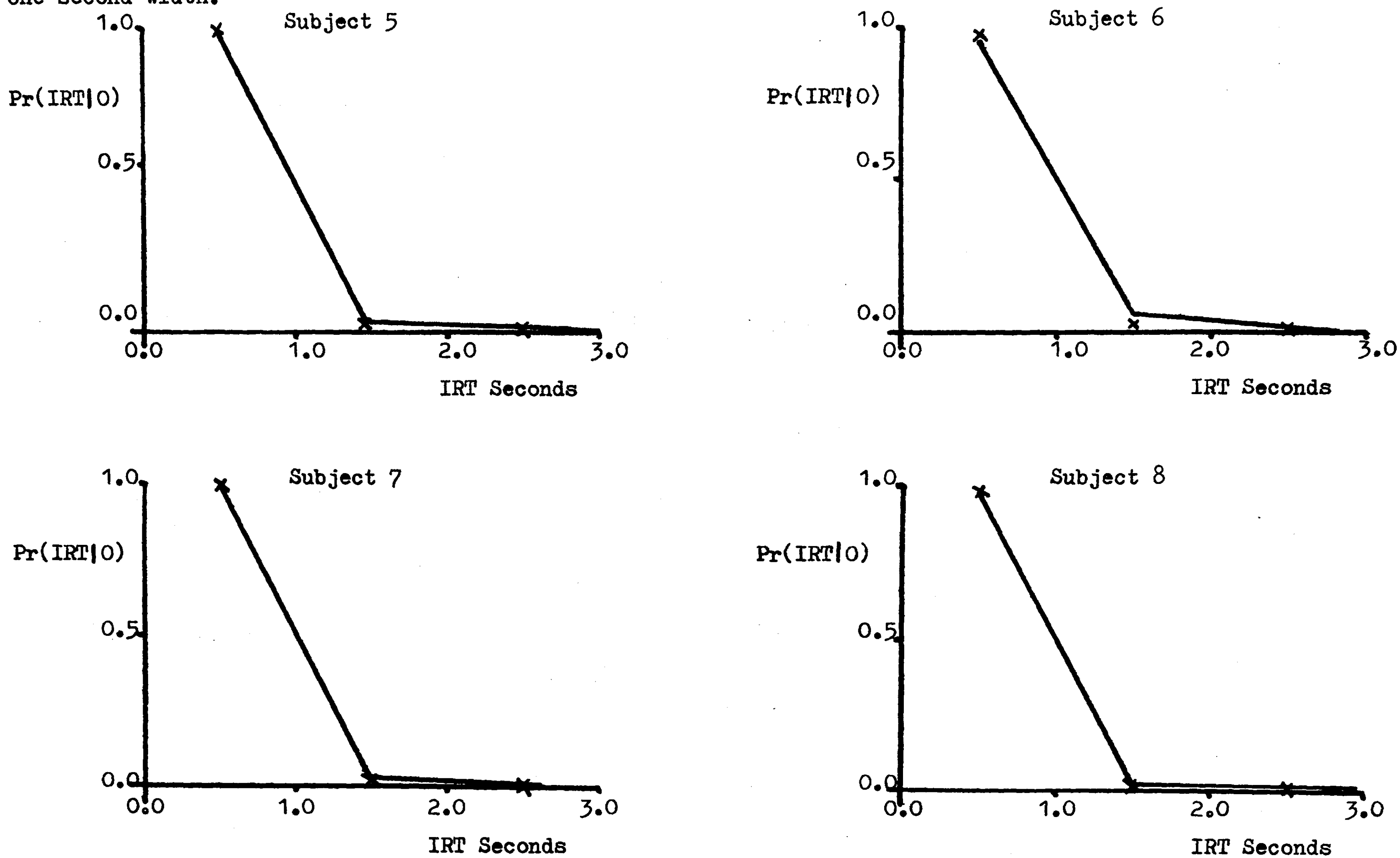


Figure 3.4.8 -- Random ratio schedule. p = 0.1. A comparison of data and predictions for the asymptotic IRT distribution conditional on the previous response not being reinforced for subjects 5-8. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one second width.



Subject	$\widehat{k_1/k}$	LSE.10^6	No. Data Points
5	0.85	18	4
6	0.00	86564	4
7	0.00	18	4
8	0.24	2	4

Table 3.4.(v) -- Estimates of k_1/k for subjects 5-8. The LSE column gives the least square error between the obtained and predicted values of the IRT distribution conditional on the previous response being reinforced. The last column gives the number of data points used in the parameter estimation.

reinforcement that he was pleased, and "stopped for a second" to wonder why he had received one.

Table 3.4.(v) gives the estimated values of the parameter (k_1/k), and the least square errors. If the anomalous subject 6 is ignored, then the parameter values are very small on average, suggesting that the response strength function $\emptyset(t)$ was very close to its maximum value α^1.

Figure 3.4.8 gives a comparison of the data and predictions for the asymptotic distribution of IRT's that followed a non-reinforcement. As for subjects 1-4, these predicted values are based not on the least

Subject	$\widehat{k_0/k}$	SE.10^6	No. Data Points
5	-0.09	1	4
6	-0.00	1054	4
7	-0.00	2	4
8	-0.03	21	4

Table 3.4.(vi) -- Estimates of the parameter k_0/k for subjects 5-8. SE column gives the square error between the obtained and predicted values of the IRT distribution conditional on the previous response not being reinforced. The last column gives the number of data points used in the calculation of SE. The values of k_0/k given here are not estimated from the conditional IRT distribution, but from k_1/k.

square fits, but are based on the parameter values of (k_0/k) calculated from equation 3.2.b.

The fits this time are all quite good, in contrast to the results for subjects 1-4. As might be expected, the fit for subject 6 is the poorest. Generally all the fits for subjects 5-8 are better than those obtained for subjects 1-4, perhaps reflecting the extra training subjects 5-8 had. The good fits to the sequential statistics, and the small values of the parameters k_1/k and k_0/k, taken together seem to suggest that the sequential effects are on the whole small, so that $\emptyset(t)$ must lie close to its maximum value of α^1.

It is possible that the very good fits obtained so far are partially due to the small number of data points available for parameter estimation. (Usually four points were used, and a single parameter estimated.) The small number of data points result from the very high rate of responding produced by all the subjects. It may thus be of interest to investigate the data when it is class -ified on the basis of intervals smaller than one second. The data was thus reclassified into half second intervals, (0.0-0.5, 0.5-1.0, etc.) and subject to a re-analysis.

Subject	$\widehat{\rho k}$	LSE.10^6	No. Data Points
1	1.54	278172	7
2	1.52	523164	7
3	1.35	492529	7
4	1.39	765236	7

Table 3.4.(vii) -- Estimates of the parameter ρk and a list of the least square errors (LSE) between the obtained and predicted values of the asymptotic IRT distribution for subjects 1-4 when the data is classified into half second intervals. The last column gives the number of data points used in the parameter estimation.

Figure 3.4.9 gives the data from subjects 1-4 when it is class -ified into half second intervals. (Asymptotic data only). These results are distinctly different from those given in figure 3.4.2. They show a peak, not in the first interval, but in the second

Figure 3.4.9 -- Random ratio schedule. p = 0.5. A comparison of the data and predictions for the asymptotic IRT distribution for subjects 1-4. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one half second width.

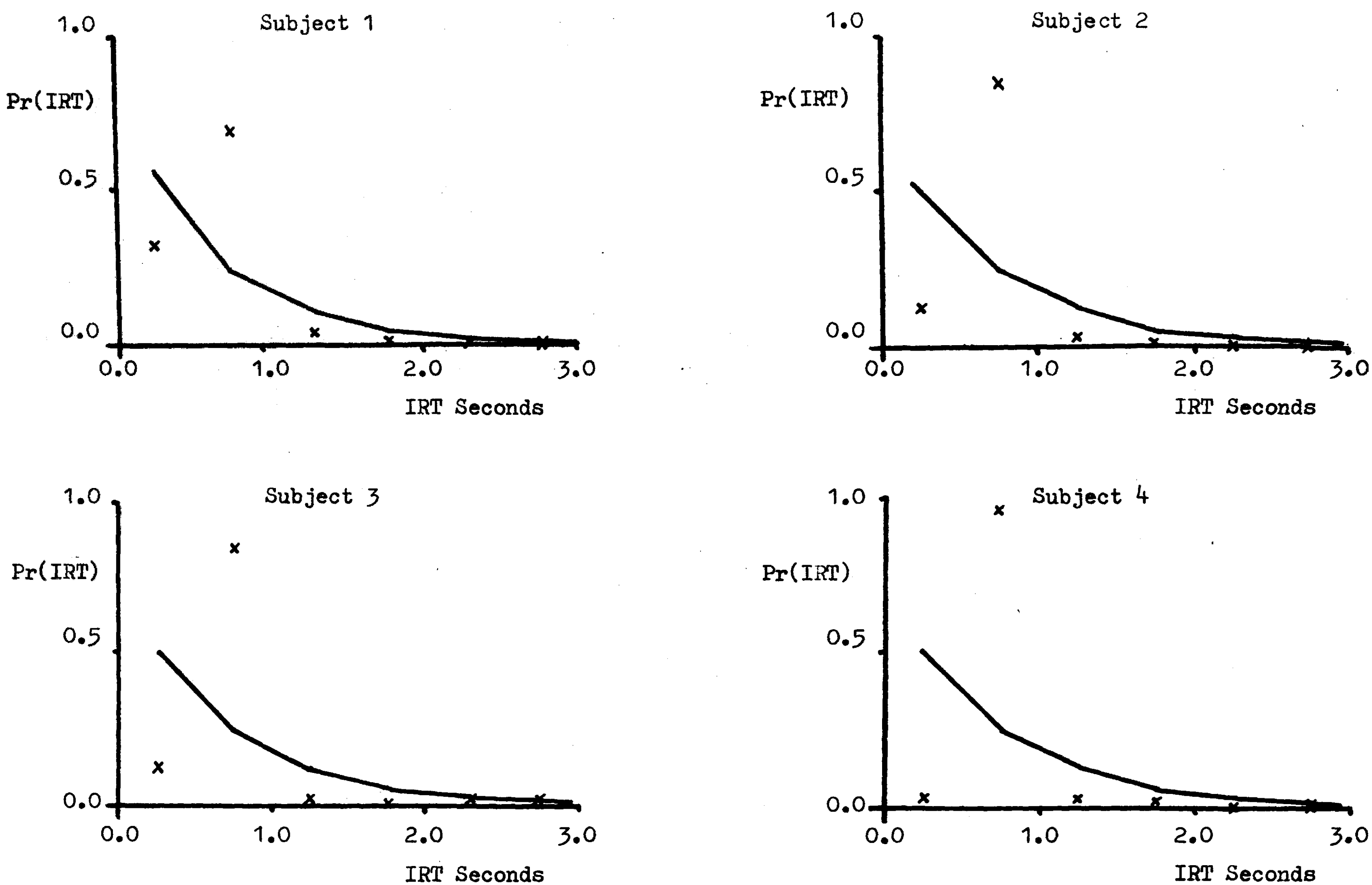
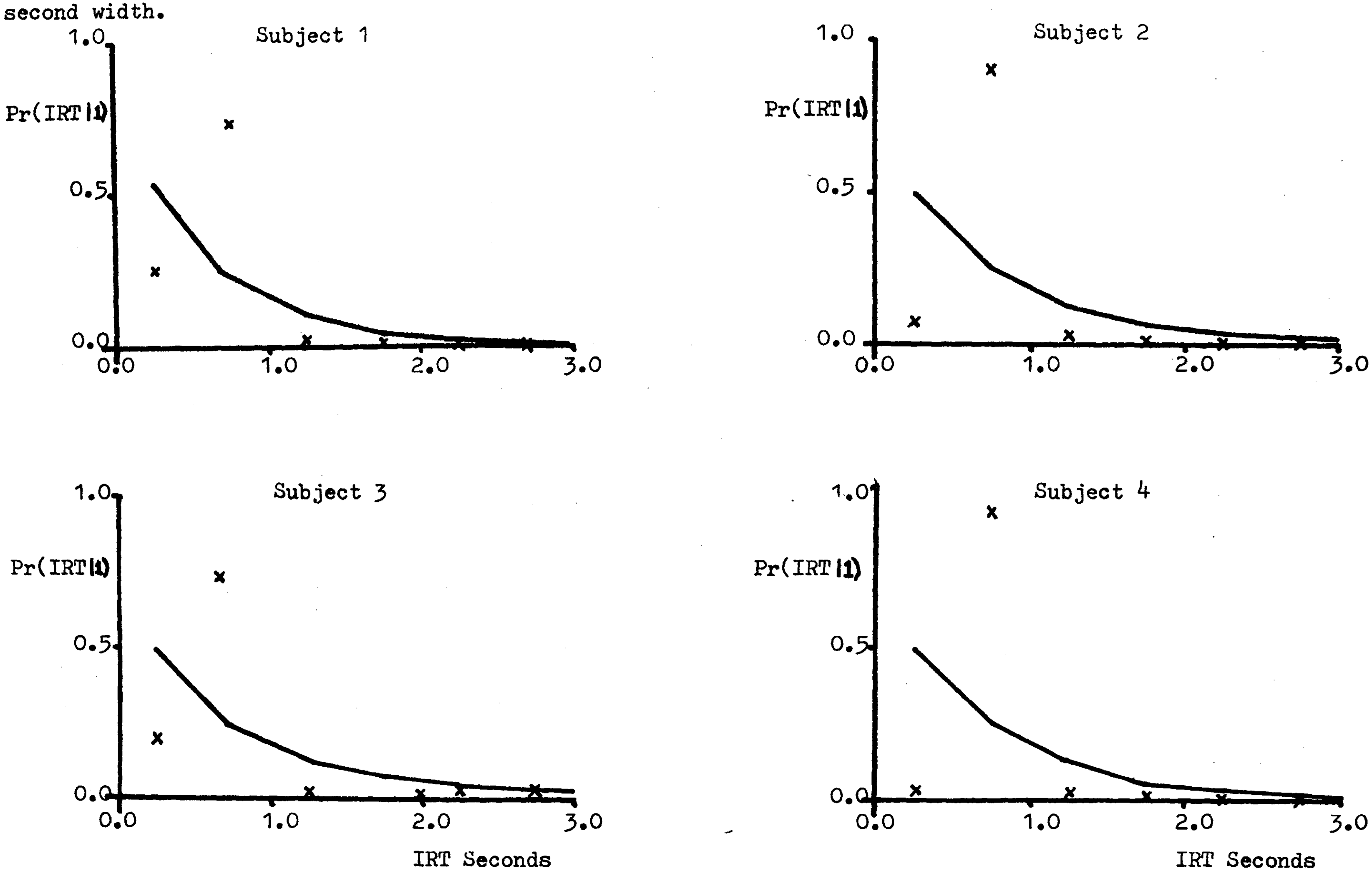


Figure 3.4.10 -- Random ratio schedule. p = 0.5. A comparison of the data and predictions for the asymptotic IRT distribution conditional on the previous response being reinforced for subjects 1-4. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one half second width.



interval. Table 3.4.(vii) shows the parameter values and the least squares fit of the model to this data. The parameter values ρk are characterised by their uniformity and the atrocity of the fit to the data.

Not surprisingly, attempts to fit the conditional statistics (figure3.4.10) are equally poor. Table 3.4.(viii) gives estimates of k_1/k. These are all zero. showing the insensitivity of the model to variations in the data points, when the fit is intrinsically so poor.

Subject	$\widehat{k_1/k}$	LSE.10^6	No. Data Points
1	0.00	344540	7
2	0.00	650323	7
3	0.00	379789	7
4	0.00	748492	7

Table 3.4.(iii) -- Estimates of k_1/k for subjects 1-4. The LSE column gives the least square error between the obtained and predicted values of the IRT distribution conditional on the previous response being reinforced, when the data is classified into half second intervals. The last column gives the number of data points used in the parameter estimation.

When the data for subjects 5-8, with half second grouping is inspected, a slightly different picture appears. (Figure 3.4.11). Subjects 5 and 8 still show a peak in the first interval, while subjects 6 and 7 are like subjects 1-4 (figure 3.4.9) in having a peak in their IRT distribution in the second interval. Attempts to fit the model now fall into two categories. Fits to the data from subjects 6 and 7 are very poor, (see table 3.4.(ix)) while those to data from subjects 5 and 8 are still quite good. However the fits for subjects 5 and 8 are worse under the half second analysis than they were under the one second analysis.

When the conditional statistics are inspected (figure 3.4.12) the picture becomes more confused. A good fit is still obtained for subject 8, but subject 5 is now classified with subjects 6 and 7 and shows a peak in the second interval, rather than in the first. Table 3.4.(x) gives the estimates of k_1/k and the least square errors. The parameters are again all zero (compare with table 3.4.(viii))

Figure 3.4.11 -- Random ratio schedule. p = 0.1. A comparison of the data and predictions for the asymptotic IRT distribution for subjects 5-6. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one half second width.

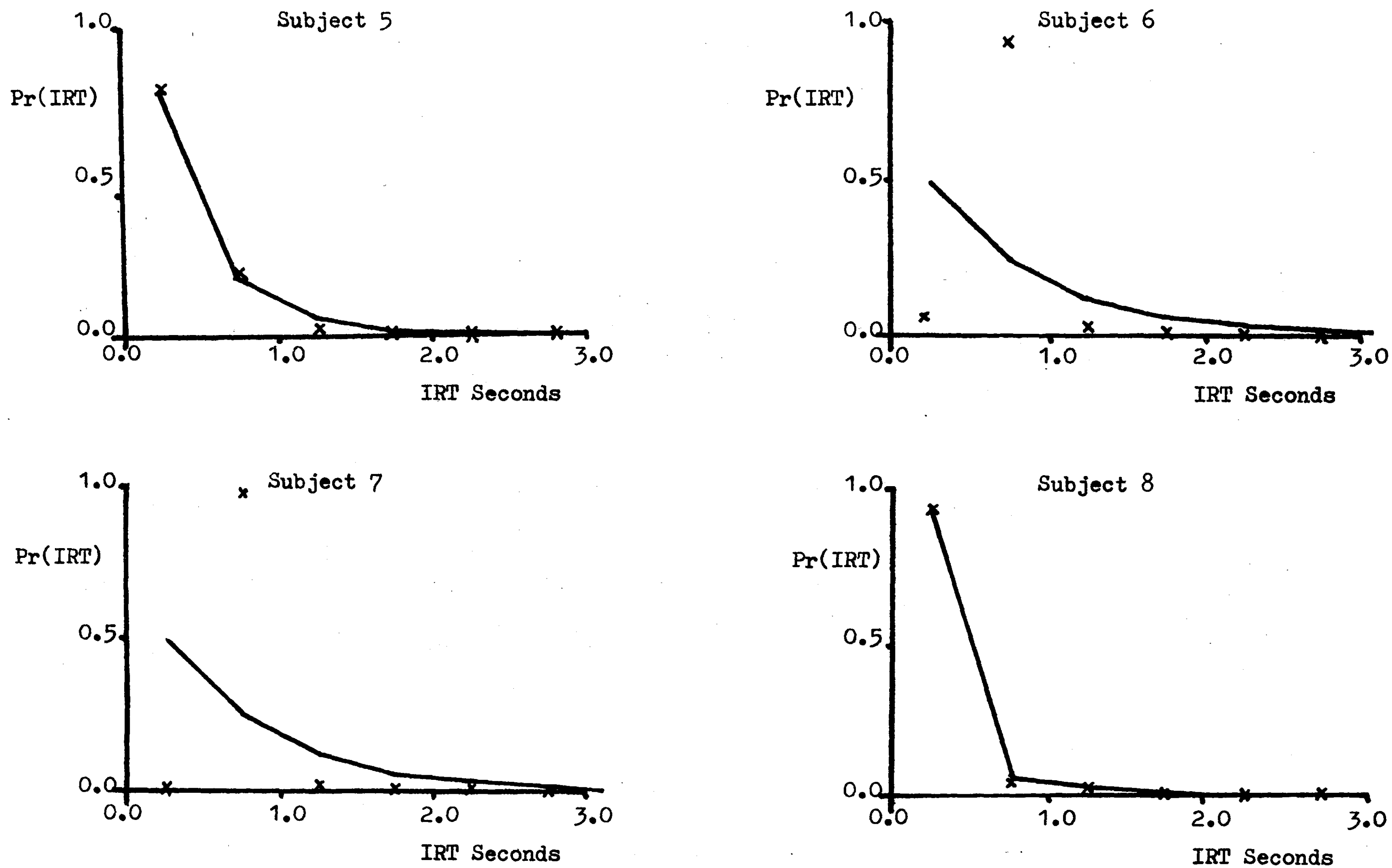
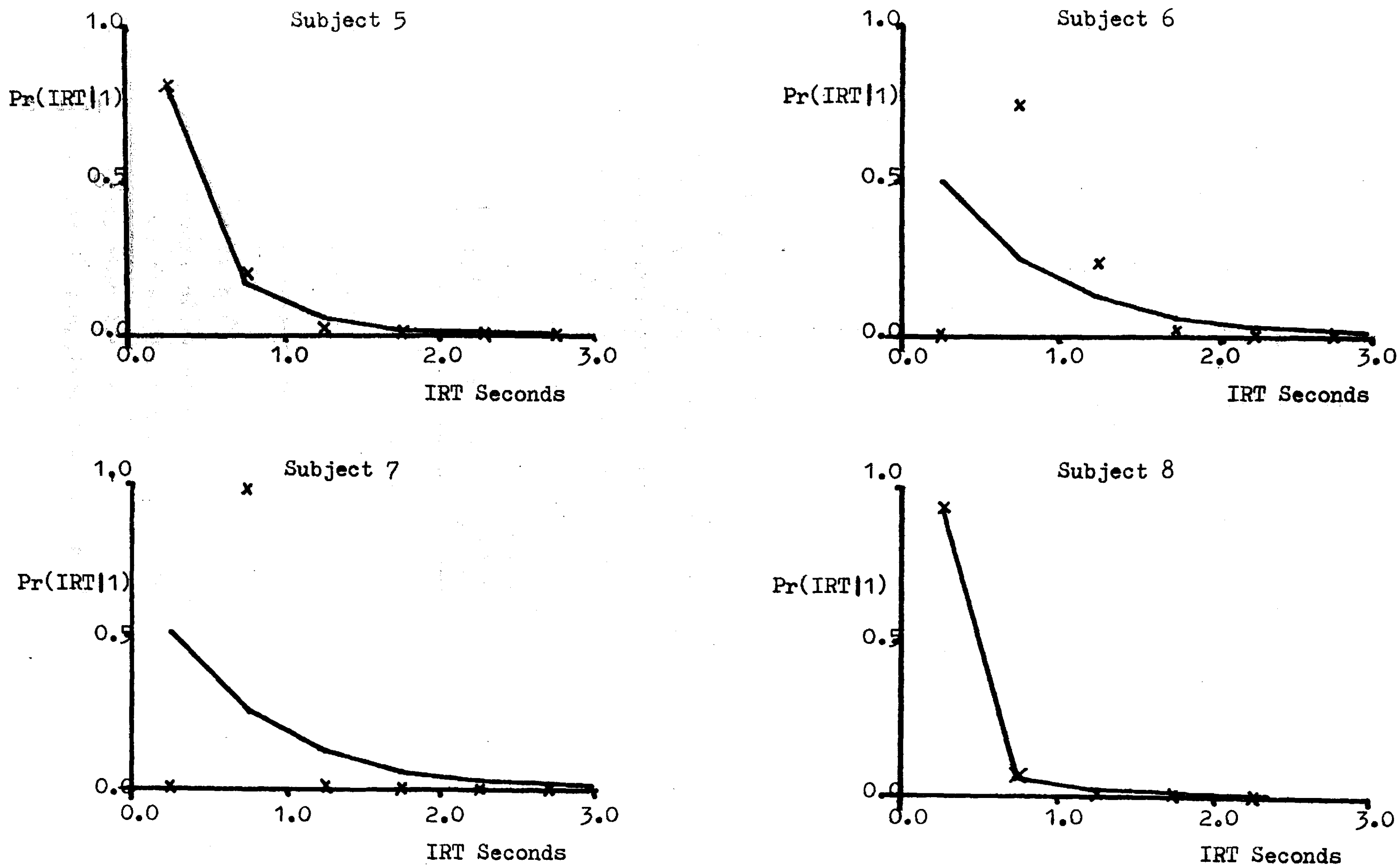


Figure 3.4 12 -- Random Ratio schedule. p = 0.1. A comparison of the data and predictions for the asymptotic IRT distribution conditional on the previous response being reinforced for subjects 5-8. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one half second width.



Subject	$\widehat{\rho k}$	LSE.10^6	No. Data Points
5	3.27	685	7
6	1.36	753855	7
7	1.38	721189	7
8	5.18	434	7

Table 3.4.(ix) -- Estimates of the parameter ρk and a list of the least square errors (LSE) between the obtained and predicted values of the asymptotic IRT distribution for subjects 5-8 when the data is classified into half second intervals. The last column gives the number of data points used in the parameter estimation.

though presumably the zero obtained for subject 8 arises from a different source than those obtained for subjects 5-7.

In view of the generally poor fits to the data, further conditional statistics are not investigated. It is obvious from the half second analysis of the data that the IRT distribution in the random ratio case contains substantial contributions from sources apart from those considered in the model. The following section investigates a particular, and probably the most important, source for this contribution.

Subject	$\widehat{k_1/k}$	LSE.10^6	No. Data Points
5	0.00	559419	7
6	0.00	510182	7
7	0.00	810817	7
8	0.00	41	7

Table 3.4.(x) -- Estimates of k_1/k for subjects 5-8. The LSE column gives the least square error between the obtained and predicted values of the IRT distribution conditional on the previous response being reinforced, when the data is classified into half second intervals. The last column gives the number of data points used in the parameter estimation.

3.5 Response Speed Effects

Ray and McGill (1964) studied the changes in the IRT distribution obtained from pigeons on a random ratio schedule when the intervals by which the IRT's were classified were progressively reduced in size. They found that as the intervals grew smaller, that the distribution finally settled out into one with two peaks, both close to zero, and tentatively ascribed these two peaks to two different response movements: a very short movement involving only the head, and a slightly longer one involving both the head and the shoulders. These results, considered together with the very high rates of responding obtained from most of the subjects used in the present experiments, suggest that an important characteristic of the IRT distribution in the random ratio case is described by the time it takes to actually make the response. This suggests that an allowance for the effects of response time should be added to the model.

The simplest approach seems to be to assume that the time taken to execute the response is basically constant, but subject to random fluctuations. A distribution with approximately these qualities is the gamma distribution. Thus, if s(t) denotes the density distribution of response times, then

$$s(t) = \frac{\beta(\beta t)^{v-1}}{\Gamma(v)} \mathrm{Exp}(-\beta t), \qquad (3.5.a)$$

where β and v are constants.

The IRT is now considered to have two components: a basic response time part - i.e. a time between the decision to initiate a response and its actual termination - and a 'decision' component, due to the learning process. (See figure 3.5.1). The 'decision' component is that described by the model given in chapter 2. This component is termed the 'decision' component because at each sampling the subject "decides", albeit in a probabilistic sense, whether or not to respond. This part of the IRT will be denoted as having density distribution r(t).

The response time is assumed to depend only on the physical parameters of the actual response required, and is independent of the decision time. It is assumed to be unaffected by learning, at least after a short practice period. The decision time is that aspect of the IRT that is affected by the reinforcement schedule.

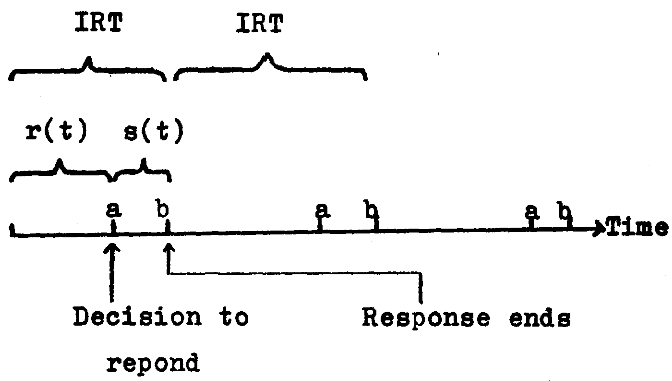


Figure 3.5.1 -- Diagramatic representation of how each IRT is divided into two components.

It is manipulated by learning, in the way given in chapter 2. Notice that the decision time is not the time required to make a choice between respond/not respond, but the time to a decision to respond, from the end of the previous response.

If q(t) is the density distribution of the IRT's, it is given by the convolution of r(t) and s(t). I.e.,

$$q(t) = \int_0^t s(x)r(t-x)dx.$$

Since the random ratio schedule is being used,

$$r(t) = \rho k \mathrm{Exp}(-\rho kt).$$

Thus,

$$q(t) = \int_0^t \beta^v \frac{x^{v-1}}{\Gamma(v)} \mathrm{Exp}(-\beta x)\, \rho k \mathrm{Exp}(-\rho k(t-x))dx$$

$$= \beta^v \rho k \mathrm{Exp}(-\rho kt) \int_0^t \frac{x^{v-1}}{\Gamma(v)} \mathrm{Exp}(-(\beta - \rho k)x)dx.$$

Assuming that $\beta \neq \rho k$, then this is,

$$q(t) = \left[\frac{\beta}{\beta-\rho k}\right]^v \rho k \mathrm{Exp}(-\rho kt)\int_0^t (\beta-\rho k)^v \frac{x^{v-1}}{\Gamma(v)}\mathrm{Exp}(-(\beta-\rho k)x)dx$$

$$= \left[\frac{\beta}{\beta-\rho k}\right]^v \rho k \mathrm{Exp}(-\rho kt)\int_0^{(\beta-\rho k)t} \frac{y^{v-1}}{\Gamma(v)}\mathrm{Exp}(-y)dy$$

Now $\int_0^t \frac{y^{v-1}}{\Gamma(v)}\mathrm{Exp}(-y)dy$ is the normalised incomplete gamma function which will be denoted by $G_v(t)$. Thus,

$$q(t) = \left[\frac{\beta}{\beta-\rho k}\right]^v \rho k \mathrm{Exp}(-\rho kt) G_v((\beta-\rho k)t),$$

when $\beta \neq \rho k$.

and,

$$q(t) = \beta\frac{(\beta t)^v}{(v-1)}\mathrm{Exp}(-\beta t),$$

when $\beta = \rho k$.

To obtain an idea of the relative importance of the response speed effects on the overall IRT, it is of interest to look at the contribution of the response speed to the theoretical mean and variance of the IRT distribution, and compare the theoretical mean and variance of the IRT distribution with the values actually obtained.

The calculation of the predicted values of the mean and variance, in terms of ρk, β, and v is quite simple. Let $M_q(x)$, $M_r(x)$, and $M_s(x)$ denote the moment generating functions of, $q(t)$, $r(t)$, and $s(t)$, respectively. Then the Faltung theorem says,

$$M_q(x) = M_r(x)M_s(x).$$

In the present case, from the well-known property of the gamma distribution,

$$M_s(x) = \left[\frac{\beta}{\beta - x}\right]^v$$

and,

$$M_r(x) = \frac{\rho k}{\rho k - x}.$$

From the properties of moment generating functions,

$$E(q(t)) = \left.\frac{d}{dx}M_q(x)\right|_{x=0},$$

and,

$$Var(q(t)) = \left.\frac{d^2}{dx^2}M_q(x)\right|_{x=0} - E(q(t))^2.$$

This gives,

$$E(q(t)) = \frac{1}{\rho k} - \frac{v}{\beta},$$

and,

$$Var(q(t)) = \frac{1}{(\rho k)^2} - \frac{v}{\beta^2}.$$

I.e. the means and variances are the sums of the means and variances of the two components.

The parameters ρk, β, and v were estimated from the asymptotic IRT distribution by the method of least squares. Figures 3.5.2, and 3.5.3 illustrate the comparison of the obtained and predicted IRT distributions for subjects 1-8. In all cases, whether or not the

Subject	ρk	β	v	LSE.10^6	No. Data Points
1	21.22	27.78	15.02	22	7
2	31.31	32.50	20.28	90	7
3	35.68	37.61	23.38	334	7
4	58.12	61.26	41.42	18	7
5	6.23	6.28	1.10	190	7
6	61.31	64.55	50.65	4	7
7	65.99	74.80	53.78	53	7
8	5.97	5.97	1.19	39	7

Table 3.5.(i) -- Estimates of the parameters ρk, β, and v, together with a list of the least square error (LSE) between the obtained and predicted values for the asymptotic IRT distribution when allowance is made for the response speed. The data was classified into half second intervals. The last column gives the number of data points used in the parameter estimation.

Figure 3.5.2 -- Random ratio schedule. p = 0.5. A comparison of the data and predictions for the asymptotic IRT distribution with an allowance for response speed effects, for subjects 1-4. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one half second width.

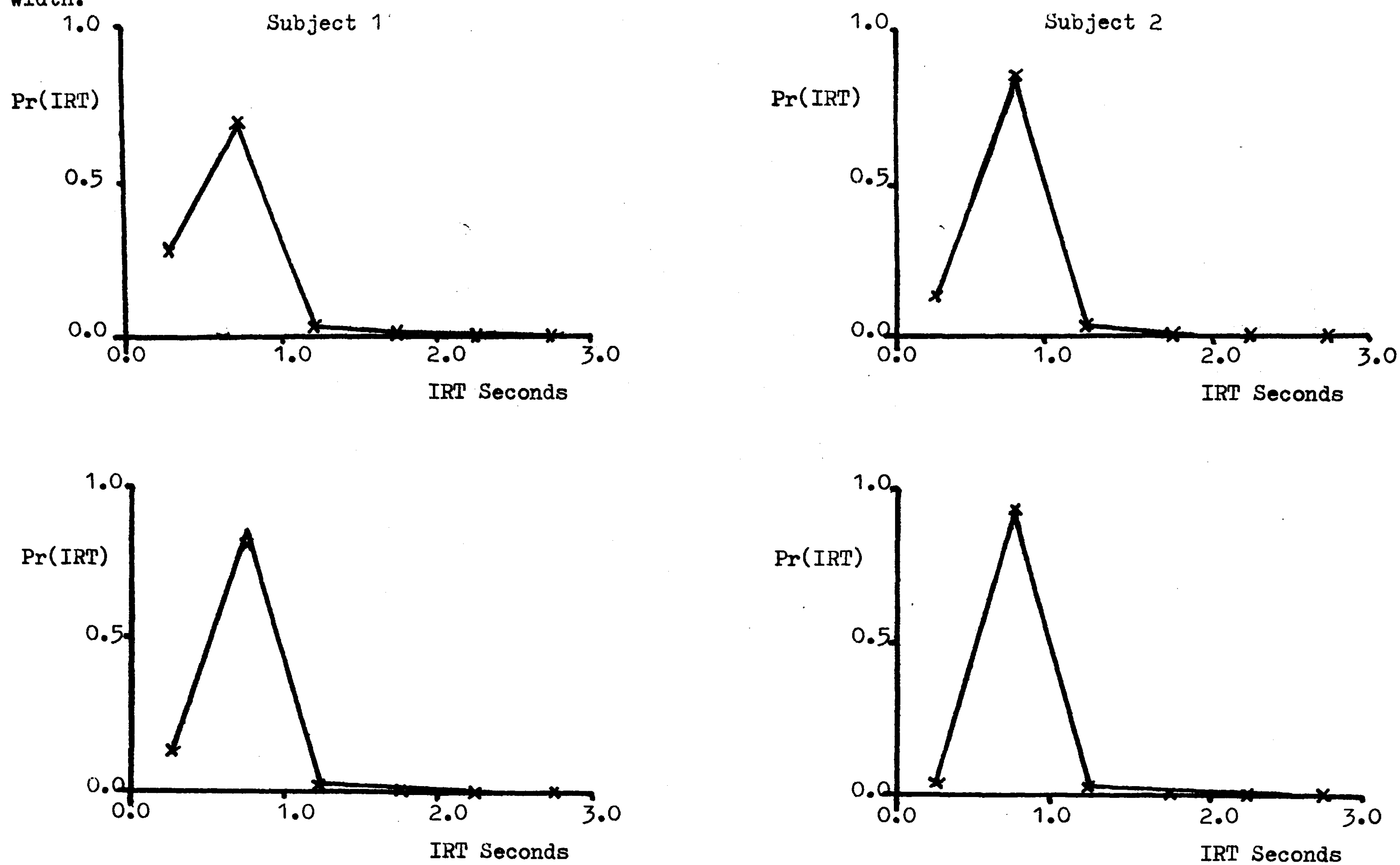
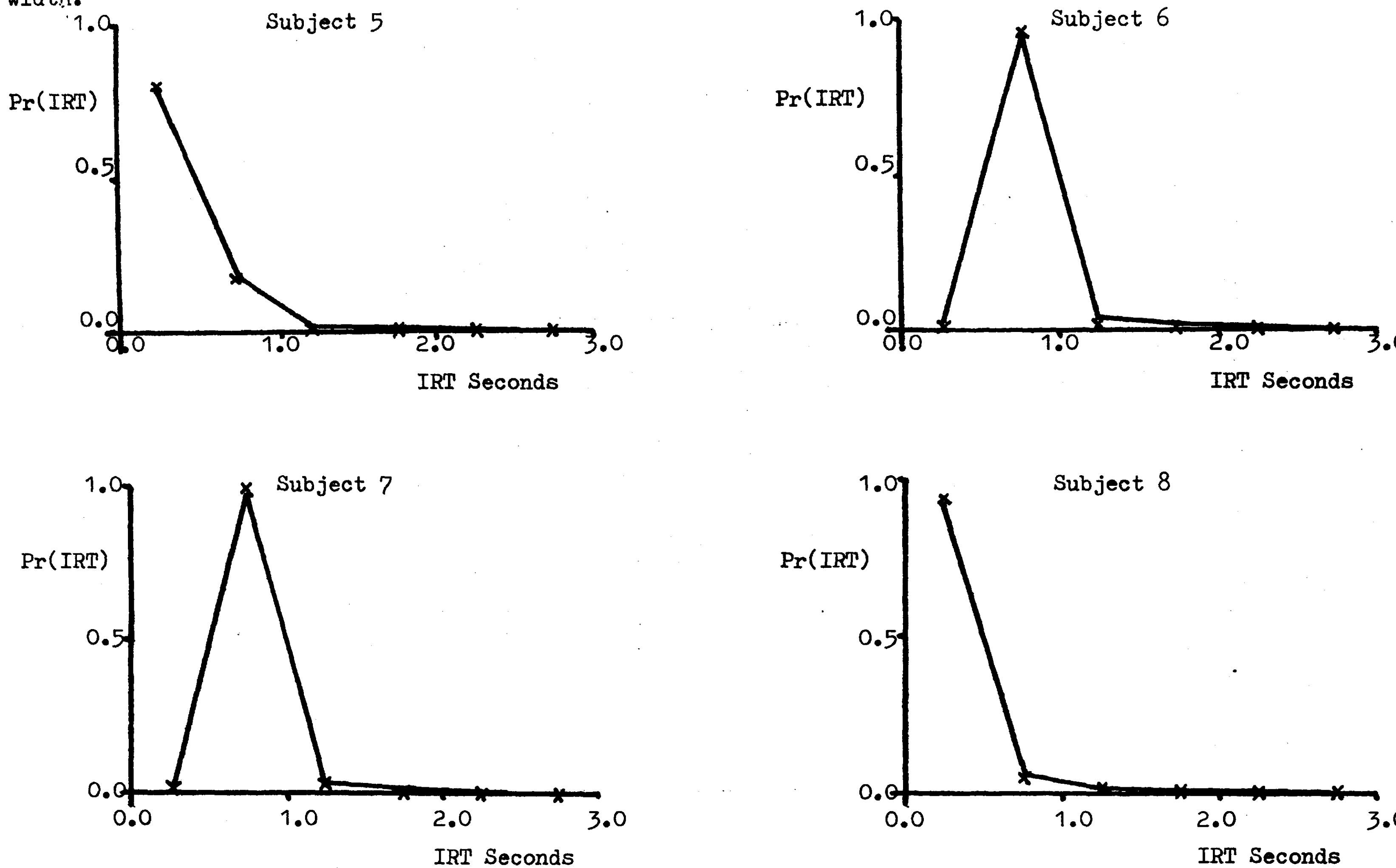


Figure 3.5.3 -- Random ratio schedule. p = 0.1. A Comparison of the data and predictions for the asymptotic IRT Distribution with an allowance for response speed effects for subjects 5-8. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of on half second width.



distributions have a peak in the second interval (all subjects except subjects 5 and 8), the fits are excellent. Table 3.5.(i) gives the obtained values of the parameters, and lists the least square error. In all cases, the errors are small. All the parameters fail to show any kind of uniformity however, and the range of values is large.

Sub	E(r(t))	E(s(t))	E(q(t))	E(IRT)	V(r(t))	V(s(t))	V(q(t))	V(IRT)
1	0.047	0.541	0.588	0.523	0.002	0.019	0.021	0.010
2	0.032	0.628	0.660	0.579	0.001	0.019	0.020	0.139
3	0.028	0.631	0.659	0.613	0.000	0.017	0.017	0.125
4	0.017	0.661	0.678	0.642	0.000	0.011	0.011	0.096
5	0.161	0.175	0.336	0.463	0.026	0.028	0.053	0.027
6	0.016	0.809	0.825	0.759	0.000	0.013	0.013	0.018
7	0.015	0.699	0.714	0.696	0.000	0.009	0.009	0.006
8	0.168	0.032	0.200	0.422	0.028	0.001	0.028	0.020

Table 3.5.(i) -- Comparison of the theoretical and actual means (E) and variances (V), together with the breakdown of the theoretical values into their two components, for subjects 1-8.

The comparison of the predicted means and variances with the ones actually obtained are given in table 3.5.(ii). They reveal some interesting details. On the whole, the obtained and predicted mean values are very close, with the largest component been given by the response time. In contrast, the variances show a large discrepancy in a number of cases, where the obtained value is much larger than the theoretical one. Close inspection of the data suggests this is due to the persistence of a small number of relatively long IRT's (about 0.5%), which few IRT's contribute substantially to the variance of the IRT distribution.

Sequential effects can be investigated with little further work, as most of the required equations are already set up. It has been shown that, (section3.2)

$$r(t|i) = (\rho k+\rho k_1 \mathrm{Exp}(-\rho kt))\mathrm{Exp}(-(\rho kt+(k_i/k)(1-\mathrm{Exp}(-\rho kt)))),$$

(where i = 1 means that the previous response was reinforced, and

$i = 0$, that it was not reinforced, and,

$$k_i = 2(\alpha^i - k)\Theta_i sh(\rho ka).)$$

The response speed effect has been defined as,

$$s(t) = \frac{\beta(\beta t)^{v-1}}{(v)} Exp(-\beta t).$$

Since the response speed is presumed unaffected by reinforcement or non-reinforcement, the total conditional IRT is given by,

$$q(t|i) = \int_0^t r(x|i)s(t-x)dx.$$

Table 3.5. (iii) gives values of k_1/k calculated by the method of least squares, for subjects 1-8. Comparison of table 3.5.(iii) with table 3.5.(i) shows that the fits to the asymptotic IRT distribution conditional on the previous response being reinforced are worse than the fits to the asymptotic IRT distribution. (Only subject 7 shows an improvement). Some subjects, e.g. 5 and 6 have data which shows a massive deterioration in the fit. Part of

Subject	$\widehat{k_1/k}$	LSE.10^6	No. Data Points
1	0.00	4108	7
2	0.00	7076	7
3	34.74	1162	7
4	2.09	86	7
5	0.00	539837	7
6	0.00	77958	7
7	0.00	21	7
8	0.00	85	7

Table 3.5.(iii) -- Estimates of k_1/k for subjects 1-8, after allowance for response speed effects. LSE column gives the least square error between obtained and predicted values of the IRT distribution conditional on the previous response being reinforced. The last column gives the number of data points used in the parameter estimation.

Figure 3.5.4 -- Random ratio schedule. p = 0.5. A comparison of the data and predictions for the asymptotic IRT distribution conditional on the previous response being reinforced and with an allowance for response speed effects for subjects 1-4. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one half second width.

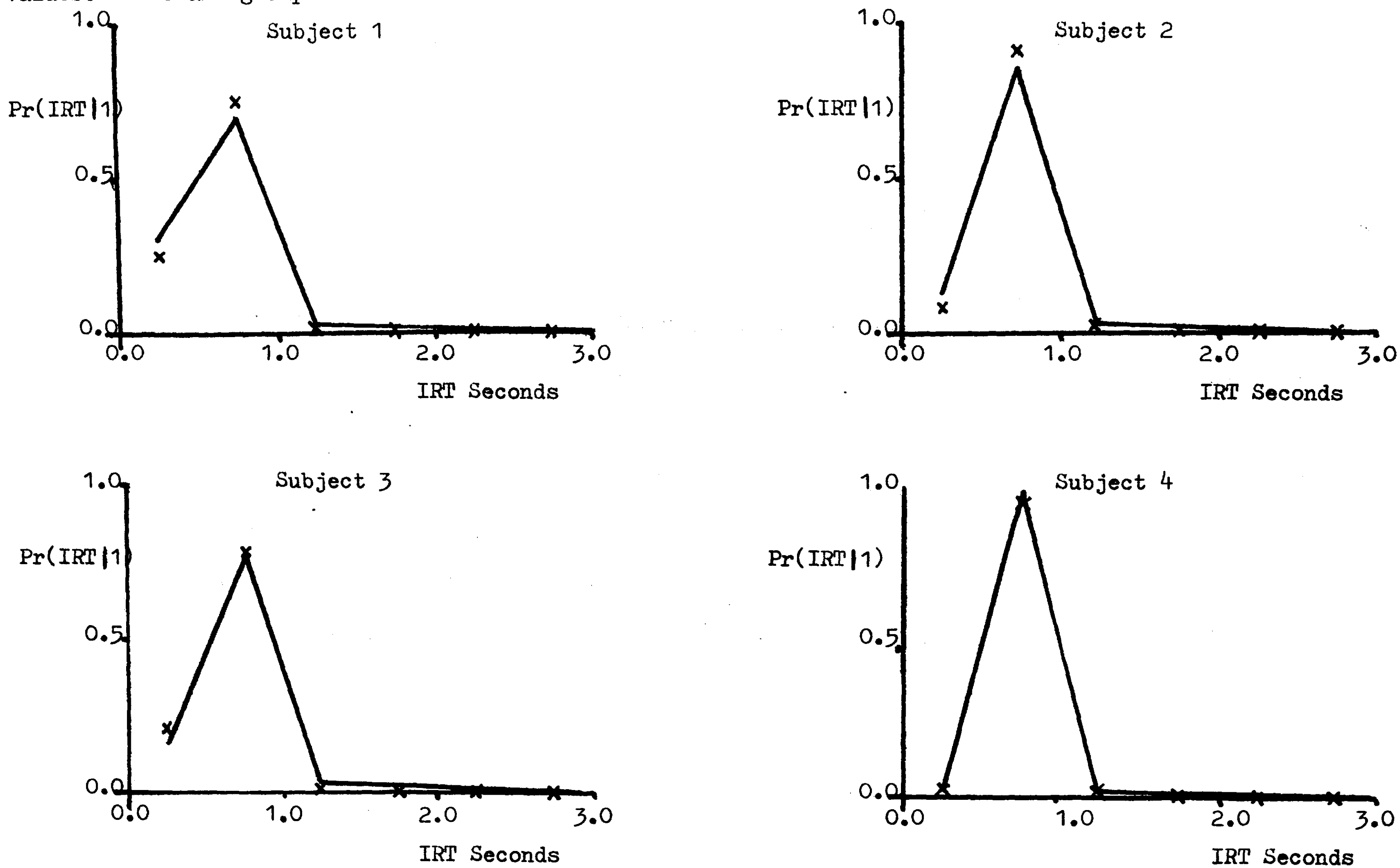
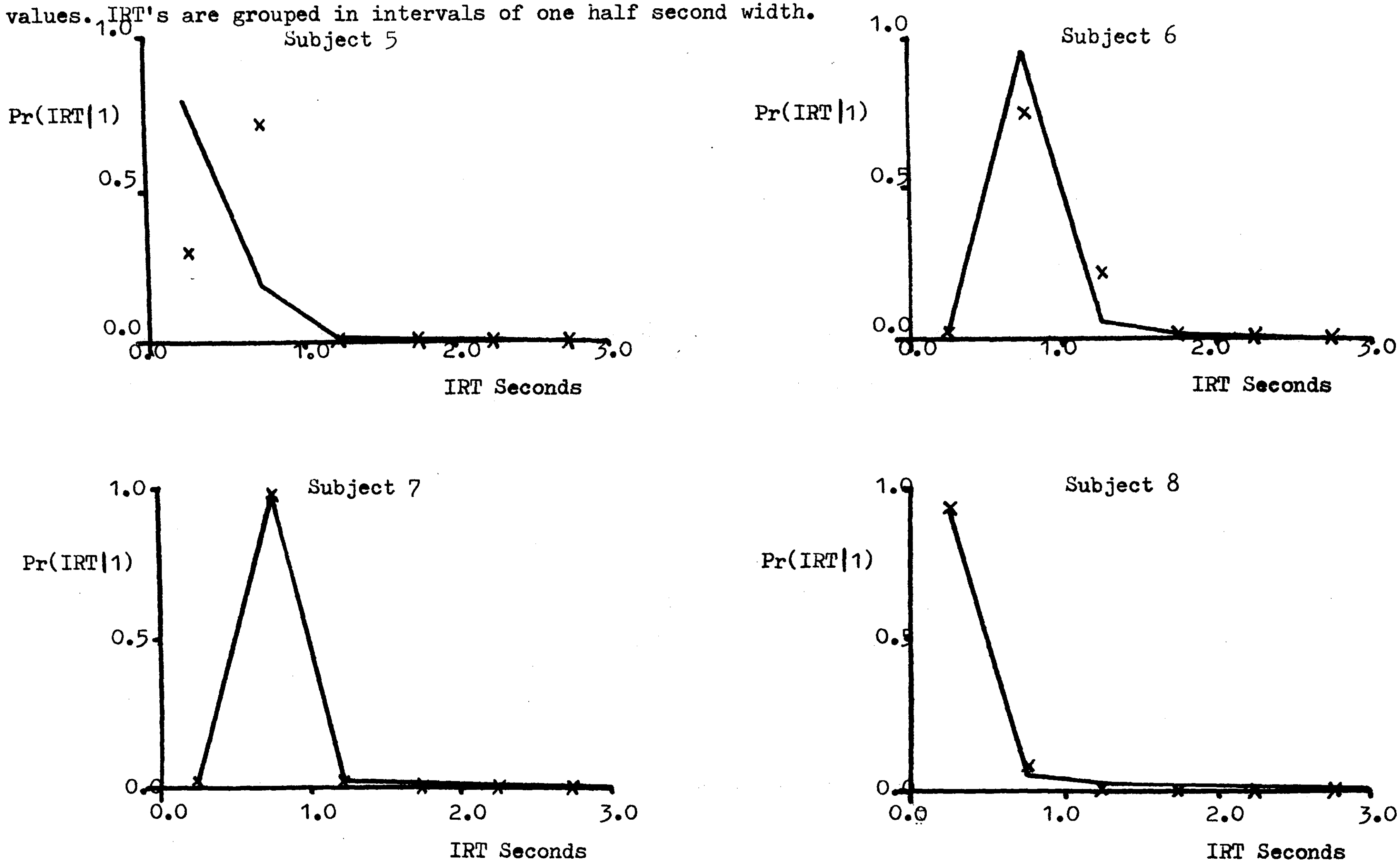


Figure 3.5.5 -- Random ratio schedule. p = 0.1. A comparison of the data and predictions for the asymptotic IRT distribution conditional on the previous response being reinforced and with an allowance for response speed effects for subjects 5-8. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one half second width.



the deterioration can be accounted for through the large number of zero entries in table 3.5.(iii) for $\widehat{k_1/k}$. These represent not, as might be expected, no sequential effects, but for subjects 1, 2, 5, and 6, situations where the sequential effects are in the opposite direction to that predicted in the model. The effect of reinforcement is, on average, predicted as a speeding up of the rate of responding, but subjects, 1, 2, 5, and 6 in fact slowed down after a reinforcement. Better fits could have been obtained by allowing k_1/k to become negative, but this would be meaningless within the framework of the model. This finding, of a slowing down effect of reinforcement stands in strong contrast to what is commonly presumed to be one of the chief effects of reinforcement, an increase in the rate of responding.

Figures 3.5.4 and 3.5.5 give a graphical illustration of the results shown in table 3.5.(iii). A comparison of these figures, with those of 3.5.2 and 3.5.3 shows clearly the slowing down effect after reinforcement, for subjects 1, 2, 5, and 6. The effect is especially striking for subject 5, where the peak moves from

Subject	$\widehat{k_0/k}$	$SE.10^6$	No. Data Points
1	-0.00	4732	7
2	-0.00	9500	7
3	-34.74	1922	7
4	-2.09	474	7
5	-0.00	13102	7
6	-0.00	1202	7
7	-0.00	34	7
8	-0.00	38	7

Table 3.5.(iv) -- Estimates of k_0/k for subjects 1-8, after allowance for response speed effects. SE column gives the square error between obtained and predicted values of the IRT distribution conditional on the previous response being reinforced. The last column gives the number of data points used to calculate the SE.

---

between 0.0-0.5 seconds to between 0.5-1.0 seconds for responses that follow reinforcement.

The parameter k_0/k can be found from $\widehat{k_1/k}$ by the use of relation

Figure 3.5.6 -- Random ratio schedule. p = 0.5. A comparison of data and predictions for the asymptotic IRT distribution conditional on the previous response not being reinforced for subjects 1-4. There is an allowance for response speed effects. Solid lines join the predicted points. Crosses mark the obtained data values. IRT's are grouped in intervals of one half second width.

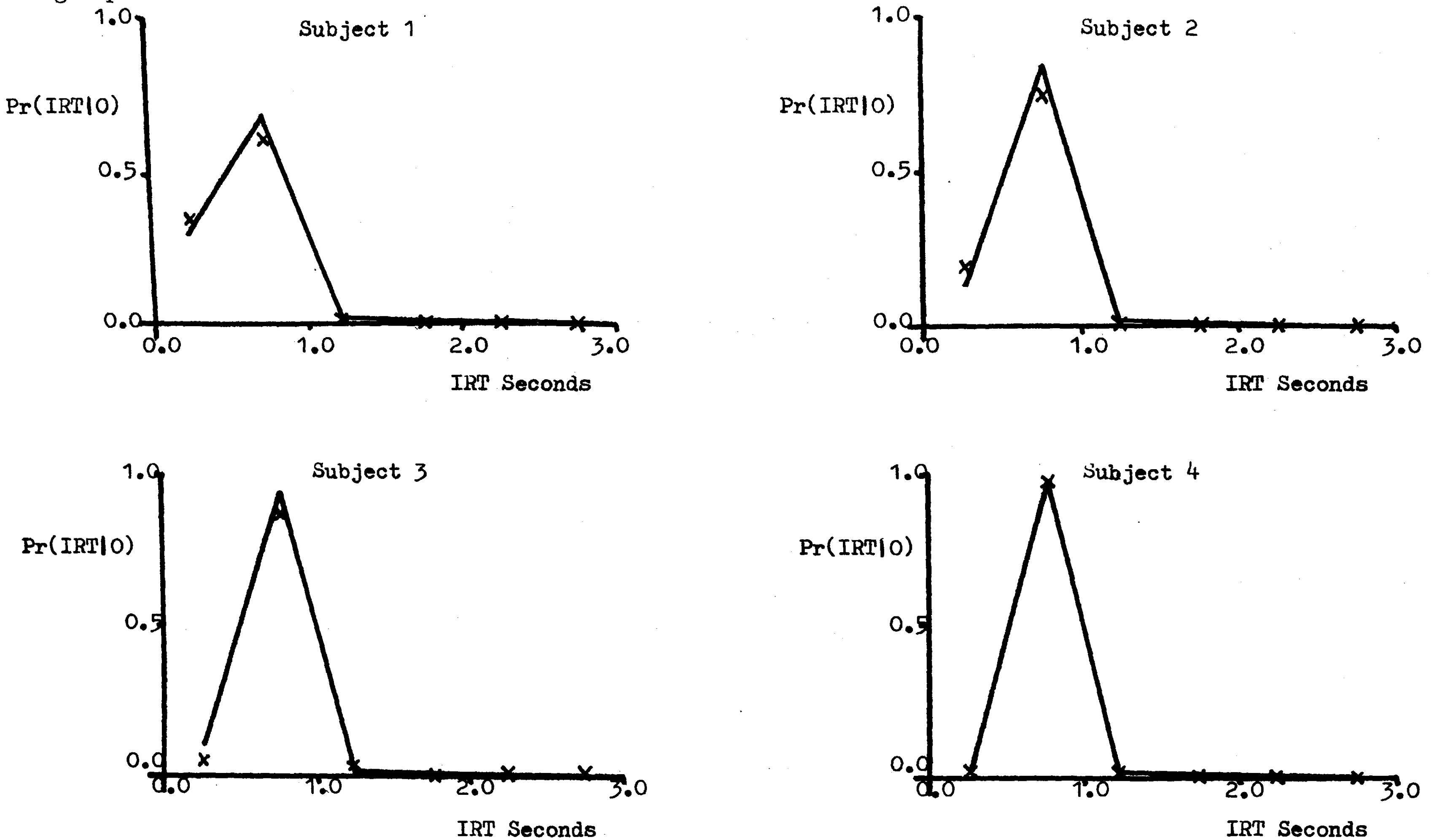
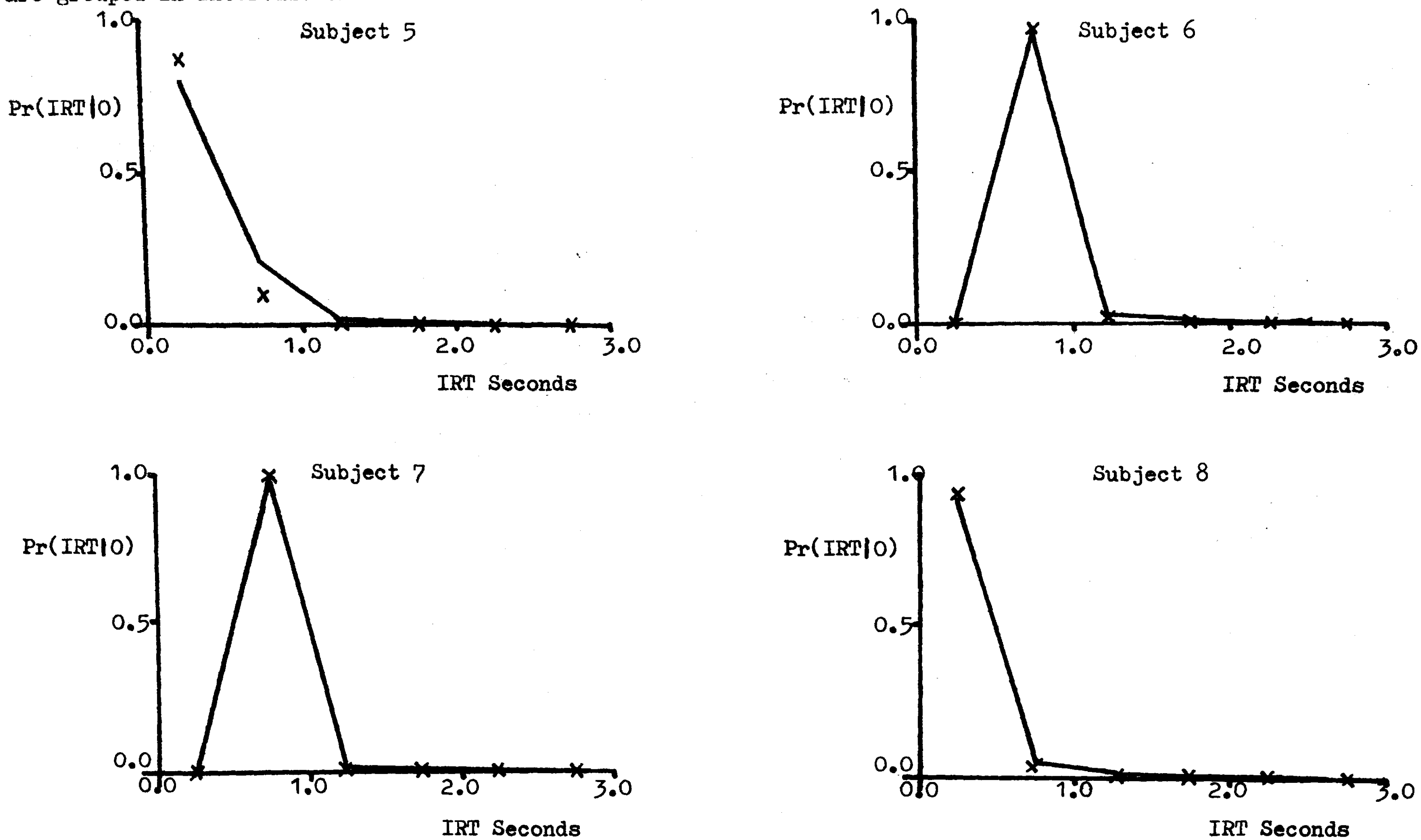


Figure 3.5.7 -- Random ratio schedule. p = 0.1. A comparison of data and predictions for the asymptotic IRT distribution conditional on the previous response not being reinforced for subjects 5-8. There is an allowance for response speed effects. Solid lines join predicted points. Crosses mark the obtained data values. IRT's are grouped in intervals of one half second width.



3.2.b, and the appropriate results are given in table 3.5.(iv). These values were used to calculate the asymptotic IRT distributions conditional on the previous response not being reinforced, and these distributions are illustrated in figures 3.5.6 and 3.5.7. Table 3.5.(iv) also gives the square error between the obtained and the predicted values. (These are not fitted curves. k_0/k is not estimated from this data). On the whole, the fits are in fact better than those for the distribution conditional on the previous response being reinforced. This may in part be a reflection of the fact that for subjects 5-8, there are approximately nine times as many non-reinforcements as reinforcements, and thus the probabilities estimated from the data for IRT's conditional on non-reinforcement have a much lower variance than those probabilities estimated for IRT's conditional on a reinforcement for these subjects.

3.6 Conclusion

A number of points arise from the previous sections. The first is that when dealing with a data continuum (in this case time), decisions as to how to separate the data up into class intervals can seriously affect the character of the results obtained. In general, the finer the interval width, the better. However, constraints are placed upon this by the accuracy of the observations and the number of observations available.

The second point is that, at least for the schedule values used here, (0.5 and 0.1) the time taken to produce the response occupies by far the largest part of any measured IRT. This means that effects dealt with by the model, are, at best, very small. This in turn implies that so long as the response speed effects are adequately described, almost any model that has negligable significance for some values of its parameters may produce a reasonable fit.

Viewed in this light, it is difficult to assess the model. With allowance for response speed, the fits are good. However, if attention turns to the conditional statistics, the results are, on the whole mediocre. The worst of the conditional fits describe cases where the effects of reinforcement and non-reinforcement seem to have been inverted, and the experimental results do not seem to lie in line with any previous reports, particularly of animal behaviour. It may be possible however to ascribe this effect to some temporary influence on the sampling behaviour. (Immediately after a reinforcement 'don't bother looking' (sampling) for a while, while after a non-reinforcement, 'look harder'.)

The overall conclusion is that the results are reasonable, and interesting, but not outstanding.

Chapter 4

DIFFERENTIAL REINFORCEMENT OF LOW RATES OF RESPONDING

4.1 Previous Experimental Findings

The differential reinforcement of low rates of responding form the most obvious class of IRT reinforcement schedules. They are usually defined by,

$$u(t) = \begin{cases} 1.0, & t \geq d \\ 0.0, & t < d, \end{cases}$$

where d is a constant. Such a schedule essentially reinforces pauses of duration d, to produce low rates of responding, hence the name.

These schedules are part of a more general class of discontinuous schedules, which may be defined as follows,

$$u(t) = \begin{cases} p, & t \geq d \\ q, & t < d, \end{cases}$$

where d, p, and q are constants. This schedule will be written in the form DRLd;p,q. The form DRLd;p will imply that $q=0.0$, while the form DRLd will imply that $p=1.0$, and $q=0.0$. When $p=q$, the schedule is simply a random ratio schedule. It is to be assumed that $p \geq q$. If $p<q$, the schedule is usually referred to as the differential reinforcement of high rates of responding, (DRH), (Shimp, 1967).

Early work on DRL schedules was done by Sidman (1956). His initial observation of the development of a mode in the IRT distribution near to the cutoff value d has since been confirmed by many other experimenters. The other characteristic of DRLd schedules is the persistence of a maximum close to $t=0.0$ in the IRT distribution. Not all experimenters have observed this phenomenon, though the majority seem to have done so.

Little work has been done on the properties of sequential effects in DRL schedules. However, Farmer and Schoenfeld (1964) have reported that reinforced responses are more likely to be followed by a reinforced response than non-reinforced responses by a reinforced response, and this has been confirmed by Ferraro, Schoenfeld and Snapper (1965).

4.2 Predictions from the Model

The derivation of the asymptotic IRT distribution is complicated in the DRL case by the discontinuity in the reinforcement function u(t) at d. Using the same spread and sampling functions as previously, then r(t) takes different forms, depending on the value of t.

i) $t \leq d-a$.

From equation 2.6.b,

$$\phi(t) = \frac{\alpha^0 \theta \int_{t-a}^{t+a} (1-q)r(y)dy + \alpha^1 \int_{t-a}^{t+a} qr(y)dy}{\theta \int_{t-a}^{t+a} (1-q)r(y)dy + \int_{t-a}^{t+a} qr(y)dy}$$

$$= \frac{\alpha^0 \theta(1-q) + \alpha^1 q}{\theta(1-q) + q} .$$

This is a constant, that will be denoted by k(q). r(t) is then given by,

$$r(t) = \rho k(q) \mathrm{Exp}(-\rho k(q)t).$$

ii) $t \geq d+a$.

Here the situation is parallel to that for $t \leq d-a$, except that p replaces q. Hence,

$$\phi(t) = k(p),$$

and,

$$r(t) = A \rho k(p) \mathrm{Exp}(-\rho k(p)(t-d+a)),$$

where A is a constant given by,

$$A = \mathrm{Exp}\left(-\int_0^{d-a} \rho\phi(y)dy\right).$$

iii) $d-a<t<d+a$.

From equation 2.6.b, and dividing each integral into two parts at the discontinuity d,

$$\phi(t) = \frac{\alpha^0\theta\int_{t-a}^{d}(1-q)r(y)dy+\alpha^0\int_{d}^{t+a}(1-p)r(y)dy+\alpha^1\int_{t-a}^{d}qr(y)dy+\alpha^1\int_{d}^{t+a}pr(y)dy}{\theta\int_{t-a}^{d}(1-q)r(y)dy+\theta\int_{d}^{t+a}(1-p)r(y)dy+\int_{t-a}^{d}qr(y)dy+\int_{d}^{t+a}pr(y)dy}$$

Let,

$$d(t) = \frac{\int_{t-a}^{d} r(y)dy}{\int_{d}^{t+a} r(y)dy},$$

then $\phi(t)$ can be written,

$$\phi(t) = \frac{\alpha^0\theta((1-q)d(t)+(1-p)) + \alpha^1(qd(t)+p)}{\theta((1-q)d(t)+(1-p)) + (qd(t)+p)}. \qquad 4.2.a$$

This is not explicitly soluble for $\phi(t)$. For the interval $d-a<t<d+a$, $r(t)$ must be found by using a version of the SAGENAR programme, described in appendix I.

Since $\phi(t)$ is a constant, outside $(d-a,d+a)$, and,

$$r(t) = \rho\phi(t)\mathrm{Exp}\left(-\int_0^{t}\rho\phi(y)dy\right),$$

then on interval $(0,d-a)$, $r(t)$ has a maximum at $t=0.0$ and a minumum at $t=d-a$. Likewise, on interval $(d+a,\infty)$, $r(t)$ has a maximum at $d+a$ and a mimumum at infinity. This means that $r(t)$ as a whole has maxima at $t=0.0$ and at some point in the interval $(d-a,d+a)$, including its end points. I.e. there are maxima around zero and around the cutoff point d. Thus $r(t)$ has the required kind of bimodality.

However, the situation is not as good as it might appear.

Figure 2.7.1 illustrates a predicted IRT distribution where α^0=0.02 and q=0.0. Although the distribution is bimodal, the slope of the curve from t=0.0 to t=4.5 is far too gentle, and the values of r(t) in the range t=1.5 to t=4.5 are too high to reliably describe the short IRT's which occur in practice on DRL schedules. It thus seems likely that they arise from other sources than the one considered in the model, (Essentially the model accounts for their existence by setting α^0>0.0.)

There appears to be at least two other ways in which it is possible to account for the peak at short IRT's. These are,

i) Suppose that any sample of stimuli can be divided into two parts, these being,

a) Stimuli generated by the preceeding response.

b) Other stimuli. (Background stimuli.)

It is possible to consider that the background stimuli consist of stimuli from a whole series of preceeding responses, and which are thus characteristic of 'long' IRT's. If a sample happens to contain a large proportion of these stimuli, which are presumably conditioned to the response, then a response will be made, as the subject effectively misreads the time. To some extent the assumption that α^0>0.0 is equivalent to this. However if most of the background stimuli are specifically presumed to come from the preceeding one or two responses, then the result of a burst of short responses is to effectively wipe out the stimuli concerned with long IRT's. Regarded in this way, this idea is then very similar to the idea of Sidman (1956) that a burst of short IRT's resets the internal clock.

ii) An alternative way of tackling the problem is to look at the idea of 'resetting the internal clock' in a somewhat different light than above. The model described in chapter 2 assumed that after each response, that when the subject next samples, the sample is taken from near X(0). (I.e. the clock 'resets', or the new stimulus trace wipes out the old one.) This need not happen however. This may fail to happen and the next sample be taken from X(t+t'), (t is the previous IRT, t' is the time since the last response.) rather than from X(t'). If t $\simeq$d, then a response is likely to follow very shortly, as at the next sampling of stimuli, ϕ(t) is likely to be large. A burst of responses could ensue before the subject reverts to the beginning of the X-continuum.

If say the two previous IRT's were t_1 and t_2, then the effective value of $\emptyset(t)$, $\emptyset_e(t)$, is either $\emptyset(t)$, or $\emptyset(t+t_1)$, or $\emptyset(t+t_1+t_2)$, or, etc. If these are averaged over all the possible values of t_1, t_2, etc., they become,

$$\emptyset(t), \text{ or } \int_0^\infty \emptyset(t+t_1)r(t_1)dt_1, \text{ or } \int_0^\infty \emptyset(t+t_1+t_2)r(t_1)r(t_2)dt_1dt_2,$$

etc.

Let z be the probability that the 'clock is reset'. The average value of $\emptyset_e(t)$ is then given by,

$$\emptyset_e(t) = z\emptyset(t) + (1-z)\int_0^\infty \emptyset(t+t_1)r(t_1)dt_1 + \ldots \text{ etc.}$$

Thus in general,

$$\emptyset_e(t) = z\emptyset(t) + \sum_{k=1}^{\infty}(1-z)^k\int_0^\infty \emptyset(t + \sum_{i=1}^{k}t_i)\prod_{j=1}^{k}r(t_j)dt_j.$$

When $\emptyset(t)$ is a constant, as in the random ratio case, this simply reduces to,

$$\emptyset_e(t) = \frac{(1-z-z^2)\emptyset(t)}{z},$$

which is also a constant. In this situation, the proposed effect is undetectable. Consider the DRL case however with $\alpha^0=0.0$ and $q=0.0$. In this case, $\emptyset(t)$ is zero when $t<d-a$, and the simple model, $(z=1.0)$ predicts that no IRT's are less than d-a long. When the case $z \neq 1.0$ is considered, however, the limits of integration become $(d-a-t,\infty)$ in the expression for $\emptyset_e(t)$. Inspection of the expressions given for $\emptyset(t)$ at the beginning of this section show that $\emptyset(t)$ is monotonic non-decreasing. Thus, as t increases from 0.0 to d-a, the range of integration in the expression for $\emptyset_e(t)$ increases, and so does $\emptyset_e(t)$. Thus over interval (0,d-a), $\emptyset_e(t)$ is monotonic non-decreasing. Hence r(t), which is now given by,

$$r(t) = \rho\emptyset_e(t)\text{Exp}(-\int_0^t \rho\emptyset_e(y)dy),$$

is monotonic decreasing on this interval. I.e. there is a peak for short IRT's, even if $\alpha^0=0.0$ and $q=0.0$.

The introduction of the z-effect into the model introduces great difficulties into the mathematics, and must unfortunately be ignored in later sections. It was included here to illustrate along with (i), the many ways in which it is possible to account for certain aspects of a given set of experimental results. This particular explanation will be neglected from now on.

One aspect of DRL performance that has received a great deal of attention is the problem of collateral behaviour. Many researchers have suggested that the timing behaviour (i.e. a maximum near to $t=d$ in the IRT distribution) is not a true timing behaviour, but a by-product of a stereotyped chain of responses that have grown up more or less fortuitously. The model in use here makes no predictions or comments about collateral behaviour. The existence of collateral behaviour simply implies that the subject has never learnt to sample from the appropriate stimulus continuum. It has been assumed throughout that subjects do select the appropriate stimulus continuum. If they do not, then the model makes no predictions about the ensuing behaviour.

As $\phi(t)$ and $r(t)$ are not capable of explicit description, the derivation of explicit formula for the sequential probabilities is also impossible. It is, though, possible to derive expressions for $r(t|1)$ and $r(t|0)$ solely in terms of $r(t)$ and the parameters of the model . This is done by eliminating $\phi(t)$ from among the following relations.

$$\phi(t) = \frac{r(t)}{\rho(1-R(t))},$$

$$\phi(t|i) = \phi(t) - (\alpha^i - \phi(t))\Theta_i\int_{t-a}^{t+a} r(y)dy, \qquad i=0,1,$$

and using,

$$r(t|i) = \rho\phi(t|i)\text{Exp}(-\int_0^t \rho\phi(y|i)dy), \qquad i=0,1.$$

For practical calculations of these conditional IRT distributions, the programme SACONDAR (appendix I) can be used.

It is possible to derive conditions under which the observations of Farmer and Schoenfeld (1964), that reinforced responses are more likely to follow reinforced responses than non-reinforced responses is true. The situation considered was one where p=1.0 and q=0.0. Let the cutoff value be d. Farmer and Schoenfeld's conclusion is thus restated as,

$$\frac{\int_d^\infty\int_d^\infty r(t|t')r(t')dt'dt}{(1 - R(d))} > \frac{\int_d^\infty\int_0^d r(t|t')r(t')dt'dt}{R(d)} . \qquad 4.2.b$$

Now let,

$$B_1(t;t') = \rho(\alpha^1-\phi(t))\theta_1 w(t;t').$$

($B_1(t;t')$ thus represents the effect of a reinforced IRT of duration t', multiplied by ρ.)

Then,

$$r(t|t') = (\rho\phi(t)+ B_1(t;t'))\mathrm{Exp}(-\int_0^t \rho\phi(y)+B_1(y;t')dy),$$

and,

$$\int_d^\infty\int_d^\infty r(t|t')r(t')dt'dt$$

$$= \int_d^\infty\int_d^\infty(\rho\phi(t)+B_1(t;t'))\mathrm{Exp}(-\int_0^t \rho\phi(y)+B_1(y;t')dy)r(t')dt'dt.$$

Integrating with respect to t gives,

$$\int_d^\infty\left[-\mathrm{Exp}(-\int_0^t \rho\phi(y)+B_1(y;t')dy\right]_d^\infty r(t')dt'$$

$$= (1-R(d))\int_d^\infty \mathrm{Exp}(-\int_0^d B_1(y;t')dy)r(t')dt'.$$

Now the value of the above integral depends on the value of the exponential term, and the value of the exponential depends in turn on the function chosen for w(t;t'). The exponential has its largest value if w(t;t') is everywhere zero, except for t=t', when it must be one. In this case,

$$I_{max} = \int_d^{\infty} Exp(-\rho(\alpha^1-\phi(t')\theta_1))r(t')dt'.$$

Now the smallest value of the exponential is given when, w(t;t') is one everywhere. In this case,

$$I_{min} = \int_d^{\infty} Exp(-\int_0^d \rho(\alpha^1-\phi(y))\theta dy)r(t')dt'.$$

However, the smallest value of $\phi(y)$ is α^0. Thus the above integral is minimised with respect to variation in $\phi(y)$ if $\phi(y)$ is set equal to α^0.

In this case,

$$I_{min} = (1-R(d))Exp(-\rho\theta_1(\alpha^1-\alpha^0)d).$$

Putting together the various terms, it can be seen that the lower bound of the LHS of equation 4.2.b is given by,

$$L_{min} = (1-R(d))Exp(-\rho\theta_1(\alpha^1-\alpha^0)d).$$

Considering now the right hand side of equation 4.2.b, by analogy with the previous paragraph, the maximum value of

$$\int_0^d Exp(-\int_0^d B_0(y;t')dy)r(t')dt',$$

where $B_0(t:t')$ is obtained from $B_1(t;t')$ by replacing all the '1' indices by the index '0'. Now, as before,

$$I_{max} = \int_0^d Exp(-\int_0^d \rho(\alpha^0-\phi(y))\theta_0 dy)r(t')dt'.$$

Put $\phi(y)=\alpha^1$, as this maximises this integral with respect to variation in $\phi(y)$, then,

$$I_{max} = R(d)Exp(-\rho\theta(\alpha^0-\alpha^1)d).$$

Thus the maximum value of the RHS of equation 4.2.b is given by,

$$R_{max} = (1-R(d))Exp(-\rho\theta(\alpha^0-\alpha^1)d).$$

Inequality 4.2.b holds under the condition that,

$$L_{min} > R_{max},$$

I.e.,

$$\mathrm{Exp}(-\rho\theta_1(\alpha^1-\alpha^0)d) > \mathrm{Exp}(-\rho\theta_0(\alpha^0-\alpha^1)d)$$

$$(\rho\theta_1(\alpha^1-\alpha^0)d) < (\rho\theta_0(\alpha^0-\alpha^1)d)$$

$$\underline{\theta_1 < -\theta_0}$$

Since both θ_1 and θ_0 are always positive, then the opposite of this inequality holds. I.e. the model predicts the opposite of the Farmer and Schoenfeld results. It predicts that reinforced responses are more likely to follow non-reinforced responses, than reinforced responses. The model makes this prediction because the general effect of reinforcement is to shorten following IRT's (but short IRT's are not reinforced) and the general effect of non-reinforcement is to lengthen following IRT's (making them more likely to be reinforced).

Thus although the model makes reasonable predictions about the asymptotic IRT distributions, its predictions about conditional statistics do not seem to agree with previous experimental findings.

4.3 The Experiment

To test more closely the predictions of the model, subjects were run on a DRL schedule. Four human subjects were used. They were all undergraduates at the University of Stirling and they were paid at the rate of 30p per session. The experimental situation was exactly the same as described in section 3.3. Only the reinforcement schedule was changed. A trial run under DRL5 proved to be too easy, and the two test subjects refused to come for further sessions. The schedule was thus changed to DRL5;0.5 (a post hoc analysis shows p to have been 0.48). Two new subjects were found, and the four subjects run for 10 sessions each.

4.4 The Results

Figure 4.4.1 gives the session by session results from subjects 9-12. These results can be classified into two groups, a first group containing subjects 9-11, and subject 12. Subjects 9-11 all show similar characteristics. An initially high rate of responding (session 0, which is the first five minutes of session 1) rapidly switches (session 1) to a very low rate of responding. Some subjects e.g. subject 10, virtually stop responding at this point. The next two sessions show quite marked speeding up of responding until the majority of IRT's lie between 5 and 10 seconds duration. This is then followed by a gradual sharpening of the IRT distribution until the vast majority of IRT's are around 5 seconds duration.

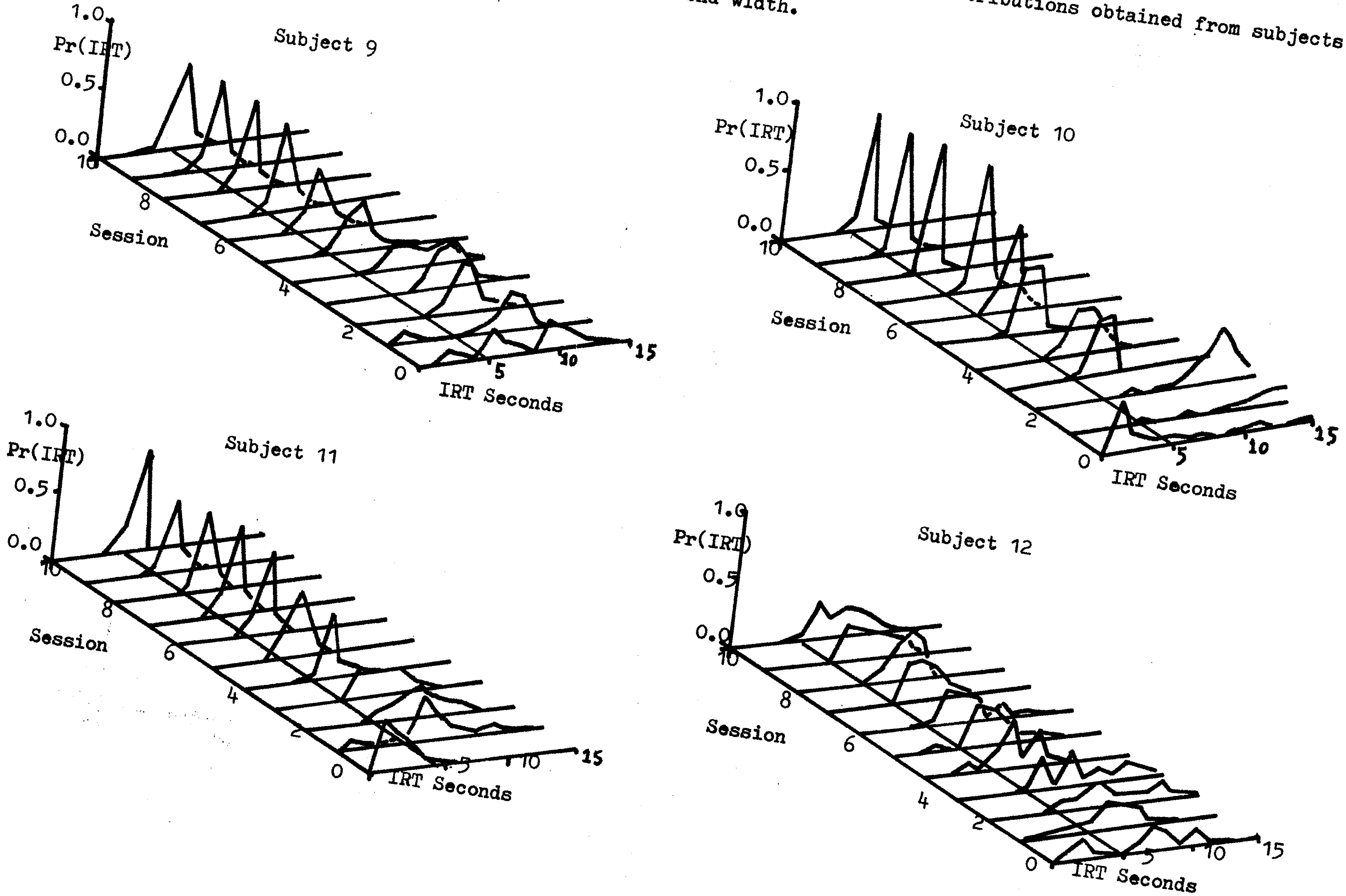
It is interesting to note, that none of subjects 9-11 have any responses within the first three seconds of the IRT distribution, and show no sign of a peak close to zero that has been characteristic of many animal experiments. All the subjects became aware that some kind of time constraint was operating, in the sense that very rapid responding was 'no good', but none of them ever characterised it exactly. This belief (in the uselessness of rapid responding) does however seem to have removed completely all the very short IRT's from these subjects' performances.

Subject 12 distinguishes himself from the others by his poor level of peformance, in terms of the obtained rate of reinforcement. There is in his results, a definite trend in the direction of responding at around 5 second intervals, but by session 10, his performance is still not very good. Whether or not the subject would have improved with more training is impossible to say. It does not seem unreasonable to assume this however, as the results do seem to be moving in the appropriate direction, and it may not be unfair to describe this subject's data as non-asymptotic.

Subject 12 also differed from subjects 9-11 in that he did show evidence of a peak at short IRT's. The graph scale (figure 4.4.1) is too small to show it. In the interval 0.0-0.5, the probability of an IRT was, for session 10, 0.002, in interval 0.5 -1.0 seconds it was 0.003, and in interval 1.0-1.5 seconds, it was 0.002, making approximately ¾% of responses in this region. Although the total effect was small, it is of interest that this behaviour should be shown by the worst performer on this schedule.

The data from sessions 9 and 10 was pooled to provide a substantial number of responses, and this data was regarded as

Figure 4.4.1 -- DRL schedule. d = 5.0, p = 0.5. The session by session IRT distributions obtained from subjects 9-12. The IRT's were grouped into intervals of one second width.



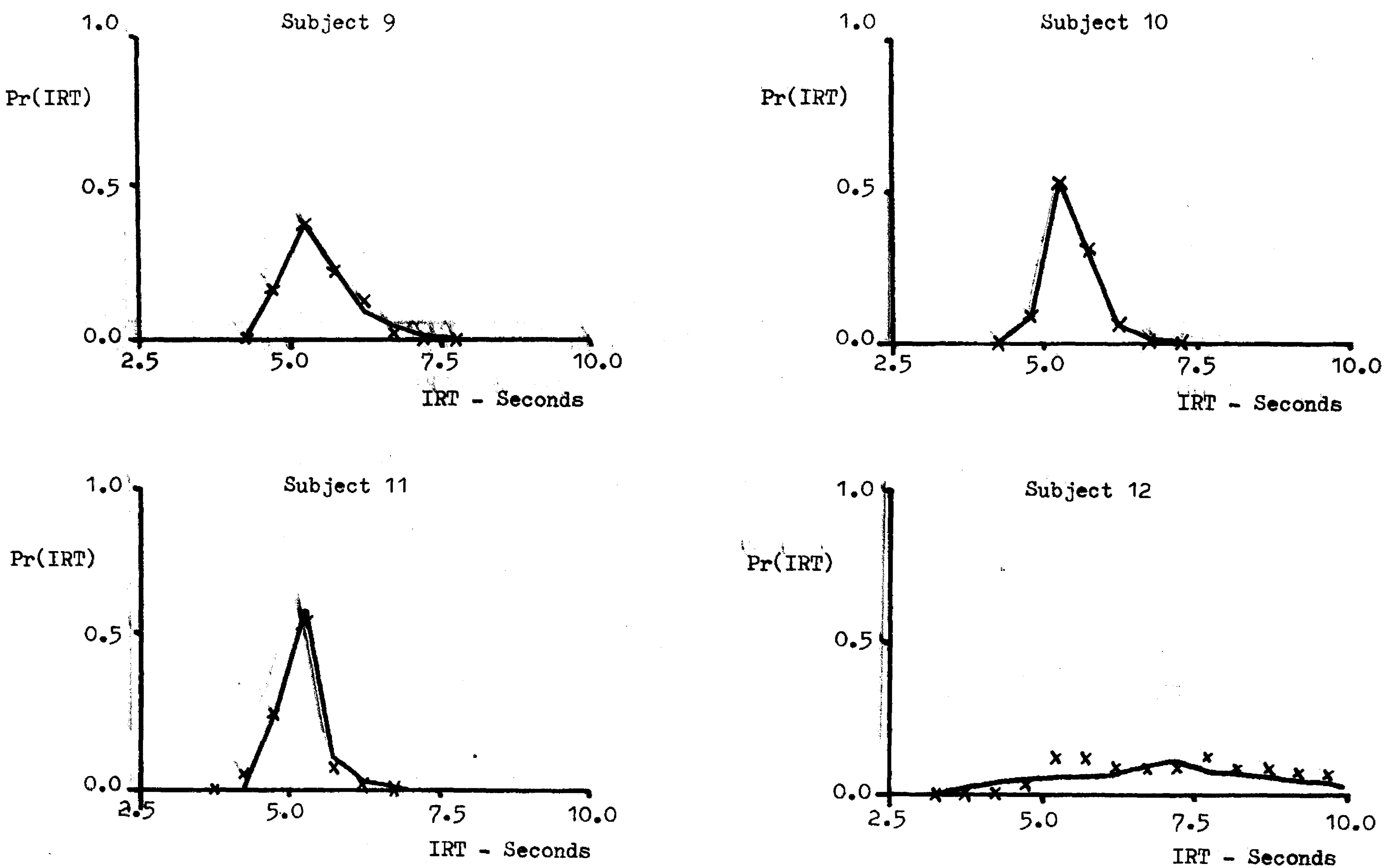


Figure 4.4.2 -- DRL schedule. d = 5.0 p = 0.48. A comparison of the data and predictions for the asymptotic IRT distribution subjects 9-12. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one half second width.

Subject	$\hat{\rho}$	$\hat{\Theta}$	$\hat{a}$	LSE.10^6	No. Data Points
9	1.65	0.38	0.31	1509	9
10	3.07	2.81	0.28	52	8
11	3.31	0.00	0.11	4591	8
12	0.41	1.13	1.72	15445	14

Table 4.4.(i) -- Estimates of the parameters ρ, Θ, and a for subjects 9-12, together with a list of the least square errors (LSE) between the obtained and predicted values of the asymptotic IRT distribution for these subjects. The last column gives the number of data points used in the parameter estimation. The IRT's were classified into half second intervals.

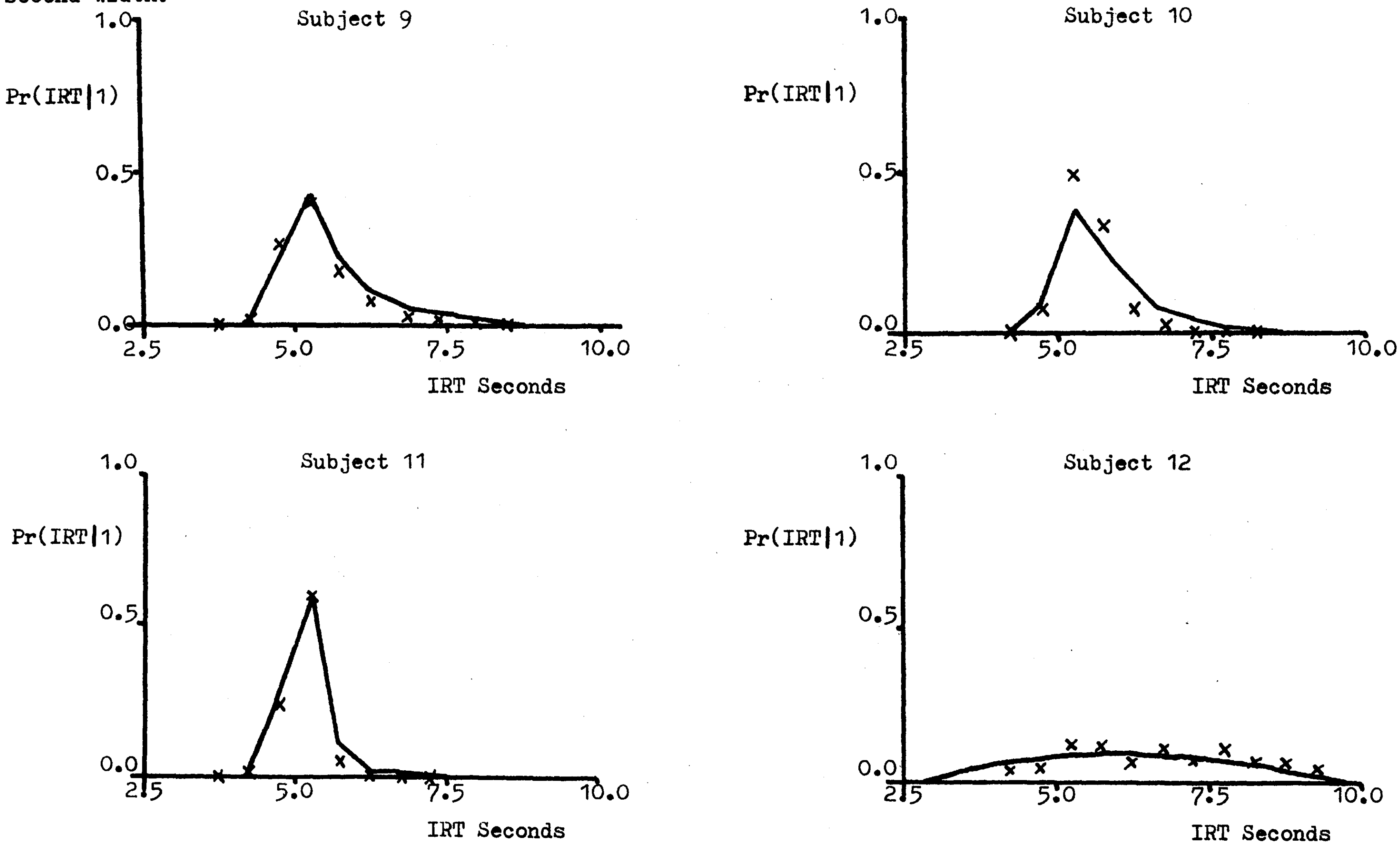
---

asymptotic and used for parameter estimation. There are three parameters to estimate for the IRT distribution, these being, ρ, Θ, and a. (In view of the failure to find a substantial peak near to t=0.0, the parameters α^0 and α^1 were set equal to 0.0 and 1.0 respectively. This simplifies considerably the search programme required to estimate ρ, Θ, and a, and reduces substantially the computing time required.) A minumum search technique was used to estimate the parameters, the criterion being the least squares fit. Table 4.4.(i) gives a list of the parameter values obtained, and the least square error.

Subject	$\hat{\Theta}_1$	LSE.10^6	No. data Points
9	1.00	5662	10
10	0.35	30850	9
11	--	5486	8
12	0.45	8050	12

Table 4.4.(ii) -- Estimates of Θ_1 for subjects 9-12. The LSE column gives the least square error between obtained and predicted values of the IRT distribution conditional on the previous response being reinforced for these subjects. The last column gives the number of data points used in the estimation of the parameter Θ_1.

Figure 4.4.3 -- DRL schedule. d = 5.0, p = 0.5. A comparison of data and predictions for the asymptotic IRT distribution conditional on the previous response being reinforced for subjects 9-12. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one half second width.



The fits for subjects 9-11 are quite good, while that for subject 12 is at least an order of magnitude worse. This poor fit to the data from subject 12 is, in a way, reassuring, for a good fit here would tend to imply that the model was so flexible that it would fit almost anything. This is not an especially desirable property for any model, as it would make it impossible to put the model to any real test of its validity. Figure 4.4.2 gives a graphical comparison of the data and the predicted values based on a least squares fit. These graphs confirm the impression obtained from table 4.4.(i), that the fits are good for subjects 9-11, but bad for subject 12.

The results of an attempted fit to the IRT distribution conditional on the previous response being reinforced are given in Figure 4.4.3. The fits, as expressed by the least square error (Table 4.4.(ii)) are quite reasonable, with the exception of subject 10. Comparison of table 4.4.(ii) with table 4.4.(i) will show that the best fit of subject 10 to the asymptotic IRT distribution becomes the worst fit to the conditional IRT distribution. It would be possible to obtain a better fit for subject 10 if the constraint applied to the value of Θ_1 by the value of Θ were ignored. (The value of Θ is the one obtained from the asymptotic IRT distribution.) This constraint is,

$$0.0 < \Theta_1 \leq \frac{1}{\Theta},$$

for,

$$\Theta = \frac{\Theta_0}{\Theta_1}$$

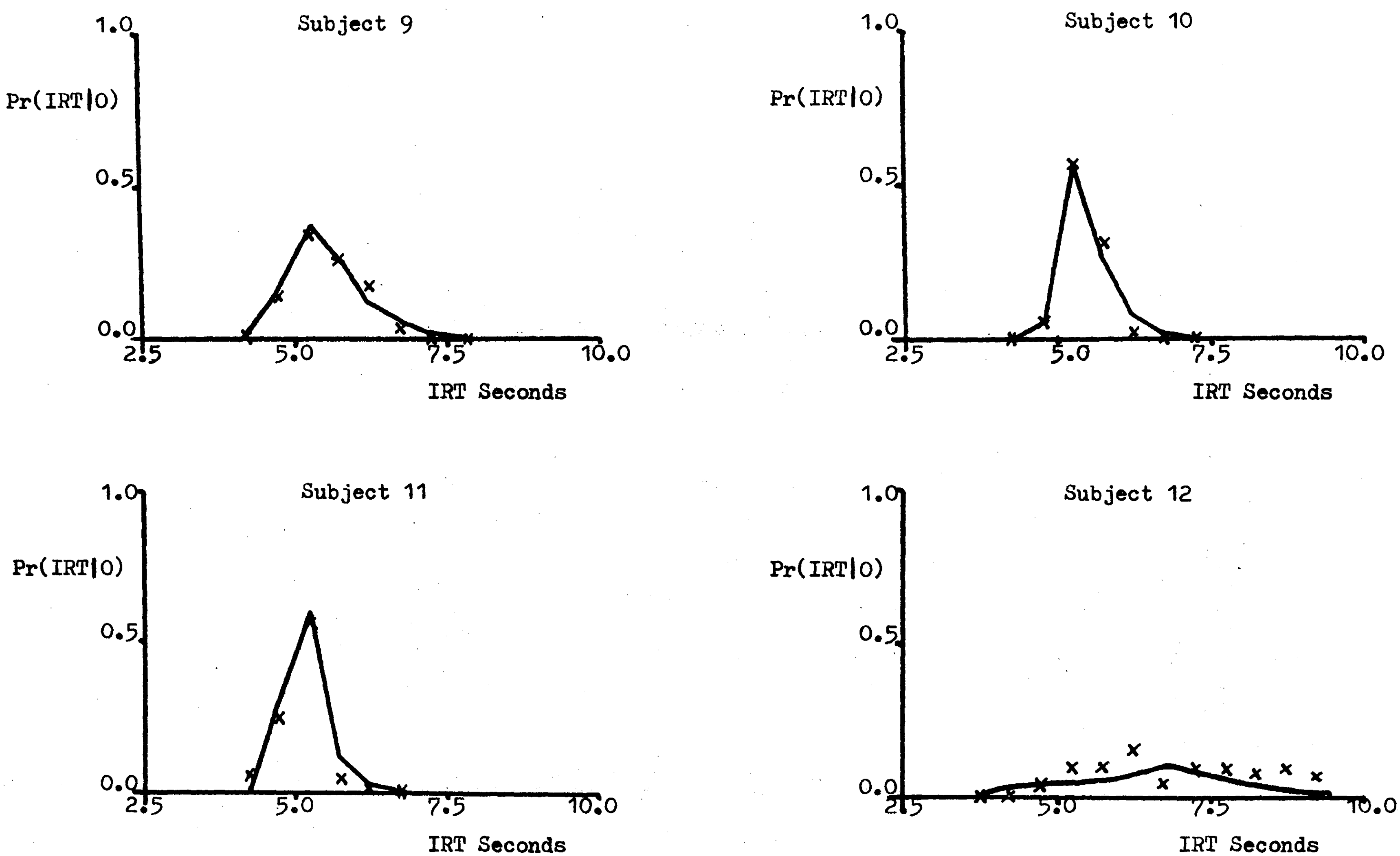
and,

$$0.0 \leq \Theta_0 \leq 1.0.$$

For subject 10, the value of Θ_1 found by the method of least squares is $1/\Theta$. If Θ_1 were allowed outside this range, a better fit could be found. The problem that then arises is one of what Θ_1 and the corresponding Θ_0 ($= \Theta.\Theta_1$) would represent.

In contrast to the very bad fit to the data of subject 10, the IRT distribution conditional on the previous response being reinforced from subject 12's data shows a much improved fit

Figure 4.4.4 -- DRL Schedule. d = 5.0, p = 0.5. A comparison of data and predictions for the asymptotic IRT distribution conditional on the previous response not being reinforced for subjects 9-12. Solid lines join the predicted points. Crosses mark obtained values. IRT's are grouped in intervals of one half second width.



over that for the corresponding asymptotic IRT distribution. This appears to be due chiefly to the bimodality of subject 12's asymptotic IRT being somewhat less pronounced in the case of the conditional IRT distribution.

It will have been observed that there is no entry for Θ_1 for subject 11. This is because, in the asymptotic IRT distribution, $\Theta = 0.0$. This result implies that $\Theta_0 = 0.0$ and that, $\phi(t) = \alpha^1$, for $t \geq d-a$. Since the conditional IRT distribution requires a term of the form $(\alpha^1 - \phi(t))\Theta_1$, which is obviously zero in this case, for any value of Θ_1, then a value for Θ_1 cannot be extracted. The asymptotic conditional IRT should be the same as the asymptotic IRT. This is used to generate the entry in the least squares column of table 4.4.(ii).

As $\Theta_0 = \Theta.\Theta_1$, it is possible to use the data from tables 4.4.(i) and 4.4.(ii) to calculate Θ_0 for each subject. This was done, and the results appear in table 4.4.(iii).

Subject	$\hat{\Theta}_0$	$SE.10^6$	No. Data Points
9	0.38	4432	8
10	1.00	1765	8
11	0.00	9433	7
12	0.51	24651	13

Table 4.4.(iii) -- Estimates of Θ_0 for subjects 9-12. SE column gives the square error between the obtained and predicted values of the IRT distribution conditional on the previous response not being reinforced. The last column gives the number of data points used to calculate the SE.

Figure 4.4.4 illustrates the fit of the model to data for the IRT distribution conditional on the previous response not being reinforced. No parameters are estimated from this data. The fits are on the whole quite good, with the not unexpected exception of the data from subject 12.

4.5 Conclusion

The results obtained from the analysis of the DRL schedule are on the whole quite reassuring. If subject 12 is ignored, on the grounds that these results do not represent asymptotic performance, then the model gives good fits to the asymptotic IRT distribution, and moderate fits to the asymptotic IRT distributions conditional on either reinforcement or non-reinforcement.

The actual values obtained for the various parameters show quite a wide range from subject to subject and reveal no particular pattern that might imply any special significance for particular values.

Chapter 5

THE RANDOM INTERVAL SCHEDULE

5.1 The Experimental Background

A random interval schedule is one where the reinforcement function is defined by,

$$u(t) = 1 - \mathrm{Exp}(-\gamma t),$$

where γ is a constant. Most variable interval schedules can be regarded as approximations to this schedule. Although a large amount of research has been done on interval schedules, relatively little has been done on random interval schedules as such. Most of the properties of interval schedules that are usually investigated turn out to be rather intractable from a mathematical point of view. There are however a number of general discussions of the nature of the IRT distribution, e.g. Anger (1956), Catania and Reynolds (1968), Farmer (1963), and Kintsch (1965).

The simplest characteristic of the IRT distribution obtained from interval schedules is given by Farmer (1963), who states that the IRT distribution is unimodal, and that the mode tends to move towards longer IRT values, as γ decreases. (I.e. lowering the mean rate of reinforcement lowers the rate of responding.) Much of the work done, e.g. by Catania and Reynolds, concerns such relations as those between rate of responding and time-since-reinforcement. Unfortunately, while it is possible to set up expressions with time-since-reinforcement as the temporal variable, the expressions are so cumbersome and complex that they have not yet yielded to manipulation. The conclusions that can be drawn from the model that relate to present knowledge of properties of behaviour under interval schedules are thus rather few. However the next section discusses some properties of the model in the random interval context, and is followed by a look at the fit of the model to some actual data.

5.2 Predictions from the Model

From equation 2.6.a, the asymptotic response strength is given by,

$$\phi(t) = \frac{\alpha^0 \theta \int_{t-a}^{t+a} \mathrm{Exp}(-\gamma y) r(y) dy + \alpha^1 \int_{t-a}^{t+a} (1-\mathrm{Exp}(-\gamma y)) r(y) dy}{\theta \int_{t-a}^{t+a} \mathrm{Exp}(-\gamma y) r(y) dy + \int_{t-a}^{t+a} (1-\mathrm{Exp}(-\gamma y)) r(y) dy}$$

Note first that if $\theta = 0.0$, (non-reinforcement has no effect), then, $\phi(t) = \alpha^1$. I.e. $\phi(t)$ is a constant and independent of γ. In this case, $r(t)$ is simply exponential.

To investigate $\phi(t)$ and $r(t)$ further, it is necessary to make some assumptions and approximations. If it is assumed that a is small, then,

$$\phi(t) \simeq \frac{\alpha^0 \theta \mathrm{Exp}(-\gamma t) + \alpha^1 (1-\mathrm{Exp}(-\gamma t))}{\theta \mathrm{Exp}(-\gamma t) + (1-\mathrm{Exp}(-\gamma t))} .$$

For simplicity, let, $\alpha^0 = 0.0$ and $\alpha^1 = 1.0$. Then, if γ is small, expanding the exponential, gives,

$$\mathrm{Exp}(-\gamma t) \simeq 1 - \gamma t,$$

or,

$$u(t) \simeq \gamma t.$$

Hence,

$$\phi(t) \simeq \frac{\gamma t}{\theta(1-\gamma t) + \gamma t}$$

$$= \frac{\gamma t}{\theta + (1-\theta)\gamma t}$$

$$\simeq \underline{\frac{\gamma t}{\theta}} .$$

This approximation for $\phi(t)$ will be used in the rest of this section.

The mode of the IRT distribution is found by setting $\frac{dr}{dt} = 0.0$, and solving for t.

$$r(t) = \rho\phi(t)\mathrm{Exp}(-\int_0^t \rho\phi(y)dy).$$

Therefore,

$$\frac{dr}{dt} = \rho\frac{d\phi}{dt}(t)\mathrm{Exp}(-\int_0^t \rho\phi(y)dy)$$
$$- \rho^2\phi(t)^2\mathrm{Exp}(-\int_0^t \rho\phi(y)dy).$$

The mode is thus given by the solution of,

$$\frac{d\phi}{dt}(t) - \rho\phi(t)^2 = 0.0.$$

Using the approximate value for $\phi(t)$, this gives,

$$\frac{\gamma}{\theta} - \rho\frac{\gamma^2 t^2}{\theta^2} = 0.0,$$

and therefore,

$$\underline{t = \sqrt{\frac{\theta}{\rho\gamma}}}$$

Thus as γ decreases, t increases. I.e. this is in agreement with Farmer's 1963 results.

The relation between the rate of responding and the rate of reinforcement has been discussed by a number of researchers, and particularly by Herrnstein (1970).

Herrnstein was considering initially the law of effect, as exemplified by choice behaviour in a two response concurrent schedule. (A concurrent schedule is simply one with two or more responses available, each response being independently reinforced. Ferster and Skinner (1957, p724) define concurrent operants as, "Two or more responses, of different topography at least with respect to locus, capable of being excecuted with little mutual interference, at the same time or in rapid alternation, under the control of separate programming devices.") If the two suffixes L and R are used to differentiate the two responses, (Left and Right) then matching is said to take place if,

$$\frac{P_L}{P_R} = \frac{R_L}{R_R}$$

or,

$$\frac{P_L}{P_L + P_R} = \frac{R_L}{R_L + R_R},$$

where P denotes the rate of responding (for Pecks, pigeons were usually the subjects in the experiments discussed by Herrnstein) and R is the rate of reinforcement.

The above relationships can be discussed in the context of a single response, by allowing one response (L) to be the response under investigation, and the other response (R) to be the set of all other possible responses, excluding L. In this simple situation, the law becomes,

$$P = kR. \qquad 5.2.a$$

(k is some constant, and the redundant suffix L has been dropped.)

As long as the total number of reinforcements $(R_L + R_R)$ is constant, there is no way to distinguish, in a single experiment,
As long as the total number of reinforcements $(R_L + R_R)$ is

$$P_L = \frac{kR_L}{R_L + R_R}. \qquad 5.2.b$$

When the sum $(R_L + R_R)$ is a constant, the only difference between equation 5.2.a and equation 5.2.b is in the value of k, which is an arbitrary constant in either case. The two equations make divergent predictions only when $(R_L + R_R)$ is not a constant, which is actually more typical. The difference is that equation 5.2.b predicts a "contrast effect", between responses, rather than a strict independence effect, as in equation 5.2.a. Generally experiments support equation 5.2.b, for contrast effects are often found, especially in concurrent procedures. (Catania, 1966).

If the suffixes are dropped in equation 5.2.b, then the matching law takes the form,

$$P = \frac{kR}{R + A},$$

where k and A are constants.

Now Herrnstein also points out that an alternative matching law is the probability matching law, in contrast to the number matching law given above. In a probability matching law, the relative rates of responding equals the relative probability of reinforcement, rather than the relative rates of reinforcement. I.e.,

$$\frac{P_L}{P_R} = \frac{\frac{R_L}{P_L}}{\frac{R_R}{P_R}}$$

or,

$$\frac{P_L}{P_R + P_L} = \frac{R_L}{\sqrt{R_L} + \sqrt{R_R}} \,.$$

If the same arguments are applied to the probability matching law, as were applied to the numbers matching law, above, then in the single response case, the law takes the form,

$$P = \frac{k\sqrt{R}}{\sqrt{R} + A},$$

where k and A are again constants, though not necessarily the same as before.

The strong attachment to rates of responding found amongst operant researchers conceals the simplicity of these relationships. If they are transformed into statements about mean interresponse times,

$$m_P = \frac{1}{P},$$

and mean interreinforcement times,

$$m_R = \frac{1}{R},$$

then they take extremely simple forms. These are,

Number matching law,

$$m_P = bm_R + c,$$

Probability matching law,

$$m_P = b\sqrt{m_R} + c,$$

where b and c are constants, though not necessarily the same in the two laws.

The point of concern here is whether or not the model used to describe random interval schedules favours either of these laws, rather than any other. To do that it is necessary to calculate values for m_P and m_R in terms of the model's parameters.

Now,

$$m_P = \int_0^\infty t r(t)dt$$

$$\simeq \int_0^\infty t\, \frac{\rho\gamma t}{\theta} \mathrm{Exp}\left(-\frac{\rho\gamma t^2}{2\theta}\right)dt.$$

Put,

$$x = \frac{\rho\gamma t^2}{2\theta},$$

therefore,

$$dt = \sqrt{\frac{\theta}{2\rho\gamma}}\, x^{-\frac{1}{2}}dx.$$

Thus,

$$m_P \simeq \sqrt{\frac{2\theta}{\rho\gamma}} \int_0^\infty x^{+\frac{1}{2}} \mathrm{Exp}(-x)dx,$$

but,

$$\int_0 x^{+\frac{1}{2}} \mathrm{Exp}(-x)dx = \Gamma(\tfrac{3}{2}) = \frac{\sqrt{\pi}}{2},$$

so that,

$$\underline{m_P \simeq \sqrt{\frac{\theta\pi}{2\rho\gamma}}}$$

This is thus the approximate value of the mean interresponse time.

m_R is somewhat more difficult to calculate, as it is not possible to calculate directly the distribution of interreinforcement times. However, if m_Q represents the mean probability that a response is reinforced, then,

$$m_R = \frac{m_P}{m_Q} .$$

m_Q is given by,

$$m_Q = \int_0^{\infty} u(t)r(t)dt$$

$$\simeq \int_0^{\infty} \gamma t \, \frac{\rho\gamma t}{\theta} \mathrm{Exp}\left(-\frac{\rho\gamma t^2}{2\theta}\right)dt$$

$$= \int_0^{\infty} \frac{\rho\gamma^2 t^2}{\theta} \mathrm{Exp}\left(-\frac{\rho\gamma t^2}{2\theta}\right)dt .$$

Put,

$$x = \frac{\rho\gamma t^2}{2\theta} ,$$

therefore,

$$dt = \sqrt{\frac{\theta}{2\rho\gamma}} \, x^{-\frac{1}{2}} ,$$

and hence,

$$m_Q \simeq \sqrt{\frac{2\theta\gamma}{\rho}} \int_0^{\infty} x^{\frac{1}{2}} \mathrm{Exp}(-x)dx .$$

Thus,

$$m_Q \simeq \sqrt{\frac{\theta\pi\gamma}{2\rho}}$$

Therefore,

$$m_R \simeq \frac{\sqrt{\frac{\pi\theta}{2\rho\gamma}}}{\sqrt{\frac{\theta\pi\gamma}{2\rho}}}$$

$$= \frac{1}{\gamma}$$

This is the result to be expected, for if the reinforcements had been scheduled as in an ordinary interval schedule, the intervals between the scheduled reinforcements would have been distributed according to $\gamma \mathrm{Exp}(-\gamma T)$. Scheduled reinforcements are usually collected shortly after scheduling, thus the mean interreinforcement time is expected to be about $1/\gamma$.

Using the above expressions for m_P and m_R,

$$m_P = \sqrt{\frac{\theta \pi m_R}{\rho}}$$

or,

$$m_P = b\sqrt{m_R}.$$

Since in practice m_P will contain components not accounted for in the model (e.g. a response speed component) this expression may be modified to read,

$$m_P = b\sqrt{m_R} + c.$$

Thus the model suggests that a probability matching law is more likely to fit data on the relationship between m_P and m_R, than is a number matching law. Herrnstein (1970) is of the opinion that the number matching law in fact gives the best representation of the data he possesses. However, he fails to make any attempts to fit the probability matching law to his data, being content to show pictorially that the number matching law gives a tolerable fit. (No measures of fit are given.)

The probability matching prediction is not only characteristic of the random interval schedule, under this model, but it is also (again approximately) characteristic of the random ratio schedule.

In the random ratio schedule, it is obvious that,

$$m_R = \frac{m_P}{p},$$

where $p = u(t)$, the reinforcement function.

Now,

$$m_P = \frac{1}{\rho k},$$

(See chapter 3)

and, taking $\alpha^0 = 0.0$ and $\alpha^1 = 1.\theta$, in the expression for k,

$$m_P = \frac{\theta(1-p) + p}{\rho p}$$

but,

$$p = \frac{m_P}{m_R},$$

therefore,

$$\rho m_P = \theta(\frac{m_R}{m_P} - 1) + 1.$$

Hence,

$$\rho m_P^2 = \theta m_R - \theta m_P + m_P$$

$$0.0 = \rho m_P^2 - (1-\theta)m_P - \theta m_R.$$

Thus,

$$m_P = \frac{(1-\theta) + \sqrt{(1-\theta)^2 + 4\rho\theta m_R}}{2\rho}$$

$$= \frac{(1-\theta)}{2\rho}\, 1 - \sqrt{(1 + \frac{4\rho\theta m_R}{(1-\theta)^2})}$$

Now if, $\frac{4\rho\theta\, m_R}{(1-\theta)^2}$ is large compared with 1.0, this becomes, approximately,

$$m_P \approx \sqrt{\frac{\theta}{\rho} m_R} + \frac{(1-\theta)}{2\rho},$$

or,

$$\underline{m_P \simeq b\sqrt{m_R} + c} \quad .$$

This relationship between m_P and m_R thus seems to partly characterise the model rather than the particular schedule, provided always, that $\theta \neq 0.0$.

It did not prove possible (see results, section 5.4) to investigate this relationship experimentally, as only one value of

γ was used in the experiments described in the following section. In many ways this is a pity. Perhaps this interesting point can be pursued experimentally at a later date. If the probability matching law is found to hold for a given subject, across changes in the schedule parameter, then this implies the following:-

The model parameters, (in this case θ and ρ) must be constant across changes in the schedule parameters. I.e. they must be psychological invariants that in some way partially characterise the subject.

Such real invariants have proved notably elusive in psychology.

Sequential statistics are extremely difficult to handle analytically, as the small a approximation for $\phi(t)$ can no longer be used. The programme SACONDAR (Appendix I) is available for numerical investigation (e.g. see figure 2.7.1) of the IRT distribution conditional on either the previous response being reinforced, or not reinforced.

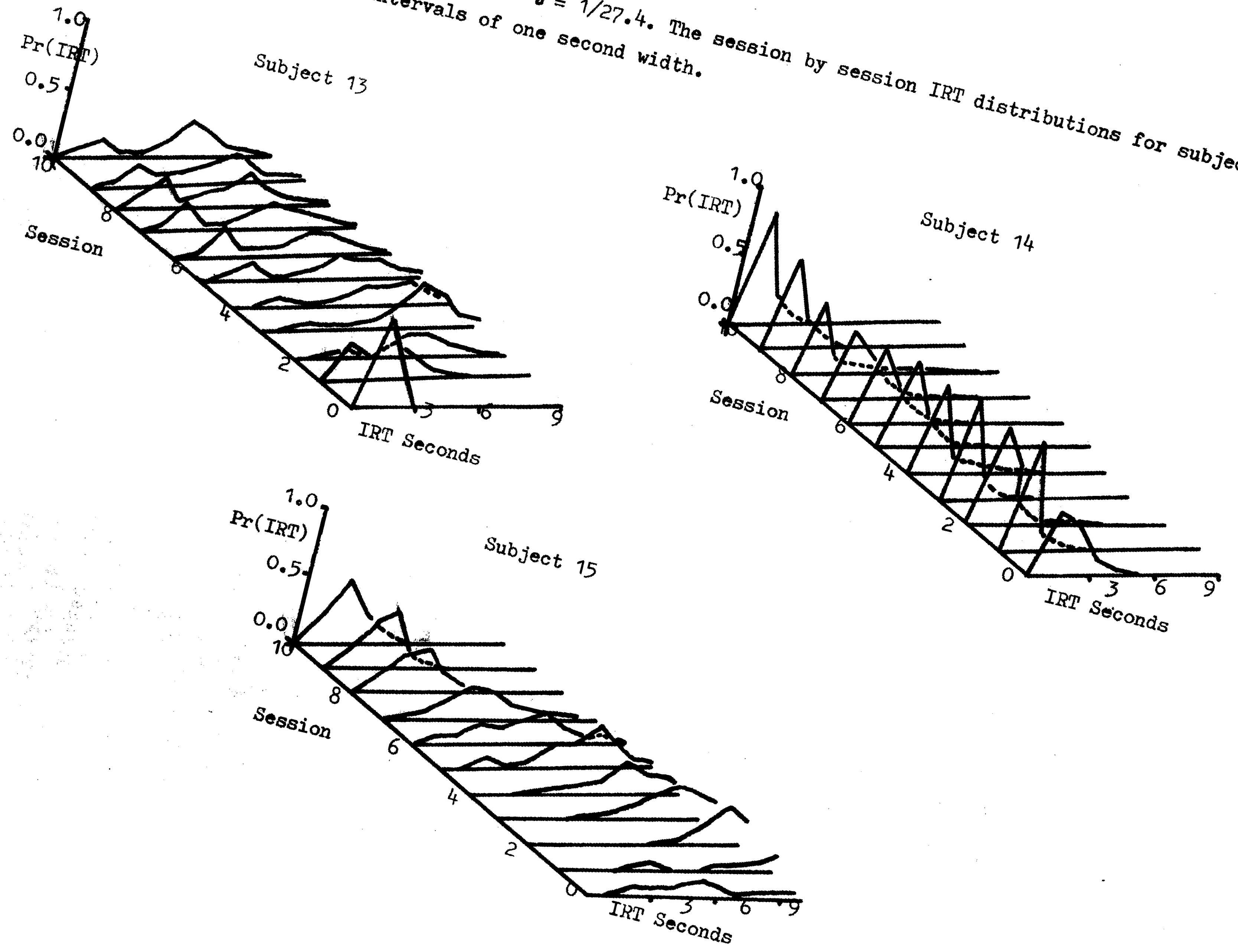
5.3 The Experiment

The same experimental apparatus as described in section 3.3 was used. Only the reinforcement schedule was changed, so that a random interval schedule was in operation. The values of the components in the schedule control were chosen so that $1/\gamma$ was approximately 30 seconds. An analysis of the distribution of reinforcements actually produced by the schedule established that $1/\gamma$ was best described by 27.40 seconds, and this value was used in all calculations. Four human subjects were used. As previously, they were paid at the rate of 30p per session. Subjects came daily for 10 sessions, excluding weekends. One subject dropped out after the first three sessions and his results are not included in the following analysis.

5.4 The Results

Figure 5.4.1 gives the session by session results for subjects 13-15. The IRT's were analysed by grouping into intervals of one second width. The distinguishing feature of these results appears to be their variety between subjects. Subject 13 begins with a

Figure 5.4.1 -- Random interval schedule. $\gamma = 1/27.4$. The session by session IRT distributions for subjects 13-15. The IRT's were grouped into intervals of one second width.



rapid rate of responding and slows down gradually over sessions 1-4. This is then followed by a slow increase in the rate of responding with the IRT distribution developing a marked bimodality. Over sessions 7-10, the peak associated with longer IRT's moves slowly toward that for short IRT's, which peak remains relatively fixed in position, but increases in magnitude. This suggests that further sessions may have resulted in the merging of the two peaks to produce a unimodal distribution.

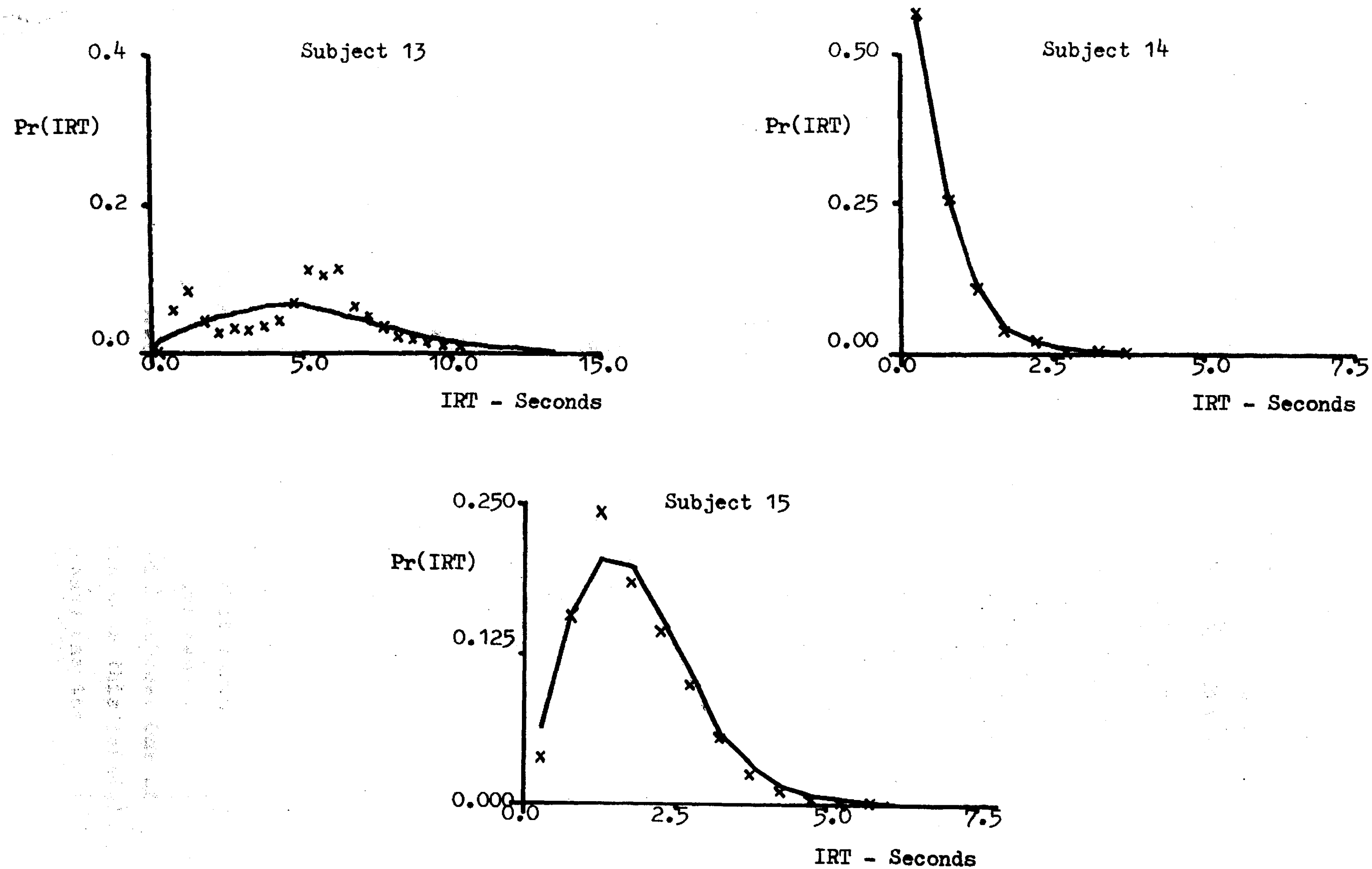
In contrast to this, subject 14 always has a relatively high rate of responding. However, as session number increases, the number of short IRT's decreases and the IRT distribution develops a long tail. This process continues upto session 9, when there is a sudden return to the very high initial rate of responding, which is then exceeded in session 10.

Subject 15 is different again, though in some ways this subject is similar to subject 13. The first session is characterised by a very low rate of responding - almost no responding. The IRT distrib -ution then begins to shift its peak from long IRT's to short IRT's, so that by session 10, the mode is in the region of 1.5 seconds. This trend is quicker and more uniform than the corresponding process for subject 13, and only a slight trace of bimodality (session 5) is to be found.

The results from session 10, for each subject, were taken as asymptotic data and re-analysed into half second intervals. The Parameters α^0 and α^1 were set at 0.0 and 1.0 respectively, to reduce the number of parameters to be estimated to a minumum, and the values of ρ, θ, and a were estimated, to two places of decimals, by the method of least squares. The programme SAGENAR (Appendix I) was used to generate the IRT distribution, rather than use the approximations developed in the previous section. Table 5.4. (i) gives the values of these parameters, for each subject, together with the least squares error.

The parameters vary widely between subjects. As is perhaps expected, θ is large for subject 13, i.e. this subject is heavily affected by non-reinforcement, and hence has a low rate of responding, while θ is small for subject 14, i.e. this subject is not much affected by non-reinforcement and so has a high rate of responding. Subject 15 lies between these two extremes. The fits

Figure 5.4.2 -- Random interval schedule, $\gamma = 1/27.4$. A comparison of the data and predictions for the asyptotic IRT distribution for subjects 13-15. Solid lines join the points giving the predicted values. Crosses mark the obtained data values. IRT's are grouped in intervals of one half second width.



obtained also range widely, being very good for subject 14, moderate for subject 15 and bad for subject 13, although 13 could be perhaps explained away by allowing that this subject's data is not really asymptotic.

Subject	$\hat{\rho}$	$\hat{\theta}$	$\hat{a}$	LSE.10^6	No. Data Points
13	5.29	4.09	1.07	16259	20
14	8.74	0.10	4.11	150	9
15	17.50	1.29	0.00	2322	13

Table 5.4.(i) -- Estimates of the parameters ρ, θ, and a for subjects 13-15, together with a list of the least square errors (LSE) between the obtained and predicted values of the asymptotic IRT distribution for these subjects. The last column gives the number of data points used in the parameter estimation.

Figure 5.4.2 shows graphs of the data for session 10 for each subject, compared with the predicted values. The reasonable results for subjects 14 and 15 can be seen. It is obvious that the poor fit to the data from subject 13 results chiefly from the bimodality of this subject's results.

No predictions or fits are made to the asymptotic conditional data, as sufficient time was not available to run a parameter search programme.

5.5 Conclusions

It is unfortunate that it was not possible to collect sufficient data to investigate the relationship between the mean rate of reinforcement and the mean rate of responding, for individual subjects. The restrictions were chiefly practical. The rather tedious nature of the task led subjects to decline the offer of further sets of 10 sessions, and, in any case, there would not have been sufficient time available to cover a whole range of γ values. One of the characteristic difficulties subjects associated with this schedule (they were not told before hand what the schedule was) was the relative insensitivity of the rate of reinforcement to their patterns of responding. The subjects did not themselves feel very successful and after a few sessions just wanted to get the whole lot finished off as quickly as possible. Perhaps it will be possible however to test the model against some animal data at some future date.

For the data obtained from a random interval schedule, the model seems to have done fairly well. It does not seem unreasonable to view the fit of the model to the data from subject 13 as partly a reflection of the non-asymptotic aspects of this subject's data. This then leaves the two reasonably good fits of the model to data from subjects 14 and 15 to characterise the success of the model in accounting for the general properties of the IRT distribution produced by random interval schedules of reinforcement.

Chapter 6

CONCLUSION

6.1 Final Remarks

The preceeding three chapters contain a fairly detailed attempt to describe the commonest types of operant reinforcement schedules by a single mathematical model. Ratio, interval and DRL schedules have all been described as special examples of schedules that prescribe reinforcement as a function of interresponse time.

In describing the relatively wide varieties of behaviour these schedules produce, only three parameters have been really necessary and each of these parameters characterise a particular aspect of the model:-

i) The response process -- ρ, the sampling parameter.
ii) The stimulus process -- a, the spread parameter.
iii) The learning process -- θ, the relative effectiveness of reinforcement and non-reinforcement parameter.

This appears to be the minumum number of parameters that could possibly be used, in any moderately complex model.

The predictions of the model are generally in accord with the known experimental results, with one exception. This exception is that result reported by Farmer and Schoenfeld (1964), that under DRL schedules, reinforced responses are more likely to follow a reinforced response than to follow a non-reinforced response. This model predicts the opposite, a prediction more in accord with the generally accepted belief that one effect of reinforcement is to speed up responding.

In most cases the fit of the model to the asymptotic IRT distribution were good. In the random ratio case however, it was found that the most important component of the IRT was the time required to produce the response. The DRL and the random interval results each produced one example of a poor fit to the asymptotic IRT distribution, but they seem to have been due to the non-asymptotic qualities of these particular sets of data. Otherwise the fit of the model to the data from these two schedules was good.

In contrast to the overall good fit to the asymptotic IRT distribution, the fits to the asymptotic conditional IRT distributions ranged from good to very bad, and show no particularly consistent patterns of fit. This suggests that sequential effects are more complex than those proposed by the model.

Taking all the results together, the model does not seem to have suffered too badly in its first confrontation with experimental data. It does at least appear to provide a reasonable theoretical foothold in the highly experimentally orientated field of operant research.

Stephen Ambler
8 November 1972

APPENDIX I

This appendix contains two very general Algol programmes that can be used to calculate asymptotic IRT distributions. It also contains examples of their output. Comments are inserted in the programmes to make them reasonably self-explanatory. The various procedures are discussed in a little more detail in the next few paragraphs.

i) SAGENAR.

This calculates the asymptotic IRT distribution for a given reinforcement schedule, sampling function, and spread function. The procedures specifying these functions need modifying, depending on the functions actually used.

a) U(T)

This gives the reinforcement schedule. The programme is in fact set up for a random interval schedule with $\gamma = 1/30$.

b) PSI(T)

This is the sampling function. The actual function used here is $\psi(t) = \rho$, in line with the choice made in section 2.5. ρ is set at 5.0.

c) OMEGA(T,X) and SETLIMITS(AL, AU, T)

These two functions together specify w(t;x), the spread function. For any value of T, SETLIMITS makes OMEGA assume values only between AL (lower limit), and AU (upper limit). Here they are both set equal to 0.05. In terms of section 2.4, a = 0.05. OMEGA(T,X) fixes the value of OMEGA, given that it is constrained by SETLIMITS. In the present case, OMEGA is set equal to 1.0. Thus together, SETLIMITS and OMEGA specify,

$$w(t;x) = \begin{cases} 1.0 & x-a \leq t \leq x+a, \\ 0.0 & \text{otherwise,} \end{cases}$$

where a = 0.05.

The details of these four procedures are up to the user. The programme can be used to investigate several aspects of the model by varying these procedures. E.g. holding U(T) and PSI(T) fixed OMEGA(T,X) and SETLIMITS(AL, AU, T) could be varied to study the effects different spread functions have on the asymptotic IRT distribution.

d) IRTDENSITY(ALPHAO, ALPHA1, THETA)

This procedure is the core of the programme. ALPHAO and ALPHA1 are the response strength limits and THETA the learning parameter. Essentially the procedure tries to find an r(t) that satisfies equations 2.2.a and 2.3.b. The procedure works in the following manner.

A finite set of values of r(t) are arbitrarily specified, (s_1). Equation 2.3.b is then used to calculate values of Ø(t), (This is done in the local procedure PHI(T).) using the set s_1 of supposed values of r(t) at different values of t. Ø(t) is then used to calculate a new set of values for the set of points of r(t), (s_2). In general, s_1 and s_2 are different. As a measure of this difference, the total square error between corresponding points of s_1 and s_2 is calculated. If this square error (SQERR) is less than a critical value, then the points of s_2 are taken to represent points of r(t). If this is not so, the s_1 values are dumped and replaced by the s_2 values and the process repeated. This is continued until the difference between two consecutive sets is less than the critical value, or until the iteration has gone through 25 cycles.

e) R(T)

This takes the set of values output by IRTDENSITY and uses linear interpolation to provide a continuous approximation to the IRT distribution.

f) DIVISOR

This computes the integral of r(t) over the total range used, and this is then printed out. If IRTDENSITY has produced a reasonable approximation, then DIVISOR should be close to 1.0.

DATA

The data must be inserted in the following order.

'TITLE' This is space for up to 100 letters or spaces. It is meant for the name of the reinforcement schedule, but may be used for any purpose. This input must be enclosed in string quotes.

UL This is the time range (0-UL seconds) over which r(t) is calculated. The initial values of r(t) at the start of the iteration are taken to be 1/UL.

S This is the increment used in selecting the finite set of r(t) values. Values of r(t) will be computed every Sth of a second, starting at zero. There are thus UL/S values in all. The critical value of SQERR in the iteration is $S^2/2$.

ALPHAO, ALPHA1 These are the response strength limits of Ø(t). Most often they will be 0.0 and 1.0 respectively.

THETA This is the learning parameter. It is the relative effectiv -ness of non-reinforcement to reinforcement.

At the end of the programme is an example of the data layout. This is followed by an example of the output.

ii) SACONDAR.

This calculates IRT distributions conditional on the previous response being either reinforced or non-reinforced. Most of the procedures are common to SAGENAR and so are not given in detail.

a) PHI(T)

This is the response strength function. Once R(T) has been calculated, a range of Ø(t) values are calculated and stored for further use in array OUTPHI. At the same time, values of $\int w(t;x)r(x)dx$ are stored in array ROMEGA.

b) CR(T,K)

This calculates the conditional IRT distribution. Interpolation is used on OUTPHI and ROMEGA to provide a continuous result.

Finally the programme outputs a set of values for r(t|0), r(t), and r(t|1) at half second intervals.

DATA

This must be inserted in the following order. It is very similar to that for SAGENAR.

'TITLE' UL S ALPHAO ALPHA1	As for SAGENAR.

THETAO, THETA1 These are the learning parameters. THETAO is the effect of non-reinforcement, while THETA1 is the effect of reinforcement. They must be between 0.0 and 1.0.

At the end of the programme is an example of the data input. This is followed by an example of the output.

```
SAGENAR;

"COMMENT" THIS PROGRAMME CALCULATES A SET OF VALUES ON
         RANGE Ø-UL OF THE IRT DISTRIBUTION. THE POINTS
          ARE TAKEN AT INTERVALS OF S APART. THE RESULTS
          ARE OUTPUT IN ARRAY OUTSR[Ø:UL/S];


"BEGIN"
"REAL" UL, S;
"INTEGER" J;
"INTEGER""ARRAY" TITLE[1:25];

J:=1;
INSTRING(TITLE,J);
"READ" UL, S;

J:=1;
OUTSTRING(TITLE,J);
"PRINT" ''L2'';

"PRINT" ''L'', SAMELINE, ALIGNED(2,2), ' TIME RANGE:
       Ø.ØØ-',UL,
         ''L'', SAMELINE, ALIGNED(1,3), ' INCREMENT IN

             ITERATION: ', S;
"PRINT" ''L2'';
"BEGIN"
"REAL""ARRAY" OUTSR[Ø:(UL)/S+2];

"REAL""PROCEDURE" U(T);
"VALUE" T;
"REAL" T;

"COMMENT" GIVES REINFORCEMENT PROBABILITY;

"BEGIN"
U:=1-EXP(-T/3Ø);
"END" OF U;

"REAL""PROCEDURE" PSI(T);
"VALUE" T;
"REAL" T;

"COMMENT" SAMPLING DISTRIBUTION;

"BEGIN"
PSI:=5.Ø;
"END" OF PSI;
```

```
 "REAL""PROCEDURE" OMEGA(T,X);
 "VALUE" T,X;
 "REAL" T,X;

 "COMMENT" SPREAD FUNCTION. THIS FUNCTION IS ASSUMED TO BE
         ZERO UNLESS X-AL<T<X+AU;

"BEGIN"
OMEGA:=1.0;
"END" OF OMEGA;

"PROCEDURE" SETLIMITS(AL,AU,T);
"VALUE" AL,AU,T;
"REAL" AL,AU,T;

"COMMENT" FIXES RANGE OUTSIDE WHICH OMEGA IS ZERO;

"BEGIN"
AL:=AU:=0.05;
"END" OF SETLIMITS;

"PROCEDURE" IRTDENSITY(ALPHA0, ALPHA1, THETA);
"VALUE" ALPHA0, ALPHA1, THETA;
 "REAL" ALPHA0, ALPHA1, THETA;

"COMMENT" OUTPUTS A FINITE NUMBER OF VALUES OF THE IRT
          DENSITY DISTRIBUTION ON 0.0 TO UL (UPPER
          LIMIT) AT POINTS S APART.
            THIS IS OUTPUT FROM ARRAY OUTSR[0:UL/S].
                THE FOLLOWING FUNCTIONS ARE NEEDED:-
           U(T) -- THE REINFORCEMENT FUNCTION.
           PSI(T) -- THE SAMPLING FUNCTION.
           OMEGA(T,X) -- THE SPREAD FUNCTION.
           SETLIMITS(AL,AU,T)  -- GIVES LIMITS OF SPREAD
                                     FUNCTION.
                 THE FOLLOWING CONSTANTS ARE GLOBAL AND
           MUST BE GIVEN VALUES BEFORE IRTDENSITY IS
          CALLED --- UL,S.
            ARRAY OUTSR[0:UL/S] MUST BE DECLARED
                AFTER THESE;

"BEGIN"
"REAL" A, AU, AL;
"INTEGER" I, L;
SETLIMITS(AL,AU,UL);
A:=AU;
L:=ENTIER((UL+A)/S)+1;
SETLIMITS(AL,AU,0);
A:=AL;
```

```
"BEGIN"
"REAL""ARRAY" SRA[-(A/S+2):L];

"REAL""PROCEDURE" PHI(T);
"VALUE" T;
 "REAL" T;

"COMMENT" RESPONSE STRENGTH FUNCTION;

"BEGIN"
"REAL" X, Y, X1, Y1;
"INTEGER" INC;
X:=Y:=0.0;

"IF" A<0.000001 "THEN"
     "BEGIN"
     X:=1.0;
     Y:=U(T);
     "END"
"ELSE"
     "BEGIN"
     SETLIMITS(AL,AU,T);
     "FOR" INC:=(T-AL)/S "STEP" 1 "UNTIL" ((T+AU)/S-1) "DO"
          "BEGIN"
          X:=X+
          (OMEGA(T,INC*S)*SRA[INC]+OMEGA(T,(INC+1)*S)
           *SRA[INC+1])*S/2;
          Y:=Y+(OMEGA(T,INC*S)*U(INC*S)*SRA[INC]+
                OMEGA(T,(INC+1)*S)*U((INC+1)*S
                     )*SRA[INC+1])*S/2;
           "END";
    "END";
X1:=ALPHA0*THETA*(X-Y) + ALPHA1*Y;
Y1:=THETA*(X-Y) + Y;
"IF" Y1<0.000001 "THEN" PHI:=ALPHA0
"ELSE" PHI:=X1/Y1;
"END" OF PHI;

"FOR" I:=-(A/S+2) "STEP" 1 "UNTIL" 0 "DO"
SRA[I]:=0.0;

"FOR" I:=0 "STEP" 1 "UNTIL" L "DO"
SRA[I]:=1/UL;

 "COMMENT" FROM THE IMPLICIT RELATION IN SR THAT
          SR MUST SATISFY, A SET OF VALUES OF SR
          AT POINTS S APART ARE CALCULATED,
          ASSUMING THAT THE VALUES SRA ACTUALLY
          SPECIFY SR. THE SETS SRA AND SRA1 ARE
          THEN COMPARED. IF THEY ARE ALIKE IT IS
          ASSUMED THAT THEY GIVE POINTS OF SR.
```

```
            IF THEY ARE NOT ALIKE, SRA1 IS USED AS
            A NEW SRA AND THE PROCESS IS REPEATED.
            THIS CONTINUES UNTIL A STABLE SET OF
            VALUES IS FOUND, OR MORE THAN 25 CYCLES
            IS MADE, WHICHEVER IS THE SOONER;

        "BEGIN"
        "REAL" SQERR, Z, P1, P2;
        "INTEGER" COUNT;
        "REAL""ARRAY" SRA1[0:L];
        Z:=P1:=P2:=0.0;
        COUNT:=0.0;
        "PRINT" ' SQERR VALUES ARE: `, ''L2``;

    LOOP:

        COUNT:=COUNT+1;
        "IF" COUNT>25 "THEN"
            "BEGIN"
            "PRINT" ''L``, SAMELINE,
                    'NO STABILITY AFTER 25 LOOPS, SQERR IS `,
                    SQERR, ''L``;
            "GOTO" FINISH;
            "END";

        "FOR" I:=0 "STEP" 1 "UNTIL" L "DO"
            "BEGIN"
            P1:=PHI(I*S)*PSI(I*S);
            "IF" I=0 "THEN" Z:=0 "ELSE"
            Z:=Z+(((P1+P2)/2)*S);
            SRA1[I]:=P1*EXP(-Z);
            P2:=P1;
            "END";

        SQERR:=0.0;
        "FOR" I:=0 "STEP" 1 "UNTIL" L "DO"
        SQERR:=SQERR+(SRA[I]-SRA1[I])*(SRA[I]-SRA1[I]);
        "PRINT" SAMELINE, SQERR;
        "IF" SQERR<S*S/2 "THEN"
        "GOTO" FINISH;
        "FOR" I:=0 "STEP" 1 "UNTIL" L "DO"
        SRA[I]:=SRA1[I];
        "GOTO" LOOP;

FINISH:
       "FOR" I:=0 "STEP" 1 "UNTIL" UL/S "DO"
        OUTSR[I]:=SRA1[I];
        "END" OF LOOP;
    "PRINT" ''L``;
    "END";
"END" OF IRT DENSITY;
```

```
"FOR" J:=Ø "STEP" 1 "UNTIL" ENTIER((UL)/S)+1 "DO"
OUTSR[J]:=Ø.Ø;

"BEGIN"
"REAL" ALPHAØ,ALPHA1,THETA,E;

"REAL""PROCEDURE" R(T);
"VALUE" T;
"REAL" T;

"COMMENT" LINEAR INTERPOLATION OF OUTSR. R IS A CONTINUOUS
          APPRØXIMATIØN TO THE IRT DISTRIBUTION;

"BEGIN"
"REAL" X;
"INTEGER" I;
I:=ENTIER(T/S);
"IF" T < Ø.Ø "THEN"
X:=Ø.Ø;
"IF" T "GE" Ø.Ø "AND" T "LE" UL "THEN"
    "BEGIN"
    "IF" I=ENTIER(UL/S) "THEN"
    X:=OUTSR[I]-OUTSR[I]*(T/S-I) "ELSE"
    X:=OUTSR[I]+(OUTSR[I+1]-OUTSR[I])*(T/S-I);
    "END";
 "IF" T > UL "THEN"
    "BEGIN"
    "PRINT" ''L' CAUTION BOUNDS OF R EXCEEDED 'L'';
    X:=Ø.Ø;
    "END";
R:=X;
"END" OF R;

"READ" ALPHAØ, ALPHA1, THETA;
"PRINT" ''L'', SAMELINE, ALIGNED(1,3), ' RESPONSE
          STRENGTH LIMITS: ',
         ' ALPHAØ ', ALPHAØ, ''S2'', ' ALPHA1 ', ALPHA1,
        ''L'', SAMELINE, ALIGNED(1,3), ' LEARNING
          PARAMETER:  ',
         ' THETA ', THETA,
         ''L3'';

IRTDENSITY(ALPHAØ,ALPHA1,THETA);

"BEGIN"
"REAL" DIVISOR;
"INTEGER" J;
```

```
"COMMENT" DIVISOR SHOULD BE APPROXIMATELY 1.0. THIS PROVIDES
          A CHECK ON THE ACCURACY OF THE PROGRAMME;

DIVISOR:=0.0;
FOR" J:=0 "STEP" 1 "UNTIL" ENTIER(UL/S) "DO"
DIVISOR:=DIVISOR+OUTSR[J]*S;
"PRINT" ''L'';
"PRINT" SAMELINE, 'DIVISOR', DIVISOR, ''L2'';
"END";


"PRINT" ''L'';

"PRINT" '    T           R(T)', ''L2'';
"FOR" E:=0.000 "STEP" 0.5 "UNTIL" 10 "DO"
"BEGIN"
"PRINT" SAMELINE, ALIGNED(3,3), E, ''S3'', R(E), ''L2'';
"IF" R(E)<0.0005 "AND" E>UL/2 "THEN"
    "BEGIN"
    "PRINT" ''L'', 'STOP', ''L'';
    "GOTO" STOP;
    "END";
"END";

"END";

"END";

STOP:
"PRINT" ''L10'';
"END" OF PROGRAMME;
```

THE DATA LAYOUT IS AS FOLLOWS

SCHEDULE IS
' RANDOM INTERVAL SCHEDULE, GAMMA IS 1/30 '

TIME RANGE
10 SECONDS

INCREMENT
0.01 SECONDS

RESPONSE STRENGTH LIMITS
0.0   1.0

LEARNING PARAMETER
1.0

END OF DATA LAYOUT.

AN EXAMPLE OF THE OUTPUT IS:

```
 RANDOM INTERVAL SCHEDULE. GAMMA IS 1/30


 TIME RANGE:    0.00 - 10.00
 INCREMENT IN ITERATION:  0.010


 RESPONSE STRENGTH LIMITS:  ALPHA0  0.000   ALPHA1  1.000
 LEARNING PARAMETER:   THETA  1.000


 SQERR VALUES ARE:

 7.5607990        .00000000

DIVISOR .99943031
```

T	R(T)
0.000	0.000
0.500	0.081
1.000	0.151
1.500	0.203
2.000	0.233
2.500	0.241
3.000	0.230
3.500	0.206
4.000	0.174
4.500	0.140
5.000	0.107
5.500	0.078
6.000	0.055
6.500	0.037
7.000	0.024
7.500	0.015
8.000	0.009
8.500	0.005
9.000	0.003
9.500	0.002
10.000	0.000

STOP

```
SACONDAR;

"COMMENT" THIS PROGRAMME CALCULATES A SET OF VALUES AT
            POINTS OF THE IRT DISTRIBUTION, ON Ø -UL
             AT INTERVALS S APART AND PUTS RESULTS IN
         ARRAY OUTSR[Ø:UL/S]. THESE RESULTS ARE THEN
          USED TO CALCULATE THE IRT DISTRIBUTIONS
         CONDITIONAL ON THE PREVIOUS RESPONSE BEING (A)
           REINFORCED, AND (B) NOT REINFORCED;
BEGIN"
"REAL" UL, S;
"INTEGER" J;
"INTEGER""ARRAY" TITLE[1:25];

J:=1;
INSTRING(TITLE,J);
"READ" UL, S;

J:=1;
OUTSTRING(TITLE,J);
"PRINT" ''L2'';
"PRINT" ''L'', SAMELINE, ALIGNED(2,2), ' TIME RANGE:
            Ø.ØØ -', UL,
        ''L'', SAMELINE, ALIGNED(1,3), ' INCREMENT IN
            ITERATION: ', S;
"PRINT" ''L2'';

"BEGIN"
"REAL""ARRAY" OUTSR[Ø:(UL)/S+2];

"REAL""PROCEDURE" U(T);

"REAL""PROCEDURE" PSI(T);

"REAL""PROCEDURE" OMEGA(T,X);

"PROCEDURE" SETLIMITS(AL,AU,T);

"PROCEDURE" IRTDENSITY(ALPHAØ, ALPHA1, THETA);

"FOR" J:=Ø "STEP" 1 "UNTIL" ENTIER((UL)/S)+1 "DO"
OUTSR[J]:=Ø.Ø;

"BEGIN"
"REAL" ALPHAØ, ALPHA1, THETAØ, THETA1, THETA, E;
"INTEGER" I;
"REAL""ARRAY" ROMEGA, OUTPHI[Ø:UL/S+2], RF[Ø:1];
```

```
"COMMENT" OUTPHI STORES VALUES OF THE RESPONSE STRENGTH
           FUNCTION PHI AT INTERVALS OF S APART ON THE RANGE
            Ø-UL.
          ROMEGA STORES VALUES OF THE INTEGRAL
           R(X)*OMEGA(T,X)  WITH RESPECT TO X BETWEEN Ø AND UL.
            VALUES ARE STORED AT INTERVALS S APART, AS T RANGES
             BETWEEN Ø AND UL;

 "REAL""PROCEDURE" R(T);


 "REAL""PROCEDURE" PHI(T);
 "VALUE" T;
  "REAL" T;

 "COMMENT" RESPONSE STRENGTH FUNCTION;

 "BEGIN"
 "REAL" X, Y, X1, Y1, AL, AU, A;
 "INTEGER" INC;
 SETLIMITS(AL,AU,T);
 A:=AL+AU;
 X:=Y:=Ø.Ø;
 "IF" A<Ø.ØØØØØ1 "THEN"
     "BEGIN"
     X:=1.Ø;
     Y:=U(T);
     "END"
 "ELSE"
     "BEGIN"
     "FOR" INC:=(T-AL)/S "STEP" 1 "UNTIL" ((T+AU)/S-1) "DO"
         "BEGIN"
         X:=X+
         (OMEGA(T,INC*S)*R(INC*S)+OMEGA(T,(INC+1)*S)
             *R((INC+1)*S))*S/2;
         Y:=Y+(OMEGA(T,INC*S)*U(INC*S)*R(INC*S)+
               OMEGA(T,(INC+1)*S)*U((INC+1)*S)
               *R((INC+1)*S))*S/2;
         "END";
     "END";
 X1:=ALPHAØ*THETA*(X-Y) + ALPHA1*Y;
 Y1:=THETA*(X-Y) + Y;
 "IF" Y1<Ø.ØØØØØ1 "THEN" PHI:=ALPHAØ
 "ELSE" PHI:=X1/Y1;
  "END" OF PHI;
```

```
"REAL" "PROCEDURE" CR(T,K);
"VALUE" T,K;
"REAL" T;
"INTEGER" K;

"COMMENT" CALCULATES THE IRT DISTRIBUTION CONDITIONAL
            ON REINFORCEMENT (K=1), OR NON-REINFORCEMENT (K=0)
            OF THE PRECEEDING RESPONSE;

"BEGIN"
"REAL" CONDPHI, FPHI, FPHI1, FPHI2;
"INTEGER" I,L;
FPHI:=FPHI1:=FPHI2:=0.0;
CONDPHI:=0.0;
L:=ENTIER(T/S);

"IF" K=1 "THEN"
    "BEGIN"
    "FOR" I:=0 "STEP" 1 "UNTIL" L-1 "DO"
    CONDPHI:=CONDPHI+PSI(I*S)*
              (OUTPHI[I]+(ALPHA1-OUTPHI[I])
               *THETA1*ROMEGA[I])*S/2
              + PSI((I+1)*S)*
              (OUTPHI[I+1]+(ALPHA1-OUTPHI[I+1])
                 *THETA1*ROMEGA[I+1])*S/2;

    FPHI1:=OUTPHI[L]+(ALPHA1-OUTPHI[L])*THETA1*ROMEGA[L];
    FPHI2:=OUTPHI[L+1]+(ALPHA1-OUTPHI[L+1])*THETA1*ROMEGA[L+1];
    "END";

"IF" K=0 "THEN"
    "BEGIN"
    "FOR" I:=0 "STEP" 1 "UNTIL" L-1 "DO"
    CONDPHI:=CONDPHI+PSI(I*S)*
              (OUTPHI[I]+(ALPHA0-OUTPHI[I])*
               THETA0*ROMEGA[I])*S/2
               + PSI((I+1)*S)*
               (OUTPHI[I+1]+(ALPHA0-OUTPHI[I+1]
               )*THETA0*ROMEGA[I+1])*S/2;

     FPHI1:=OUTPHI[L]+(ALPHA0-OUTPHI[L])*THETA0*ROMEGA[L];
    FPHI2:=OUTPHI[L+1]+(ALPHA0-OUTPHI[L+1])*THETA0*ROMEGA[L+1];
    "END";

    FPHI:=FPHI1+(FPHI2-FPHI1)*(T/S-L);
    FPHI:=FPHI*PSI(T);
    CONDPHHI:=CONDPHI+FPHI*(T/S-L)*S;
CR:=FPHI*EXP(-CONDPHI);
"END" OF CR;
```

```
"READ" ALPHAØ, ALPHA1, THETAØ, THETA1;
THETA:=THETAØ/THETA1;
"PRINT" ''L'', SAMELINE, ALIGNED(1,3), ' RESPONSE
         STRENGTH LIMITS:        ',
        ' ALPHAØ ', ALPHAØ, ''S2'', ' ALPHA1 ', ALPHA1,
        ''L'', SAMELINE, ALIGNED(1,3), ' LEARNING
         PARAMETERS:  ',
       ' THETAØ ', THETAØ, ''S2'', ' THETA1 ', THETA1, ''S2'',
        ' THETA ', THETA,
        ''L3'';

"FOR" I:=Ø "STEP" 1 "UNTIL" UL/S "DO"
OUTPHI[I]:=ROMEGA[I]:=Ø.Ø;

IRTDENSITY(ALPHAØ,ALPHA1,THETA);

"BEGIN"
"REAL" DIVISOR;
"INTEGER" J;
DIVISOR:=Ø.Ø;
RF[1]:=RF[Ø]:=Ø.Ø;
"FOR" J:=Ø "STEP" 1 "UNTIL" ENTIER(UL/S) "DO"
    "BEGIN"
    DIVISOR:=DIVISOR+OUTSR[J]*S;
    RF[1]:=RF[1]+OUTSR[J]*U(J*S)*S;
    OUTPHI[J]:=PHI(J*S);
    "BEGIN"
    "REAL" P,Q;
    "INTEGER" K,L,M;
    SETLIMITS(P,Q,J*S);
    "IF" (J-P/S+1)<Ø "THEN" L:=1 "ELSE" L:=(J-P/S+1);
    "IF" (J+Q/S-1)>(UL/S) "THEN" M:=UL/S "ELSE" M:=(J+Q/S-1);
    "FOR" K:=L "STEP" 1 "UNTIL" M "DØ"
    ROMEGA[J]:=ROMEGA[J]+
                 OMEGA(J*S,K*S):*OUTSR[K]*S;
    ROMEGA[J]:=ROMEGA[J]
               + OMEGA(J*S, (L-1)*S)*OUTSR[L-1]*S/2
               + OMEGA(J*S, (M+1)*S)*OUTSR[M+1]*S/2;
    "END";
    "END";
RF[Ø]:=1 - RF[1];

"PRINT" ''L'';
"PRINT" SAMELINE, 'DIVISOR', DIVISOR, ''L2'';
"PRINT" SAMELINE,  ' MEAN PROBABILITY OF REINFORCEMENT: ',
         ALIGNED(1,3), RF[1], ''L2'';
"END";
"PRINT" ''L'';
```

```
"PRINT" '    T          CR(T,1)      R(T)        CR(T,0)', ''L2'';
"FOR" E:=0.000 "STEP" 0.5 "UNTIL" UL "DO"
"BEGIN"
"PRINT" SAMELINE, ALIGNED(3,3), E, ''S3'', CR(E,1), ''S3'',
        R(E), ''S3'', CR(E,0), ''L2'';
"IF" R(E)<0.0005 "AND" E>UL/2 "THEN"
    "BEGIN"
    "PRINT" ''L'', 'STOP', ''L'';
    "GOTO" STOP;
    "END";
"END";

"END";

"END";

STOP:
"PRINT" ''L10'';
"END" OF PROGRAMME;
```

AN EXAMPLE OF THE DATA LAYOUT FOLLOWS

```
SCHEDULE IS
' RANDOM INTERVAL SCHEDULE, GAMMA IS 1/30 '

TIME RANGE
10 SECONDS

INCREMENT
0.01

RESPONSE STRENGTH LIMITS
0.0   1.0

LEARNING PARAMETERS
0.5  0.5

END OF DATA LAYOUT
```

REFERENCES

ANDERSON, N.H., "Linear Models for Responses Measured on a Continuous Scale", Journal of Mathematical Psychology, 1964, 1, 121-142.

ANGER, D., "The Dependence of Interresponse Times upon the Relative Reinforcement of Different Interresponse Times", Journal of Experimental Psychology, 1956, 52, 145-161.

ANGER, D., "The Role of Temporal Discrimination in the Reinforcement of Sidman Avoidance Behaviour", Journal of the Experimental Analysis of Behaviour, 1963, 6, 477-506.

BLOUGH, P.M., & BLOUGH, D.S., "The Distribution of Interresponse Times in the Pigeon During Interval Reinforcement", Journal of the Experimental Analysis of Behaviour, 1968, 11, 23-27.

BRANDAUER, C.M., "The effects of Uniform Probabilities of Reinforcement on the Response Rate of the Pigeon", Unpublished Doctoral Dissertation, Columbia University, 1968.

BRUNER, A., & REVUSKY, S.H., "Collateral Behaviour in Humans", Journal of the Experimental Analysis of Behaviour, 1961, 4, 349-350.

BUSH, R.R., & MOSTELLER, F., "Stochastic Models for Learning", New York: Wiley, 1955.

CATANIA, A.C., "Concurrent Operants", in Operant Behaviour: Areas of Research and Application, Honig, W.K., (Editor), New York: Appleton-Century-Crofts, 1966, pp243-270.

CATANIA, A.C., "Reinforcement Schedules and Psychophysical Judgements", in The Theory of Reinforcement Schedules, Schoenfeld, W.N., (Editor), Appleton-Century-Crofts: New York, 1970, pp1-42.

CATANIA, A.C., & REYNOLDS, G.S., "A Quantitative Analysis of Responding Maintained by Interval Schedules of Reinforcement", Journal of the Experimental Analysis of Behaviour, 1968, 11, 327-383.

CLARK, R., "Some Time-correalated Reinforcement Schedules and their Effects on Behaviour", Journal of the Experimental Analysis of Behaviour, 1959, 2, 1-22.

CUMMING, W.W., & SCHOENFELD, W.N. "Some Data on Behaviour Reversability in a Steady State Experiment", Journal of the Experimental Analysis of Behaviour, 1959, 2, 87-90.

CUMMING, W.W. & SCHOENFELD, W.N., "Characteristics of Responding under a Temporally defined Schedule of Long Cycle Length", Journal of the Experimental Analysis of Behaviour, 1961, 4, 73-80.

ESTES, W.K., "Toward a Statistical Theory of Learning", Psychological Review, 1950, 57, 94-107.

ESTES, W.K., "Statistical Theory of Spontaneous Recovery and Regression", Psychological Review, 1955, 62, 145-154.

ESTES, W.K., & BURKE, C.J., "A Theory of Stimulus Variability in Learning", Psychological Review, 1953, 60, 276-286.

FARMER, J., "Properties of Behaviour under Random Interval Reinforcement Schedules", Journal of the Experimental Analysis of Behaviour, 1963, 6, 607-616.

FARMER, J., & SCHOENFELD, W.N., "Effects of a DRL Contingency added to a Fixed Interval Reinforcement Schedule", Journal of the Experimental Analysis of Behaviour, 1964(a), 7, 391-399.

FARMER, J., & SCHOENFELD, W.N., "Interreinforcement times for the Bar Pressing Response of White Rats on two DRL Schedules", Journal of the Experimental Analysis of Behaviour, 1964(b), 7, 119-122.

FERRARO, D.P., SCHOENFELD, W.N., & SNAPPER, A.G., "Sequential Response Effects in the White rat During Conditioning and Extinction on a DRL Schedule", Journal of the Experimental Analysis of Behaviour, 1965, 8, 255-260.

FERSTER, C.B., & SKINNER, B.F., "Schedules of Reinforcement", New York: Appleton-Century-Crofts, 1957.

FLESHLER, M., & HOFFMAN, H.S., "A Progression for generating Variable Interval Schedules of Reinforcement", Journal of the Experimental Analysis of Behaviour, 1962, 5, 529-530.

HEARST, E., "The Behavioural Effects of some Temporally Defined Schedules of Reinforcement", Journal of the Experimental Analysis of Behaviour, 1958, 1, 45-55.

HERRNSTEIN, R.J., "Secondary Reinforcement", in Operant Behaviour: Areas of Research and Application, Honig, W.K., (Editor), New York: Appleton-Century-Crofts, 1966, pp33-52.

HERRNSTEIN, R.J., "On the Law of Effect", Journal of the Experimental Analysis of Behaviour, 1970, 13, 243-266.

HODOS, W., ROSS, G.S., & BRADY, J.V., "Complex Response Patterns during Temporally Spaced Responding", Journal of the Experimental Analysis of Behaviour, 1962, 5, 473-79.

KELLEHER, R.T., FRY, W., & COOK, L., "Interresponse Time Distribution as a Function of Differential Reinforcement of Temporally Spaced Responses", Journal of the Experimental Analysis of Behaviour, 1959, 2, 91-106.

KENDALL, M.G., & STUART, R., "The Advanced Theory of Statistics", Vol 2, London: Griffin and Co., 1952.

KINTSCH, W., "Frequency Distribution of Interresponse Times during VI and VR Reinforcement", Journal of the Experimental Analysis of Behaviour, 1965, 8, 347-352.

LATIES, V.G., & WEISS, B., "Human Observing Behaviour after Signal Detection", Journal of the Experimental Analysis of Behaviour, 1960, 3, 27-33.

LATIES, V.G., & WEISS, B., "Effects of a Concurrent Task on Fixed Interval Responding in Humans", Journal of the Experimental Analysis of Behaviour, 1963, 6, 431-443.

LATIES, V.G., WEISS, B., CLARK, R.L., & REYNOLDS, M.D., "Overt 'Mediating' Behaviour during Temporally Spaced Responding", Journal of the Experimental Analysis of Behaviour, 1965, 8, 107-116.

LATIES, V.G., WEISS, B., & WEISS, A.B., "Further observations on Overt 'Mediating' Behaviour and the Discrimination of Time", Journal of the Experimental Analysis of Behaviour, 1969, 12, 43-57.

MALOTT, R.W., & CUMMING, W.W., "Schedules of Interresponse Time Reinforcement", Psychological Record, 1964, 14, 211-252.

MALOTT, R.W., & CUMMING, W.W., "Concurrent Schedules of IRT Reinforcement: Probability of Reinforcement and Lower Bounds of Reinforced IRT Intervals", Journal of the Experimental Analysis of Behaviour, 1966, 9, 317-325.

MILLENSON, J.R., "Some Behavioural Effects of a two-valued Temporally Defined Reinforcement Schedule", Journal of the Experimental Analysis of Behaviour, 1958, 2, 191-202.

MILLENSON, J.R., "Probability of Response and Reinforcement in a Response defined Analogue of an Interval Schedule", Journal of the Experimental Analysis of Behaviour, 1966, 9, 87-94.

MILLENSON, J.R., "Random Interval Schedules of Reinforcement", Journal of the Experimental Analysis of Behaviour, 1963, 6, 437-443.

MORSE, W.H., "Intermittent Reinforcement", in Operant Behaviour:

Areas of Research and Application, Honig, W.K., (Editor), New York: Appleton-Century-Crofts, 1966, pp52-108.

MUELLER, C.V., "Theoretical Relationships among some Measures of Conditioning", Proceedings of the National Academy of Sciences, 1950, 36, 308-311.

NORMAN, M.F., "An Approach to Free-responding on Schedules that Prescribe Reinforcement Probability as a Function of Interresponse Time", Journal of Mathematical Psychology, 1966, 3, 235-268.

RAY, R.C., & McGILL, W., "Effects of Class Interval Size upon Certain Distributions of Inter-response Times", Journal of The Experimental Analysis of Behaviour, 1964, 7, 125-127.

REVUSKY, S.H., "Mathematical Analysis of the Durations of Reinforced Inter-response Times During Variable Interval Reinforcement", Psychometrica, 1962, 27, 307-314.

SCHICK, K., "Operants", Journal of the Experimental Analysis of Behaviour, 1971, 15, 413-423.

SCHOENFELD, W.N., CUMMING, W.W., & HEARST, E., "On the Classification of Reinforcement Schedules", Proceedings of the National Academy of Sciences, 1956, 42, 563-570.

SHIMP, C.P., "The Reinforcement of Short Interresponse Times", Journal of the Experimental Analysis of Behaviour, 1967, 10, 425-434.

SHIMP, C.P., "Magnitude and Frequency of Reinforcement and Frequencies of Interresponse Times", Journal of the Experimental Analysis of Behaviour, 1968, 11, 525-535.

SHIMP, C.P., "The Concurrent Reinforcement of Two Interresponse Times: The Relative Frequency of an Interresponse Time equals its Relative Harmonic Length", Journal of the Experimental Analysis of Behaviour, 1969, 12, 403-411.

SIDLEY, N.A., & SCHOENFELD, W.N., "Behaviour Stability and Response Rate as a Function of Reinforcement Probability on Random Ratio Schedules", Journal of the Experimental Analysis of Behaviour, 1964, 7, 281-283.

SIDMAN, M., "Time Discrimination and Behavioural Interaction in a Free Operant Situation", Journal of Comparative and Physiological Psychology, 1956, 49, 469-473.

STUBBS, A., "The Discrimination of Stimulus Duration by Pigeons", Journal of the Experimental Analysis of Behaviour, 1968, 11 223-238.

SKINNER, B.F., "The Behaviour of Organisms", New York: Appleton-Century-Crofts, 1938.

SUPPES, P., "A Linear Model for a Continuum of Responses", in Studies in Mathematical Learning Theory, Bush, R.R., & Estes, W.K., (Editors), Stanford:Stanford University Press, 1959, pp400-414.

SUPPES, P., & DONIO, J., "Foundations of Stimulus Sampling Theory for Continuous-time Processes", Journal of Mathematical Psychology, 1967, 4, 202-225.

SUPPES, P., "Stimulus Sampling Theory for a Continuum of Responses", In Mathematical Methods in the Social Sciences, Arrow, J., Karlin, S., & Suppes, P., (Editors) Stanford: Stanford University Press, 1959, pp348-365.

SUPPES, P., & FRANKMANN, R.W., "Test of a Stimulus Sampling Theory for a Continuum of Responses with Unimodal Non-contingent Determinate Reinforcement", Journal of Experimental Psychology, 1961, 57, 122-132.

SUPPES, P., & ROUANET, H., "A Simple Discrimination Experiment with a Continuum of Responses", in Studies in Mathematical Psychology, Atkinson, R.C., (Editor), Stanford: Stanford University Press, 1964.

SUPPES, P., ROUANET, H., LEVINE, M., & FRANKMANN, R.W., "Empirical Comparison of Models for a Continuum of Responses with Non-contingent Bimodal Reinforcement", in Studies in Mathematical Psychology, Atkinson, R.C., (Editor), Stanford: Stanford University Press, 1964, pp358-379.

SUPPES, P., & ZINNES, J.L., "Stochastic Learning Theories for a Response Continuum with Non-determinate Reinforcement", Psychometrica, 1961, 26, 373-390.

SUPPES, P., & ZINNES, J.L., "A Continuous-Response Task with Nondeterminate, Contingent Reinforcement", Journal of Mathematical Psychology, 1966, 3, 197-216.

TERRACE, H.S., "Stimulus Control", in Operant Behaviour: Areas of Research and Application, Honig, W.K., (Editor), New York: Appleton-Century-Crofts, 1966, pp271-345.

WEINER, H., "Some Effects of Response Cost upon Human Operant Behaviour", Journal of the Experimental Analysis of Behaviour, 1962, 5, 201-208.

WEINER, H., "Response Cost and Fixed Ratio Performance", Journal of the Experimental Analysis of Behaviour, 1964(a), 7, 79-81.

WEINER, H., "Response Cost Effects During Extinction Following Fixed Interval Reinforcement in Humans", Journal of the Experimental Analysis of Behaviour, 1964(b), 7, 333-335.

WEINER, H., "Conditioning History and Human Fixed Interval Performance", Journal of the Experimental Analysis of Behaviour, 1964(c), 7, 383-385.

WEINER, H., "Controlling Human Fixed Interval Performance", Journal of the Experimental Analysis of Behaviour, 1969, 12, 349-373.

WEISS, B., "The Fine Structure of Operant Behaviour during Transition States", in Theory of Reinforcement Schedules, Schoenfeld, W.N., (Editor), New York: Appleton-Century-Crofts, 1970, pp277-311.

WILLIAMS, D.R., "The Structure of Response Rate", Journal of the Experimental Analysis of Behaviour, 1968, 11, 251-258.

WILSON, M.P., & KELLER, F.S., "Selective Reinforcement of Spaced Responses", Journal of Comparative and Physiological Psychology, 1953, 46, 190-193.

WYCKOFF, L.B. Jr., "The Role of Observing Responses in Discrimination Learning", Psychological Review, 1952, 59, 431-442.

ZURIFF, G.E., "Collateral Responding during Differential Reinforcement of Low Rates", Journal of the Experimental Analysis of Behaviour, 1969, 12, 971-976.