OPERANT REMEDIAL TECHNIQUES WITH
READING RETARDATION IN CHILDREN

Thesis submitted in fulfilment
of the requirements for the
Degree of Doctor of Philosophy.

Helen Russell WINEFIELD
B.A. (Hons. 1st class) Adel. 1967
Department of Psychology
University of Adelaide
South Australia

Date submitted February 1973.
# TABLE OF CONTENTS

## SECTION I. LITERATURE REVIEW AND THEORETICAL DISCUSSION

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Introduction</td>
<td>7</td>
</tr>
<tr>
<td>II</td>
<td>Reading as operant behaviour</td>
<td>14</td>
</tr>
<tr>
<td>III</td>
<td>Reinforcers suitable in educational settings</td>
<td>40</td>
</tr>
<tr>
<td>IV</td>
<td>Token reinforcement for reading</td>
<td>65</td>
</tr>
<tr>
<td>V</td>
<td>Reinforcement and human learning</td>
<td>89</td>
</tr>
<tr>
<td>VI</td>
<td>Other approaches to reading retardation and its remediation</td>
<td>109</td>
</tr>
</tbody>
</table>

## SECTION II. RESEARCH REPORT

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>VII</td>
<td>Fact finding</td>
<td>124</td>
</tr>
<tr>
<td>VIII</td>
<td>Experiment I. Effects on retarded readers' progress of extra teaching, tokens, kind of backup reinforcement, and response-cost for errors</td>
<td>148</td>
</tr>
<tr>
<td>IX</td>
<td>Experiments II and III. Followups of the response-cost effect with multiple-choice material</td>
<td>176</td>
</tr>
<tr>
<td>X</td>
<td>Experiment IV. Response-cost with a paired associate learning task</td>
<td>188</td>
</tr>
<tr>
<td>XI</td>
<td>Experiment V. Effects of contingent and noncontingent material reinforcement on reading fluency and error rate</td>
<td>199</td>
</tr>
<tr>
<td>XII</td>
<td>Experiment VI. Effects of extrinsic and intrinsic backup reinforcement on fluency and error rate</td>
<td>220</td>
</tr>
<tr>
<td>XIII</td>
<td>Summary of main findings and conclusions</td>
<td>243</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>A: Technical diagram of apparatus used for all experiments</td>
<td>260</td>
<td></td>
</tr>
<tr>
<td>B: Experiment I: Teacher's rating scales</td>
<td>261</td>
<td></td>
</tr>
<tr>
<td>Subjects' post-treatment Questionnaire, and Examples of Teaching Materials</td>
<td>262</td>
<td></td>
</tr>
<tr>
<td>C: Experiment I: Tables of means, and significant analyses of variance</td>
<td>265</td>
<td></td>
</tr>
<tr>
<td>D: Materials used for Experiments II and III</td>
<td>268</td>
<td></td>
</tr>
<tr>
<td>E: Detailed instructions used for Experiments II, III, IV and V</td>
<td>274</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BIBLIOGRAPHY</th>
<th></th>
</tr>
</thead>
</table>

277
Summary

The last decade has seen a steadily increasing interest on the part of psychologists, in applying the principles of behaviour modification in educational settings. Token reinforcement in particular has been found appropriate and effective in schools. The work to be reported aimed to explore in detail which are the aspects which are critical to success in using tokens to acquire control of complex cognitive processes in children.

The selection of reading as a target behaviour involved problems of definition, observation and measurement, both theoretical and practical, which are discussed. The relevant literature is reviewed from the fields of operant-based research with children in the classroom, education and educational psychology, and the experimental psychology of human learning.

An interrelated series of three exploratory studies and six experiments was undertaken with the aim of investigating the following aspects of token reinforcement for reading achievement:

a) the dynamics and effects of response-cost in the form of token-deduction for incorrect reading responses.

b) a comparison of tangible positive reinforcers with social reinforcement and information on progress alone, with non-contingent appetitive stimulation, and with "intrinsic" reinforcers. The subjects were middle-class, native English-speaking children of normal intelligence, mostly of 9 - 10 years of age and experiencing difficulties in learning to read by the conventional methods. The experimental design included both within- and between-subject comparisons.
Main results are cautionary with regard to the efficacy of token reinforcement for reading. Knowledge of results and mild intensities of social reinforcement were found to be as effective as additional material reinforcers. The provision of a response-cost contingency for errors led to superior performance while the token system was operating and a relative decrement thereafter. Material rewards were associated with apparently greater tension during lessons and more incorrect responses.

Several suggestions are made for future consequent research.
Author's Statement

This thesis contains no material which has been accepted for the award of any other degree or diploma at The University of Adelaide or any other university. To the best of my knowledge and belief, the thesis contains no material previously published or written by any other person except as due reference is made in the text of the thesis.

Signed

Helen R. Winefield
Acknowledgements

My thanks are due to my supervisor, Dr. F. R. Dalziel, for his stimulation and encouragement. I am also very grateful to the Headmistress, Headmasters, and teachers of Adelaide metropolitan primary schools, who granted access to their pupils for my use as subjects.
Chapter I. Introduction

During the last decade, clinical applications of psychological knowledge have been greatly extended. Many more psychologists have accepted the challenge and the responsibility of seeking to alter behaviour. The most important cause of this expansion of therapeutic activity on the part of psychologists, is the growth of interest and experience in applying the principles of learning in a systematic manner, to real-life extra-laboratory human behaviour.

Behaviour modification, or more specifically the technique of behaviour analysis and shaping originating from the operant approach of B.F. Skinner, has particular suitability in work with children. Behaviour disorders in the young are aversive to guardians, and susceptible to retraining to an extent which makes both parents and researchers eager to use modification techniques; these are being continually refined and diversified at present.

Before 1962 little attempt had been made by learning theorists to apply laboratory-derived principles to unacceptable behaviour in children, with the exceptions of bedwetting and phobias. As experimental evidence of therapeutic efficacy has accumulated at an accelerating rate, attention has been turned not only to overt, motor behaviours such as tantrums, disruptive classroom behaviour
and deficits in developmental skills, but also to symbolic and cognitive forms of behaviour.

Paramount amongst the cognitive skills demanded for success in our technologically-advanced society, are those involved in educational achievement, and of these the ability to read takes first place in importance. Many vocational opportunities are closed, and much pleasure and information is denied, to the illiterate. It was predictable, therefore, that techniques of behaviour change originally developed for psychiatric disorders would eventually be applied to the reading difficulty experienced by many otherwise apparently normal citizens.

Reading is a learned behaviour, even more obviously than irrational fears or maladaptive modes of relating interpersonally are learned behaviours. The literature contains several reports of experimenters trying to remedy deficiencies in the learning-to-read process, by analysing it in operant terms and then by reinforcing the crucial components of the skill. This thesis is concerned with the problem of how the person whose reading is poor can be helped to read better. Many remedial methods have been proposed and tried by educators in the past and the proportion of retarded readers has, as far as can be judged, remained similar. Accordingly, it is the possible contribution of psychologists, basing their interventions on the operant view of human learning, which is of special
interest here. The successes with systematic reinforcement of desired reading behaviour which are noted above, seem to hold the promise of a new and logical method of attacking an old and disabling problem, and thus to deserve further investigation.

A prerequisite is the analysis of reading in learning theory terms. Chapter II carries out such an analysis and demonstrates how, due to the length and complexity of the associative learning task whereby printed stimuli acquire discriminative control over verbal responses, the question of what reinforcement maintains the learning process assumes great importance. Understandably, the emphasis in behaviour modification studies of children has tended to be placed on control by positive reinforcement (and its withdrawal) rather than on control by the physically aversive treatments often used with adult clinical patients (e.g. "aversion therapy"); aversive operant techniques involving electric shock such as in the work of Lovaas (1967) provide a noted exception.

Behaviour modification is peculiarly suitable for use in educational settings. Its emphasis on control by appetitive rather than aversive stimuli, the long-standing interest in educational technology of its founder, and its growing application as a preventive approach in natural environments (Goodall, 1972) all make the developing, learning child an ideal subject for its use.
In fact, an extensive literature has accumulated, which deals with attempts to modify the behaviour of children in educational settings by manipulating reinforcement contingencies. Chapter III surveys the reinforcers which have been found effective in a variety of schoolroom situations, with particular reference to the use of token programmes.

Token reinforcement procedures have been applied mainly to overt, easily measurable behaviours such as disruptive talk and attention-getting, but also to the more subtle area of cognitive skills. Reading, a complex skill which is both a significant aspect of an individual's functioning, and a challenge to the theoretical and practical ingenuity of the behaviour modifier, presents problems of definition, observation and measurement of the behaviour. These problems are described in Chapter II; the question of suitable research designs where therapy is intended to be permanent, is discussed in Chapter III; and the literature on token reinforcement for reading is surveyed in Chapter IV. The main gaps in understanding identified in Chapter IV, can be foreshadowed as throwing doubt on claims that token programmes affect reading skill. In particular, existing studies display the following deficiencies:

a) They do not distinguish the informational role of tokens from their function as signals of future material reinforcement.
b) They have failed to include control subjects who have received other (non-token) forms of remedial instruction in reading, and have been deficient in follow-up.

c) They leave untouched the question of response-cost for undesirable responses and the effects of aversive stimulation in general on the progress of cognitive behaviour change.

These are the deficiencies which motivated the current research, described in Section II. However, to gain as complete an understanding of the problem as possible, contributions from the literature in more traditional areas of psychological research were also sought.

Reading exemplifies the human ability to develop automatic cognitive skills which then have high utility over a generality of situations. Experimental psychologists have for a long time been concerned with covert behaviour of this kind. Memory, discrimination, association, and the coding and transfer of information, are traditional areas of investigation in experimental psychology. The ability to read can, therefore, be seen as an accomplishment which casts interesting light on human mental functions. The discovery of what characteristics of the learning situation may hinder or accelerate the process, has theoretical significance in addition to its practical educational implications.
Chapter V is concerned with the contributions of experimental psychology which might prove relevant to an understanding of the acquisition of cognitive skills. In particular, studies are reviewed which investigate the effects of punishment (verbal and non-verbal) on human performance and verbal learning, and the effects of comparing different kinds of positive reinforcers (e.g. tangible and social) with knowledge-of-results alone, in children of different age, social class and other characteristics. It is concluded from this review that the operation of the low-intensity "reinforcers" suitable for experiments with children can largely be understood in terms of their informational content.

Educational psychologists and others have long been concerned with the nature of reading skill and possible explanations for its non-acquisition. Chapter VI assesses the results of this approach which are, on the whole, seen to be disappointing in both conceptualization of the causes of reading backwardness, and prescriptions for its remedy. Whilst these prescriptions unanimously advise the elimination of aversive stimulation from the re-training situation, specification of how to reinforce progress, and the relative likelihood of success with different methods, has been vague.

The experimental work described in Section II aimed to extend earlier studies on token reinforcement for reading,
towards the most serious of the identified information gaps. The effects of systematic penalties for incorrect responses, and the informational vs. reward value of tokens, provided the initial focus for this work.

Several exploratory studies were first carried out, concerned with the mechanics of token reinforcement procedures, the incidence and possible measurement of reading disability in South Australia, and the relationship between the "natural frequency" of reading and reading skill. A long-term matched-groups experiment was then performed, to explore the questions raised above by the critical review of previous studies. Several shorter studies arose from the results of the first main experiment; the main findings and conclusions, and their implications for future research, appear in Chapter XIII.
Chapter II. Reading as operant behaviour

1. The nature of reading

Before looking in detail at attempted modifications of reading behaviour or skill, it is appropriate to discuss how reading itself may be conceptualized. For the purpose of any therapeutic intervention it is first necessary to describe the behaviour in question as clearly as possible. As will be demonstrated, the analysis of the nature of reading raises various theoretical questions in addition to those of practice.

Reading is clearly a symbolic and cognitive skill which is dependent upon the prior acquisition of speech. It is a non-universal yet unusually grammatical form of verbal behaviour, which is not expected to develop spontaneously in the absence of formal training (as language often is). Reduced to essentials, reading involves the control of language by visual stimulation. In the presence of speech or writing, an appropriate linguistic response, which need not be vocalized, is made by the reader.

At present it is fashionable amongst linguists to blur the distinctions between reading and other forms of verbal behaviour, for instance by denying the representational function of orthography in favour of an analysis in terms of its independent abstract structure. There is a tendency to focus interest upon the performance
of the mature, skilled reader, rather than upon that of
the beginner (Williams 1970). This movement, while of
theoretical interest to particular persons, promises
little for those concerned with the social and educational
problems of reading instruction. Indeed, it appears
plausible that the earlier "grapheme-phoneme correspondence"
approach to reading has fallen into disfavour not because
it had yielded up its riches, but because of the lure of
novelty in competing fields of research.

The maintenance of a distinction between reading and
other varieties of language enables reading to be discussed
with clarity while the controversy over language is fought
out separately. Such a viewpoint does not of course deny,
for instance, the importance of comprehension to skill in
reading, but does seek to capitalize on the fact that the
beginning reader has by and large mastered his language
system before reading instruction begins. Hence a
repertoire of speech and syntactical skills, and a
substantial and expanding vocabulary, predate the control
of such language by print which we label as reading. It
appears parsimonious therefore to explore what is distinctive
in the reading process.

Reading is in fact characterized by the manner in which
existing verbal behaviour comes under the control of
discriminative visual stimuli. For the skilled reader
these stimuli constitute a sample merely of the available visual array; the eyes move with long saccades, and inferences about intervening content are made strategically, based on the reader's experience of the redundancy of English spelling, grammar, and the forms of expression common to a given subject. The beginning reader on the other hand, like the skilled reader faced with material unfamiliar in style or content, is obliged to adopt a much more thorough sampling technique, even to the extent of proceeding letter by letter.

The novice is severely disadvantaged in this procedure by some fortuitous deficiencies in the orthographic system. Firstly, the inconstant correspondence between letters and sounds in English, although it may well in principle be describable by sufficiently esoteric rules, in practice renders the correct response to a given visual stimulus unpredictable. Additionally, non-correspondence of written and spoken English derives from the failure of the printed version to reliably indicate syllabification (e.g. mis-led versus misle-d), grammatical usage (e.g. lead as verb versus noun), or stress pattern. Because of these irregularities, "what is learned" in learning to read can be an association between visual stimuli and linguistic responses of variable sizes (see Section II, 1. and 3. in this chapter). The past experiences and accumulated knowledge of the reader will in every case influence the exact content of the connection formed. It may also be
noted here that direct modelling of the appropriate response will often be the only feasible method of facilitating the association.

Whether the stimulus is a letter or a phrase, it acquires with the learner's progress the power to control the performance of a particular response. The theoretical ideal orthographic system would allow perfect pronunciation of entirely novel words; in practice for the reasons discussed above the probability of correct rendition of an unfamiliar word will depend on the reader's general fluency, and on chance. The point remains that understanding of, or belief in the accuracy of the printed material is not a prerequisite for successful reading performance. (In fact the tutor will usually explain the meaning of novel words with the dual purpose of expediting retention of pronunciation, and of enlarging the reader's vocabulary.)

Skinner (1957, p.66) referred to the performance of linguistic responses under the control of discriminative visual stimuli ($S^D$s) as texting. In the acquisition of reading skill, two kinds of "performance" need to be distinguished:

a) the emission of verbal responses appropriate to the visual stimulus (i.e. discriminative control, as a result of associative learning), and

b) the learner's effort and attention to continue the training process. It will be shown that it is this latter, "motivational" aspect of reading which forces the instructor
to consider what are the effective reinforcers for success in the learning-to-read situation. As will be discussed more fully later, the visual associative learning process is a protracted and demanding one; definition of the reinforcement for skill-development is of corresponding importance.

Although Skinner's hypotheses in *Verbal Behaviour* (1957) have been mordantly criticised by Chomsky (1959) and as vigorously defended by Mac Corquodale (1969, 1970), it has been left to others to amplify the operant analysis of reading. Goldiamond and Dryud (1966, 1968), Hively (1966) and Staats and Staats (1963) have extended Skinner's arguments with regard to reading.

Goldiamond and Dryud (1966) include the notion of "behavioural control" as an aspect of reading. Whereas educational and lay writers stress the importance to reading skill of the reader's comprehension of what is read, texting is independent of the speaker's motivations such that the verbal responses made, need not make sense at all to the performer; in principle one could "read" aloud in any foreign language which used the Latin alphabet. Evaluation of the reader's understanding, i.e. the relation between the content of the text and his future behaviour, poses a problem from any theoretical viewpoint and becomes part of the overall issue of how to measure skill in reading (see II.4, IV.1).
II. Difficulties in the operant analysis of reading

1. Specification of the discriminative stimulus

What researchers in remedial reading have primarily aimed to do is presumably to increase the control of the discriminative stimulus ($S^D$), so that a given verbal response occurs only when the occasion has been set for it. The effect of this would be to decrease the number of errors and thus increase the accuracy of reading. The competence acquired could in principle be measured by a simple formula (Millenson, 1967, p.215) comparing rate under $S^D$ with the sum of rates under $S^D$ and under other stimuli ($S^A$). In practice, however, this is an unworkable ideal, due to the non-occurrence of the convenient but artificial laboratory situation where one discriminative response is learned to a given particular stimulus.

In the early stages of learning to read by a phonic method, the visual stimulus 'g' might come to control verbal responses of [g] or [j] depending on the following letter (i.e. whether e or i or not). Supposedly the stimulus unit has become more complex (e.g. "ight") when the correct response is controlled by bright, and more specific still when through, thought, though, bough, dough and rough produce correct responses. In these stimuli only the whole word is sufficient information for the reader to respond appropriately to 'g' or even to 'ough'.
2. **Specification of reinforcers**

"Indeed, textual behaviour is so strongly reinforced that one is likely to find oneself reading not only letters, books, and newspapers, but unimportant labels on packages, subway advertisements, and billboards."

(Skinner, 1957, p.66).

Several sources of reinforcement maintain, and presumably support increasing competence in, textual behaviour. In the case of children, the educational system, within which a large proportion of waking time is spent, is committed to the premise that reading, and the complementary skill of writing, are not only desirable but essential behaviours for the acquisition, retention and display of knowledge. The child's responsiveness to this pressure will be partly determined by how well he has been trained by his parents and subculture to value education. The responses of peers and later of potential employers, will be influenced by educational achievement, in turn directly dependent on reading competence.

Concurrent with but distinguishable from this official encouragement of reading skill, there is the intrinsic reinforcement consequent upon reading: the gaining of information or, in the case of fiction, perhaps the pleasure of mild degrees of emotional arousal due to uncertainty and unpredictability of the stimulus. This intrinsic or "automatic" (Skinner, 1957, p.66) reinforcement of reading is
necessarily private, idiosyncratic, and variable in intensity. How effectively a given passage reinforces a given reader in this way, such that he may read it more quickly, frequently, intently or with greater "interest and enjoyment", will depend on his past reinforcement history in a way which is assumed to be lawful but in the human case, cannot generally be predicted. The contribution of intrinsic reinforcers cannot be specified in advance in individual cases, and is thus vulnerable to confusion with the effects of extrinsic reinforcement, particularly when, for example, extrinsic reinforcers have been presented on an intermittent schedule (see Chapter XII).

3. Specification of learning effects

Providing reinforcement for reading behaviour has often been expected to "increase reading skill", in operant-based research such as the studies reviewed in Chapter IV. A double outcome is in fact predicted; that the more reliable stimulus control of text reading responses achieved by training will not only increase their accuracy, but also "strengthen" reading behaviour in the sense of increasing its probability. Unlike the pigeon learning to peck different keys under different conditions of illumination, human subjects have many (indeed continual) opportunities to engage in reading outside the training situation. How much generalization occurs, and whether more time is in fact spent in voluntary reading during or after participation in a training programme, is just not known.
No relevant data are available, probably due to the monumental difficulties of observation and measurement of the response.

Thus texting is behaviour of variable form and amplitude, maintained by a diversity of extrinsic and intrinsic reinforcers, and controlled after appropriate chaining by $S^D$s of similarly variable form.

Athey (1971) has summarized practical contributions of the behaviourist analysis of reading in terms of the importance of (1) knowing the individual's past reinforcement history so that the most effective incentives can be provided for a given person or group, and (2) knowing how to order and rationalize presentation of the controlling stimuli, which in turn necessitates an awareness on the teacher's part of the child's abilities and past achievements.

Athey believes (1971, p.98) that "optimum scheduling of reinforcement for rapid learning and for retention" needs to be investigated, as does comprehension "the crux of both spoken and written language". (The emphasis on comprehension as an integral aspect of reading appears to data from Thorndike's 1917 paper "Reading as Reasoning"; see Otto, 1971, and Stauffer, 1971).

Athey's last comment raises an important practical point. However operant theorists may choose to consider or not consider "meaning", some evaluation of the degree of the understanding achieved by the reader is necessary in assessing
the growth or improvement of reading skill. The problems involved in measurement of the reading response are elaborated in Section IV.1. of this chapter.

III. Operant analysis of reading failure

1. Reasons for failure

Reading involves not only discriminating between visual symbols, but also associating the correct verbal response to each (Bandura, 1969). The latter process is made much harder by the variable relationship between visual stimulus and verbal response, in nonphonetic languages such as English. Staats and Staats (1962) considered various reasons for the relative difficulty experienced in learning to read compared with learning to speak. Firstly, training in reading has a sudden and intense onset, and "escape" behaviours are likely to be strengthened. Secondly, particularly where the child has not been trained to find the approval of the teacher and evidence of academic progress rewarding, the reinforcers for reading are likely to be weak. Further, such reinforcers as do occur are likely to be too late, too infrequent, or not contingent on the correct reading response. Yates (1970) observed that a third possible explanation is that the child can encounter in natural situations no modelling of reading, as he continually does for speech. It is conceivable, further, that some valued reinforcers may be lost as reading competence is gained, in that the pleasant, relaxed, often
intimate setting of being passively read to by parents and others is likely to decline in frequency.

Experiments such as those of Zimmerman & Zimmerman (1962) and Hasazi & Hasazi (1972) described in Chapter III, highlight the role of adult attention in maintaining apparent learning deficits. The academically disabled child may be modelling himself on a similarly afflicted parent, or may be using his lack of educational progress as an aversive reinforcer for his parents, or to demonstrate that parental and perhaps social-authority values have been rejected and do not control his behaviour; many examples are possible of positive consequences contingent for the child upon "not learning".

The majority of children do learn to read during their first five years at school. In seeking to discover whether the teaching process could be streamlined and increased in predictability, cases of its evident failure (i.e. of reading retardation) may provide insight into the functions of reinforcement.

2. The role of punishment

Where an aversive stimulus is presented following an originally neutral one, a conditioned emotional response (CER) often regarded as the experimental analogue of anxiety, is produced. The result is a conditioned suppression of responding during the presentation of the warning signal (Estes & Skinner, 1941), with correlated autonomic and skeletal (respondent) changes (Millenson, 1967, p.441 ff.; Church, 1972,
p. 712). If print should become a signal for punishment, whether due to early failure by a poorly-taught, sickly or unintelligent child, or to over-sensitivity on the child's part to correction by the teacher, anxiety and a decrease in reading responses could be predicted. A "vicious circle" might then occur, with successful escape and avoidance behaviours on the child's part effectively removing him from opportunities to make good the original deficiency in learning. Non-reading behaviours during the reading lesson could range from day-dreaming and doodling, to disrupting the class or truanting from school, and would be expected to generalize, ensuring the absence of out-of-school reading and perhaps the presence of negative attitudes towards other aspects of education.

Remedial help in such a case would necessarily involve deconditioning the CER to print, for example by replacing the punishment signalled by print with appetitive stimulation.

Theories about the extinction of fear responses rely mainly on inhibition, counterconditioning (by the learning of some antagonistic response such as relaxation), or some combination of these. "Usually there is no basis for selecting one of these alternative explanations over the other", if we believe McAllister & McAllister (1971, p.159). The effects of punishment are extremely complex, depending on the reinforcement schedule prior to and during punishment, the availability of alternative responses, and the intensity,
immediacy, frequency and persistence of the punishing stimulus (Dinsmoor, 1970, p.41). However, in the print-avoidance case discussed above, the effects of further aversive consequences for incorrect reading might be hypothesized to be counter-productive of better reading.

Some evidence to support this hypothesis is available from the experimental remediation of reading retardation by token reinforcement (see Chapter IV). As will be seen there, however, the effects of aversive consequences for errors have not been systematically studied.

3. Material Reinforcement

If punishment is eliminated from training sessions, the further presentation of textual stimuli, perhaps preferably in an individual or small group situation to ensure the child's attention, should extinguish the CER and permit learning to proceed. The punishment may, however, be self-mediated (e.g. "feelings of failure") as well as extrinsic (physical or verbal abuse or criticism). For some children, success alone will act as strong positive reinforcement for effort; such children may need no further reward than clearly-presented informational feedback about progress. Others may need different incentives, such as the teacher's praise, stars, candy, or evidence of parental delight, corresponding to whichever primary and secondary reinforcers characteristically influence their behaviour.
At this point it seems appropriate to distinguish clearly the discriminative control, from the contingencies which may be in operation. Cameron et al. (1972) describe a highly individualized teaching method, more efficient than the conventional classroom technique due to its greater sensitivity to personal needs (e.g. for extra instruction on points of ignorance). Such a programme like many others, provides a logical succession of learning opportunities, in a chain of manageable accretions of knowledge. "The major brunt of change is borne by the programme itself. The consequences are used to maintain the behaviour through the programme" (Goldiamond & Dryud, 1968, p.73).

If behaviour is maintained therefore, some positive consequence must be assumed, and in the absence of a superimposed consequence such as money or candy or even systematic praise, it must be assumed to be (a) intrinsic to the activity, and/or (b) an invisible conditioned reinforcer of the kind sometimes described as "self-control". In discussing Cameron's paper, Ferster (1972, p.201) explicitly raises these questions:

"For whose benefit is the child reading? ....
Is the competence in reading a derived reinforcement? Or conversely, is it possible that the reading is a reinforcer independent of what it makes possible outside the classroom?"
The match between subject and optimal reinforcer is both important and in practice, determinable only empirically. One may suspect that for some children material reinforcement increases motivation too far, such that in line with the Yerkes-Dodson Law, performance is disorganized. For others, social and symbolic consequences have minimal behavioural control until specifically conditioned to do so.

The next chapter will show that present understanding of token effects does not specify which aspects of the reinforcement are necessary (rather than sufficient) to modify the cognitive behaviours summarized as "reading".

4. Generalization of effects over time

While the aim of therapeutic intervention is generally to change behaviour permanently, there are sadly few follow-up investigations in the published literature. According to the present analysis, the maintenance of reading improvement would presumably depend on a continued non-punitive reading environment, which may be more likely after staff-participation in "research" or after a commitment to remedial action any way.

IV. Some difficulties of reading experimentation

1. Measurement of the behaviour

As previously noted, reading may be an involuntary and momentary response to any form of the print which is almost ubiquitous in an urban environment. The investigator is forced to exclude such instances of the behaviour, since he
has no way of observing them without attaching a sensitive camera to record eye-movements, to his subject's eyeballs continuously throughout the duration of the experiment. The varying amplitude and possibly entirely covert nature of reading will always present problems; the fact that the subject sits with his eyes on print and regularly turns pages does not necessarily imply reading.

Externalizing the response by asking the subject to read aloud is only a partial solution. The adequacy of word recognition and expressive fluency can now be assessed, but as with silent reading, a further dimension of the process must in some way be estimated if any overall measurement of the skill is required. This dimension is the subject's comprehension of what is read or the extent to which sufficient mechanical fluency has been acquired to allow the reader to respond (with an experience of "understanding") in roughly the way intended by the originator of the read material. A person struggling to read aloud material which is excessively complicated in form or esoteric in vocabulary, can successfully answer few questions about content afterwards.

The desirability of including some assessment of "comprehension" in the measurement of reading behaviour means that reliance on rate of correct responses alone as a dependent variable is unsatisfactory. Reading is not, in other words, like typing, where the only criterion of success is number of
accurate responses per minute; apart from anything else, there are limitations on rate of reading aloud imposed secondarily by the agility of the human vocal musculature, but primarily by the information-processing and auditory-discrimination talents of the listener. A limited assessment of comprehension can be gained by asking the reader questions about what he has been reading, taking into account the role of memory load. Such data are occasionally mentioned in the reading-training literature, but rarely reported quantitatively.

As has been pointed out, attempts to measure the natural frequency of reading behaviour are doomed to large measures of frustration, and to the writer's knowledge have never been successfully made. The usual solution is to restrict the class of reading behaviours in which one is interested, for example to "those reading responses which occur during a remedial training session of stated length and frequency."

Within this boundary, two main techniques of measurement are available, which coincide with the two aspects of reading with which behaviour modifiers are naturally concerned, namely the level of skill from moment to moment or at least lesson to lesson, and some gauge of increasing competence over a series of lessons. These measures are (a) rate within sessions, and (b) pre-post changes in scores on standardized tests.
(a) Rate of correct responses (i.e., strength of SD control). As "percent of total time spent reading" is an unsatisfactory measure for reasons already discussed, a count (or estimate, as in the work of Staats, 1965, 1967, 1968, 1969, 1970) of words read per session has enjoyed some popularity, along with variations such as words read correctly in unit time, number of correct frames of programmed material, or progress through material of graduated increasing difficulty with a constant error level.

In intra-subject replication designs, continuous data-gathering is a necessity if treatment effects are to be compared with each other or with baseline conditions. Session-by-session rate changes or distances progressed through tasks of known and uniformly increasing difficulty, seem to be the best continuous measures available. The experimenter must decide, however, whether to present a homogenous learning task, and run the risk of boredom or irrelevance for particular subjects, or whether to vary the tasks within sessions, (e.g., to include spelling, exercises and comprehension questions as well as reading aloud), and then perhaps find that only part of the data generated in each session can be meaningfully used. If the training is at all successful, he must somehow deal with the problem of drifting baselines.

(b) Standardized tests. To discover whether reading skill has in fact been affected by different treatments,
researchers have compared subjects in matched-group designs on pre- and post-treatment tests of reading. Samples of the training material provide information of limited generality, so "objective" samples of the word-universe, in the form of standardized tests, are often used. Such tests always include word-recognition, and often other aspects of reading as well. They are based on the assumption that some words are more difficult than others, and that persons who can read them successfully are correspondingly better readers than those who cannot. "Difficulty" of words usually implies increasingly infrequency of usage, and often increasing length, rather than increasing phonic irregularity alone (e.g. "ceiling" is regarded as easier than "enigma").

That there is a continuum of reading skill is very clear; the expectation that relative skill can be assessed by performance on a fixed sample of words is more questionable. Cultural and other environmental variables will determine the words with which a given individual will be familiar, in written as well as in spoken form. Within a relatively homogenous group of subjects, however, it seems not unreasonable to expect that if groups are matched for pre-treatment scores, differences in post-treatment scores (or in magnitudes of individual change) reflect differing treatment effects.

To further complicate matters, the "standardization" of available tests is often inadequate or outdated. The
"diagnostic" tests, such as the Dominion Individual Test of Word-analysis Skills (Ontario College of Education, 1947), Neale's Analysis of Reading Ability (Neale, 1958), Daniels & Diack's (1960) Standard Reading Tests, and Gates & McKillop's Reading Diagnostic Tests (1962) sample a wide range of constituent abilities but rely extensively on the teacher-administrator's experience to pinpoint relative weaknesses in performance. The "attainment" tests on the other hand, often designed for group administration or individual screening, sacrifice breadth of information obtained for rather dubious norms. Validity and reliability can often both be queried, not only for current local use but as applied to the original standardization population. The Australian Council for Educational Research, for example, derives average reading scores for South Australian children from results of a survey taken in 1946.

The best tactics are then to discover, if possible, a test which, even if coincidentally, has some validity for the potential-subject population, then to define relevant parameters of the subjects such as socio-economic status and school attended, and then to use the standard tests to compare each subject with his own record as far as possible.

On the question of measurement as a whole, the optimal strategy seems to be the employment of a selection of measures of different aspects of skill acquisition, including
for example, error rates on material of increasing difficulty, fluency, and comprehension of subject matter. It is forlorn to expect any single measure to reflect treatment effects or non-effects adequately.

2. Research designs

Conventionally, operant researchers prefer not to give groups of subjects different treatments and then test for statistically reliable differences, e.g. by analysis of variance. Sidman (1960, p. 85) states: "When an organism's behaviour can repeatedly be manipulated in a quantitatively consistent fashion, the phenomenon in question is a real one and the experimenter has relevant variables well under control."

Behaviour-therapists in general have been equally sympathetic to the single-organism study where behavioural control by the experimenter or therapist can be clearly demonstrated. Shapiro (1961, 1966) complaining of the group-centred, indirect nature of much clinical research, points out that "N = 1" can, if replicated, illustrate general laws and, if predictions are made beforehand, demonstrate behavioural control. Chassan (1967) in discussing the "intensive design", points out that, with one subject, the researcher can specify the context of any significant effect found. The whole validity of behaviour therapy lies in showing that "changes in [the patient's] behaviour are lawfully related to the experimental operations which were intended to produce them" (Yates, 1970, p. 381). Yates argues for the "base line/change/
reversal-to-baseline paradigm", and utilization of the natural history of the disorder as generally known and in the particular sufferer, to help evaluate the significance of change in the single case. Yates does acknowledge the difficulty of this design, namely, that therapeutic changes are intended to be permanent, or in Bandura's (1969, p.244) words: "In many instances the original baseline is not recoverable."

For several reasons recovery of initial baseline is an unrealistic expectation when reading training is discontinued as the stimulus control developed is intended to be irreversible. Although material reinforcement may be discontinued, the experimenter will have been attempting to strengthen the reinforcing power of social events such as praise and attention by pairing them with the tokens or candy, etc. If successful, this technique will remove the need for material reinforcement by replacing it with a conditioned reinforcer likely to be a normal feature of the environment. A further explanation of why reading should not return to baseline after the treatment (as it did not in the study by Wolf et al., 1968), is that once a basic competence in the mechanics of reading has been acquired, what Skinner calls automatic reinforcement, or the intrinsic consequences of reading, will support the behaviour (although not differentially reinforcing accuracy of reading).
Kazdin & Bootzin (1972) noted in a recent review that the ABAB design often employed is incompatible with the demonstration of long-term resistance to extinction of treatment effects.

After a comprehensive discussion of the difficulties of interpretation with single-organism reversal designs (intra-subject replications) Bandura concluded that:

"the implicit assumptions that repetitive control does not alter the modifiability of the behaviour in question, that behaviour at different levels is equally modifiable, and that reinforcement operations are unaffected by contrast in incentive conditions are all probably untenable. The relative potency of different controlling variables can, therefore, be best assessed through experimental designs involving matched groups." (p.244).

Thus, there is a place in behaviour modification research for both kinds of design (single-case and matched groups), according to the goal of the experimenter in particular situations; whether he wishes to demonstrate "understanding" of an antecedent-consequent relationship by producing it at will, or whether he wishes to contrast the operations of different variables, especially where irreversible learning is expected and desired.

Research methodologies were evaluated in Gelfand & Hartmann's (1968) review of child behaviour therapy studies.
They suggest the following criteria for acceptable small-sample research: adequate baseline measures, detailed descriptions of treatment procedures (to ensure replicability), continuous monitoring of behaviour, systematic variations of reinforcement contingencies, and rigorous follow-up observations.

Specific guides for determining response-rate in educational situations, methods of representing data graphically, and making appropriate treatment decisions on that evidence, are provided by Lovitt, Kunzelmann, Nolen & Hulten (1966). Journals such as the Journal of Applied Behaviour Analysis demonstrate the continuing popularity of result-graphs amongst operant experimenters.

A recent paper by Gentile, Roden & Klein (1972) offers a rationale and methodology for interpreting the data produced by one or a small number of subjects in reversal designs in analysis of variance terms. The application of this sensitive and flexible statistical device can help the experimenter decide, particularly in situations "such as classrooms" (p.193) where precise control of variables is difficult, whether smallish changes in the expected direction, or changed performance in only some of the subjects, do in fact represent experimental control. Perhaps the availability of this technique refutes Skinner's (1961) charge that conventional statistics cannot cope with direct manipulation of variables, but only with post hoc variation.
Summary

Despite difficulties in definition and observation, reading can be fitted into the operant framework without excessive strain if basic behaviourist premises are accepted. The main theoretical problem (and one not confined to the operant approach) is how to take account of "meaning". How well the reader understands what he reads must constitute one measure of reading ability, and is also likely to contribute to the inherent, intrinsic or "automatic" reinforcers which support the behaviour. These latter cannot as yet be satisfactorily disentangled from the more obvious extrinsic rewards for learning to perform the behaviour, which are provided by teachers and parents.

Refusal to read may have its own positive consequences, especially perhaps after early failures. The possible role of associated punishment was analysed in terms of conditioned anxiety to the visual stimulus of print. Remedial efforts would appear to be required to extinguish this maladaptive response. Whether unreinforced trials alone or extra appetitive stimulation is necessary, is an empirical question and the answer may depend on the reinforcement history of the individual involved. It is necessary in any case to distinguish between the acquisition of discriminative control (such that accurate reading becomes possible), and the willingness to expose oneself to a remedial programme
and to expend effort to take advantage of the opportunity for learning.

Measurement of reading skill is, like the choice of experimental design where permanent change is desired, a practical problem for the researcher and a multi-faceted approach has been suggested.
Chapter III. Reinforcers suitable in Educational Settings

Behaviour modification within normal classrooms has recently enjoyed a remarkable expansion, despite "the truly formidable methodological problems involved" (Yates, 1971, p.161). This chapter looks at some of the methods used and the queries raised about them from ethical, theoretical and practical points of view.

The application of operant principles to the classroom is not entirely new. Skinner has continuously advocated the use of teaching machines and programmed instruction, acknowledging his debt to the pioneering efforts in the 1920's, of S.L. Pressey (e.g. Skinner, 1961, Part III). More recently, concern has shifted from how to present materials to ensure maximally efficient learning, to the question of how to promote task-relevant behaviour in the classroom.

Lovitt (1970a) characterizes the "behaviour modifier" in the school as one who relies on the individualized application of the principles of direct observation, continuous measurement, and systematic manipulation. Future trends are seen (1970b) as (1) use of the techniques by an increasing number and variety of persons charged with the welfare and education of children; (2) an increasing appreciation of the need to separate measurement from manipulation (in order to establish norms); (3) a concentration on academic performance rather than the
maintenance of order and discipline; (4) growing readiness to measure many aspects of behaviour simultaneously and investigate the problem of generalization of treatment effects; (5) parametric or component analyses of variables responsible for behavioural changes; and lastly (6) interest in training children to record and observe their own performance, thus laying the foundations of self-management. Lovitt expresses the hope that children could be trained to design their own learning programmes and even their own contingency systems, and he expects that these procedures would increase the probability of producing "unique, independent, motivated students" (p.157).

In concluding his review of the area, Forness (1970) points out the continuing responsibility of the teacher, in the light of demonstrated potency of behaviouristic techniques, to decide the objectives of the programme and evaluate not only direct results but possible side-effects.

It appears that behaviour modification has been accepted as a powerful yet respectable educational aid (see e.g. O'Leary and O'Leary, 1972). Estimates of how widely the techniques are used by individual teachers are difficult to make, especially in Australia where, although there is a groundswell of educational opinion in favour of open units and more self-direction by pupils, opportunities for formal training of teachers in behaviour modification techniques are limited.
Upon what kind of evidence then, is this acceptance of behaviour modification based? Classroom behaviours (such as disruptiveness) which affect learning performance indirectly, and also academic behaviours other than reading are discussed here according to the kind of reinforcers found effective: primary, social, "natural", and token. Most of the contingencies commonly applied to children by behaviour therapists have also been used in the classroom. Therapists have either worked directly with the child, or, as is increasingly the case, by advising and manipulating the behaviour of teachers and parents (e.g. Christopherson & Arnold, 1971; Hall, Axelrod, Tyler, Grief, Jones & Robertson, 1971, 1972; Mira, 1970; Patterson, McNeal, Hawkins & Phelps, 1967; Smith, 1972; Wahler & Erickson, 1969).

1. Primary reinforcers

Bijou & Sturges (1959) concluded from a review of data available at that time, that the optimal reinforcers for use with children were consumables (if parental consent could be obtained and satiation avoided), and manipulatables, such as access to, or award of, toys and trinkets. Long, Hammack, May & Campbell (1958) investigated 4-8 year old children's performance of simple motor tasks, with trinkets, candy and slide-changes as official reinforcers, but with the approval of the experimenter often becoming in fact the effective
reinforcer. They concluded that despite short sessions and low levels of deprivation, intermittent schedules of reinforcement had similar effects on the children's behaviour as on that of lower animals.

One of the earliest attempts at educational instruction based strictly on operant principles, was that of Hewett (1964). Over a year and a half, his non-speaking 13 year old autistic subject developed skills progressively to the point of communicating his wishes in writing. Gumdrops acted as the reinforcer; at first they were contingent upon attention and cooperation in elementary matching tasks, and eventually upon writing a whole phrase. The major aim of the programme, a breakdown in the self-absorbed isolation of the child, was achieved simultaneously with the development of rudimentary reading and writing ability.

Greenberg & O'Donnell (1972) combined individual and group contingencies to control the highly disruptive class-behaviour of an isolated 6 year old boy. Tantrums were followed by removal to a cloakroom, and 1½ hour periods of no-tantrum by candy for all the children in the class. Tantrum rate decreased from 6 a day to 1-2 per week, peer relationships improved and school-work began to be done, the latter two effects disappearing when the group reinforcement procedure was discontinued.
Noting that most of his preschool subjects arrived without having had breakfast, Risley (1968) successfully utilized food as a reinforcer for desirable pre-academic learning (such as politeness, motor and verbal imitation of adult models and verbal fluency). The teachers dispensed a variety of desirable snacks, such as cookies, grapes, candies, and warm soup, invariably accompanying them with "smiling, warm-voiced statements of approval, appreciation, or affection" (p. 30), in order to boost the children's susceptibility to social reinforcers, the most common kind in the classroom. A similar association of candy and praise was instituted by Drass & Jones (1971) in their programme to use reading-disabled teenagers as tutors for younger sufferers. Many of the token studies to be reviewed later, also stress the important opportunity for strengthening social reinforcers, which is provided by primary rewards.

2. Social reinforcers

Methodological problems in assessing the impact of social reinforcement on children's motor behaviour have been reviewed by Parton & Ross (1965); Hill & Moely (1969) explored the effect of task instructions, and age, sex and baseline performance of the subject; and Leventhal & Fischer (1970) investigated non-verbal social reinforcers. Training in systematic use of their attention and
praise was given by Ward & Baker (1968) to teachers of four behaviour problem 1st graders. During 5 weeks of this treatment, the rate of disruptive behaviour dropped from 74% to 57%, while matched controls in the same classes increased slightly in disruptiveness (37% to 41%). No adverse ("symptom-substitution" type) effects nor generalized improvement on psychological tests were found, for either target or control children, and it was noted that the programme had greater success with the acting-out children than with one withdrawn and inattentive child who was largely unaffected.

In a classroom of well-behaved middle-primary level children, Thomas, Becker & Armstrong (1968) found that the teacher's approval (shown by smiles, praise and contact) maintained appropriate classroom behaviours, but that when the teacher ceased this behaviour, or increased her signs of disapproval (hitting, scolding and threats), disruptiveness by the children, especially noise and gross motor activity, was increased. There was evidence that "peer reinforcement (among other stimuli) takes over when social reinforcement is not provided by teacher" (p.45).

Three inexperienced teachers, with precarious or no control over their classes, were trained by Hall, Panyon, Rabon & Broden (1968) to make their attention contingent on study behaviour, with resultant increases in their efficiency.
Cormier (1970), however, found that the teacher's social reinforcement (praise and attention) was effective in increasing task-relevant behaviours and decreasing inappropriate behaviours in both target and non-target adolescents, and whether or not the reinforcement was contingent on desired behaviour.

3. "Natural" reinforcers

Cantrell, Cantrell, Huddleston & Wooldridge (1969) used contingency contracts between behaviour problem children and their teachers or parents or both. The natural reinforcers of the child's environment were made contingent on specified adaptive behaviours (or approximations to these); the authors note that the use of reinforcers already inherent in the situation, rather than the introduction of extrinsic rewards, appealed to the natural contingency managers. Home-based reinforcers were also used by Bailey, Wolf & Phillips (1970) to modify the classroom behaviour of five 11-15 year old pre-delinquent boys. A Yes/No mark on appropriateness of behaviour in school was supplied each day by the teacher, and it was found that once study behaviour was instituted, the daily feedback and reinforcement system could safely be faded out.

Further examples of the use of inherent or natural reinforcers are provided by the work of Osbourne (1969) who successfully reduced out-of-seat behaviour in 6 deaf girls by making free-time contingent on remaining seated; Wasik
(1970) who used access to a display of toys (called free-choice activity time) to control appropriate behaviour in a second grade classroom; and Wasik & Simmons (1971) who increased the staying-in-centre (and therefore attending and learning) behaviour of first-graders by contingent use of a maximum of 30 minutes of special (art and craft) activities per day.

All of these studies are based on Premack's (1965) generalization on reinforcement: that "For any pair of responses, the more probable one will reinforce the less probable one" (p.132). Thus preferred activities, i.e. those which have high likelihood of occurrence if a choice is allowed, can be used to reinforce less favoured activities. Unfortunately, follow-ups are often insufficient to make clear whether the originally less-preferred activity develops a higher probability of its own following the procedure; this could be expected perhaps in cases such as academic performance where the expenditure of initial effort might increase achievement and thus the intrinsic reinforcing properties of the activity. Also unknown is the exact probability differential which is desirable between the reinforcing and the unfavoured activity; presumably it would not be impossible for the initially highly-probable response to decrease in natural frequency if the relative aversiveness of the required antecedent behaviour was too great.
Medland & Stachnik (1972) used extra recess and free time as group consequences, in the "Good-behaviour game" applied to a disruptive reading class. Undesirable behaviours were reduced by 99% for one group and 97% for the other. Upon removal of the game contingencies, disruptive behaviours failed to return to baseline in some subjects. The authors attribute this to control having been acquired by "coincidental" or extra-experimental reinforcers.

4. Token reinforcers

In 1966 Bijou & Baer described the use of "generalized" reinforcers (e.g. a coin which would operate an array of vending machines), and how to stretch the response-reinforcement ratio by signalling the correctness of non-reinforced responses.

Token reinforcement systems fill the requirements for a method of bridging the temporal gap between performance and reward. As noted by Bandura (1969), their reinforcing value is relatively independent of momentary deprivalional states, they are not subject to satiation effects, can be easily and quickly presented, and can be exchanged for a variety of items which particular individuals find attractive.

With these advantages, it is not surprising to find how widely tokens have been employed by experimenters dealing with both adults and children above a certain mental age, in long-term therapeutic programmes. In such cases, it is desirable
to reinforce responses under many different conditions of time and place, but without either disrupting the on-going performance, (e.g. by the ingestion of a nutritional reward), or obscuring the response-reinforcement link by temporal lag. The studies by Ayllon & Azrin (1965, 1968), and Atthowe & Krasner (1968) of the shaping and maintenance of desirable behaviour in long-term mental hospital patients, have acted as an inspirational spark for a wide variety of applications of the token reinforcement technique.

An early example of a token system instituted to improve motivation, concentration, perseverance and good study habits, was that by Birnbrauer, Bijou, Wolf & Kidder (1965). Their mentally retarded subjects would not work steadily for social and informational reinforcement only, but after 5 months 7 out of 8 became "good students" despite an average value of backup reinforcement of only 5c per week. Bijou, Birnbrauer, Kidder & Tague (1966) used the token system (of gummed stars and later "marks") with diverse backups, to improve scholastic motivation and study habits in 27 retarded children. They were careful to pair token-award with social events such as a positive comment by the teacher, so that normal secondary reinforcing properties would be acquired by these stimuli. When the token system was discontinued, "some" children continued to progress, some progressed more slowly, and some performed worse, than before. Similarly, deducting marks for incorrect responses led to "some" pupils becoming more accurate
Lovitt & Curtiss (1969) found, with a replicated baseline-A-B-A design, that a 12 year old boy worked faster when he, rather than the teacher, decided the point-value of work in different academic areas. This was found not to be a function of his greater generosity to himself, for when the teacher awarded points in the same proportions as the boy had previously used, his work rate remained slower than when the contingencies were self-managed.

Total time spent working, output rate per hour, and accuracy were all improved over baseline levels by a 6-week combination of tangible backups and social reinforcers, in an experiment by Chadwick & Day (1971). Only 2 weeks after material reinforcers were discontinued however, total time spent working had returned to baseline level, it being too early for the other two indices to demonstrate clearly the effect of continued social reinforcement by itself. As will be seen, the questions of how universal and how permanent are the effects of tokens, occur in study after study.

Response cost. A further convenient attribute is that, by withholding or repossessing tokens for undesired responses the experimenter has a measure of negative control possible which does not involve physical or verbal abuse. Whether or not token-deduction is an aversive event is an empirical question which will depend on the unique characteristics of
the situation. However, if tokens reinforce behaviour positively, token-removal could well be expected to decrease the strength of antecedent behaviours. The usefulness of this would be, as Bijou & Baer (1966) pointed out, that many of the other mild forms of aversive stimulation acceptable in child studies, may even prove equivalent to positive stimulation for some children.

Points were deducted from an earned amount for aggressive statements and for poor grammar (saying "ain't"), by Phillips (1968). His pre-delinquent subjects, in a residential treatment setting, modified their behaviour accordingly. Burchard & Barrera (1972) found few differences between response-cost and timeout consequences for undesirable behaviours such as assault and swearing. They point out that response-cost has the relative advantage of not removing the subject from the opportunity to reform his behaviour.

The comparative effects of verbal feedback, tokens with deduction for errors, and a combination of these conditions, were investigated by Walker & Buckley (1972a). Their subject underwent the above treatments while working on programmed mathematics materials, and her performance could be compared with that of a control subject who received none of the treatments. Error rate was found to be more clearly influenced than was correct response rate, and most dramatically by the combined treatments condition. The authors concluded that information on progress was the most powerful aid to academic learning.
Function of tokens. Several other writers have questioned whether the efficiency of tokens lies solely in their motivating properties (i.e., whether performance of desired responses is strengthened by reinforcement, with tokens intermediary between response and back-up). Krasner's (1971) review of the state of behaviour therapy includes a section on token economies in which he stresses their great usefulness in functioning as a rationale for giving "key environmental figures such as nurses, aides, parents, and teachers" (p.500) training in behavioural observation and in self-awareness regarding their own reinforcement potential.

Ferster (1971) elaborated this point with particular regard to the educational situation. He regarded the institution of a token system as likely to promote closer attention to the child's behaviour by the teacher, which ipso facto will improve the quality and sensitivity of the teaching interaction. Tokens are thus effective because they make behaviour "more visible" (Ferster, 1972, p.6). This view replaces the secondary reinforcing role of tokens (as a signal of later money, food, etc.) with an informational feed-back conception, and was emphasized by Skinner (1968) in his advocacy of programmed learning materials. By forcing a division of the subject matter into small, logical steps, teaching machines, like tokens, allow immediate knowledge of results and a rate of progress tailored to the individual's capacity.
Non-contingency. If tokens functionally reinforce the behaviours upon which they are contingent, noncontingent token-award would be expected to have very different effects, and this has been demonstrated in several studies.

Tokens backed-up by reinforcers inherent in the preschool situation (such as snacks and play activities) were awarded by Miller & Schneider (1970) to teach handwriting. When the reinforcement was made non-contingent upon responding, virtually no responding occurred. Informal, uncontrolled observations suggested that vocabulary, attitude to school, and co-operative play, were all improved by the token system.

Bushell, Wrobel & Michaelis (1968) also found performance decrements when backup reinforcers (in their case, a "special event") were awarded non-contingently to twelve 3-6 year old children. Work behaviours dropped from 67% to 42%, rising to 64% again on the first day of contingency reinstatement and continuing this increase for four days.

Programmed instructional materials, with their great virtue of permitting a daily objective measure (frames correct) of the learner's performance, were used by Rickard, Clements & Willis (1970). Five emotionally-disturbed boys spent 1½ hours per day with successively no tokens, contingent, non-contingent and contingent tokens, working on mathematics materials, and each earning from $10-$15 worth of merchandise altogether. The initial 8 day "baseline" period showed a steady productivity decline in 4 of the 5 subjects. (In a
footnote Rickard records a similar decline lasting 24 days in a later study.) Upon this unstable baseline, the effect of the first introduction of contingent tokens was to increase response-rates; the decline was halted but overall performance was not superior to the baseline condition. When token-award was made noncontingent, productivity decreased in all subjects and became negligible in 3. The re-instatement of token-contingency (with doubled token value per unit) again increased production, although not to the level attained in the first contingent token phase. The extreme variability shown throughout, both between subjects within sessions, and within subjects between sessions even under the same treatment, is a recurrent theme in reports of token reinforcement applied to children.

Variable responses to token systems. Individual variability in response to tokens was again shown in a programme aimed at reducing disruptive behaviour in 7 second-grade children carried out by O'Leary, Becker, Evans & Saudargas (1969). Before initiating the token system, they successively introduced a list of appropriate behaviours which was reviewed daily, an increase in educational structure equivalent to that necessary for a token system to operate, and a regime whereby the teacher systematically praised good behaviour and ignored bad. None of these measures produced significant changes in disruptive behaviour from the original baseline; the combination of all 3, however, eliminated
disruptive behaviour in one child. Of the other 5 subjects, 5 decreased the frequency of disruptive behaviour when tokens were introduced, increased it when tokens were discontinued, and 4 reduced it again upon token re-instatement. A follow-up period demonstrated that the teacher could in 3 cases transfer the control obtained via tokens, to a less expensive and more usual token system of coloured stars and occasional pieces of candy. (The latter system only ever followed the former so its intrinsic effectiveness could not be judged.) No generalization of appropriate behaviour occurred between the afternoon, with token system operating, and the token-free morning. Significant differences are quoted but not overall averages for groups across treatments.

Kuypers, Becker & O'Leary (1969) used the same token procedures and reduced disruptions in 6 disturbed third- and fourth-grade children, from 54% to 27.8% (averages) when the programme was introduced, with a rise to 41.5% on token withdrawal. The teacher, who had had no preliminary instruction, and considered the original level of deviant behaviour to be close to acceptable anyway, was irritated by the observers and ended the study prematurely by refusing to co-operate further. This fact, and an examination of individual graphs showing that only 4 of the subjects had been clearly influenced for the better by the token system, led the authors to conclude that their programme had been only marginally effective. Apart from inadequate preparation
of the teacher, they attribute their "failure" to making tokens contingent upon absolute standards rather than shaping desired levels of performance, and to neglecting the use of differential social reinforcement, particularly where marks were not awarded until well after their antecedent behaviour.

As a further example of the inconsistent results which appear to be common when using token reinforcement systems, Zimmerman, Zimmerman & Russell (1969) contrasted the effects of praise versus praise plus tokens on frequency of disruptive behaviour in 7 retarded boys. Tokens were differentially effective in increasing instruction-following in 4 of the subjects. Of the others, one subject responded appropriately to most instructions throughout the study, one to none, and the seventh improved continuously regardless of treatment.

Martin, Burkholder, Rosenthal, Tharp & Thorne (1968) found that only one out of nine homebound adolescents showed steady progress (in increased academic output and decreased disruptiveness) under a token system which included social reinforcement and a wide choice of backups. More success was obtained when a complex "phase" system, whereby promotion between phases could be earned by carefully specified steps towards desired behaviour, and reinforced by increasingly social and non-concrete reinforcers, was added. For example, in Phase I the child could earn attention, points and canteen backups, praise and a letter to his parents, for attending
1-4 hours of school and behaving well for 1 hour. For regular attendance at a public school, appropriate behaviour and passing grades, the subject earned the right by Phase V to special tutoring and counselling, cash bonuses and special home based backups, after a graduation ceremony from Phase IV which included a banquet, gift and certificate for the subject and a corsage for his mother. The authors see this programme, with its successive approximations to the behavioural goal, as the social rehabilitation equivalent of an academic instructional programme, but with more reliance on modelling, and the deliberate fostering of social skills and independent decision-making.

Kazdin & Bootzin (1972), in an evaluative review of token economies, concluded that the areas most in need of further investigation are (1) programmed generalization of token-maintained effects; (2) control of language and social behaviours by tokens; and (3) the non-response of some subjects to token reinforcement.

Large-scale programmes. It is relatively rare for token studies to be very ambitious in either time span or range of behaviours attacked. One of the most widely-quoted "all inclusive" token programmes aimed at academic achievement, is that by Cohen (1968). In the correctional institution for boys where he and his associates set up the project, Cohen proceeded from the belief that it is necessary for delinquents to acquire the same academic standards and skills as free,
non-delinquent adolescents. Accordingly they designed "a 4-story, 24-hour contingency oriented educational laboratory" (p.31). Academic responses earned points which gained the young offenders material from a Sears-Roebuck catalogue, or entrance to a lounge where there were friends and a jukebox. As their attitudes to education became more positive, the opportunity to learn algebra became a prestige symbol and would effectively control preliminary learning.

Hewett, Taylor & Artuso (1969) carried out a study which is impressive for its time span. Over an entire school year, one class received tangible rewards for attention in class and achievement in reading and arithmetic. Another class (the controls) received any mixture of procedures and incentives which the teacher chose, excepting the use of tangible or token rewards. Two classes got token reinforcement for the first semester and the control procedure for the second, and a further two received the reverse order of treatments. Task attention increased when the experimental condition was introduced but also when, in the other class, it was removed. Reading achievement was not differentially affected by either condition, but arithmetic improved significantly under the experimental (token) condition.

Side-effects: the "ripple effect". Modelling might be hypothesized to be important in studies such as that by
Reppucci & Reiss (1970), and Boudin (1971), which deal with the "ripple effect" or some spread of effect from token systems to non-target class members. Reppucci used a double token system; studying was followed by the award of stars, praise and at least 5 sec. of attention, backed up on a fixed-ratio schedule by favoured activities and privileges. Disruptive behaviour on the other hand led to an equally immediate "zero", given with no special attention or verbal interaction, and backed-up with deprivation of free-time, or social isolation. The rate or disruptive responses was halved over 7 weeks for 3 groups of subjects: disruptive-target, disruptive non-target, and non-disruptive non-target. Rate of task relevant responses showed corresponding increases in all groups, though less clearly in the case of the disruptive non-target boys than with the other two groups. The authors relate their finding to that of Ward & Baker (1968, op.cit.), that the adverse effects on non-target children of directing a special token system to one or a few class-members, of which teachers often express a fear, do not occur. On the contrary, reducing the distraction caused by a highly disruptive child may produce "a major beneficial effect on the other children" (p.744).

A caution about hoping for "free" beneficial side-effects of token programmes, was underlined recently by Ferritor, Buckholdt, Hamblin & Smith (1972). Using third-grade pupils doing mathematics problems, they found that,
while behaviour contingencies reduced disruptions and performance contingencies improved percent correct problems, no generalization occurred: only when reinforcement was contingent on both attention and performance did both improve.

Wagner & Guyer (1971) used teacher's initials as tokens of small backup reinforcers: the behavioural goal was 15 minutes of consecutive concentration on a classroom task. Ratings by the teachers reflected their opinion that "discipline had improved" (p.288), but academic achievement did not. Parallel results were obtained by McLaughlin & Malaby (1972): over a school year, spelling, language, handwriting and mathematics improved under a token economy, but when "quiet behaviour" replaced assignment completion as the target, assignment completion fell off markedly.

It seems likely that some explanation for apparently contrasting experimental conclusions on the ripple effect, may lie in the differing experimental designs (matched groups versus intra-subject reversals). Findings will obviously depend on what (and who) is observed, as well as on the specificity of procedures, with regard to temporal span and task.

Generalization and ethics. The problem of how to achieve generalization of token-effects (across both situations and time) has been faced in a comprehensive review of token systems in the classroom (O'Leary & Drabman, 1971). After giving several suggestions (most of which have already been
commented on here), they conclude that token systems should only be implemented after other less complex and more "natural" reinforcers such as extra recess and privileges have been tried and failed.

Skinner (in Hall, 1972), Bijou (1970) and Ferster (1971, op.cit.) agree that "contrived" reinforcers are appropriate only when the usual reinforcers inherent in the school situation (e.g. approval, privileges, confirmation of progress) have failed to motivate a particular learner. Furthermore, the "natural" reinforcers should replace the contrived as soon as possible, where the latter have been resorted to in an effort to stimulate or initiate performance.

Demands have also been made on ethical grounds for the phasing-out of extrinsic reinforcers in favour of self-control training procedures. Lovibond (1971) feels that control of normal schoolchildren by tokens is "ethically indefensible" and that manipulation of the behaviour of the child by operant means "becomes an abuse of power... if it is not phased out in accordance with the child's developing capacity to institute rationally based control over his own behaviour" (p.178).

Walker & Buckley (1972b) evaluated maintenance strategies following a two-month period in a token-economy classroom. On return to regular classrooms their subjects underwent two months of one of the following conditions: "peer programming" (subject could regularly earn points towards privileges for the whole class); "equating stimulus conditions" (a continuation of
social and token reinforcement and the same work materials); teacher training in behaviour modification techniques; or a control condition of no attempted generalization of the token benefits. The first two strategies but not the third, produced significantly higher rates of appropriate behaviour than did the control condition, during the maintenance period.

Finally, the effects of "extra help" for symptoms of academic backwardness have been experimentally observed by Zimmerman & Zimmerman (1962) on spelling, and Hasazi & Hasazi (1972) on digit reversals. In both cases contingency reversals demonstrated that deviant performance was being differentially maintained by the special attention it attracted. The implications of these findings should be borne in mind when conventional remedial methods are considered (Chapter VI).

Summary

This review of behaviour modification techniques applied to educational settings has focussed on the kinds of reinforcers which are both appropriate and efficacious. While primary and social reinforcers and opportunities for desired happenings (based on Premack's principle) have all been shown to succeed on occasion, reinforcement by tokens has also been widely investigated, though usually recommended only as a "last resort".

Analysis of token studies has revealed the following information:
(a) There is some evidence that response-cost for errors acts as a mildly aversive consequence.

(b) A strong body of opinion exists that tokens owe their effectiveness to the changes in teaching-material, teacher-pupil interaction, and feedback on progress, which are demanded by their implementation. (Without further information on what other feedback on progress was being provided, no clarification of this point is possible. The question is discussed further in Chapter XI.)

(c) Performance falls off when tokens are awarded non-contingently.

(d) Response to token programmes is by no means homogeneous across individuals, or even over time in a given individual. The reasons for this are not clear, as investigation of non-responders has not been adequate.

(e) Evidence is conflicting about whether and how token systems generalize in their effects with regard to non-participating classmembers, non-target behaviours, and token-free time periods (including follow-ups).

(f) Queries have been raised about the ethics of reliance upon token systems for classroom control. The issue of extrinsic versus intrinsic incentives will be discussed more fully in succeeding chapters.

Thus, although token systems are widely but not universally regarded as effective means of controlling both
academic performance and motor behaviour in many children, the questions of how and why they do so, and indeed whether and when they should be employed to do so, are contentious.
Chapter IV. Token reinforcement for reading

It has so far been shown that reading is subject to analysis in operant terms, and that a range of reinforcers has been used to modify behaviour in classrooms. The concern of this chapter is to review the claims of success which have been made for the treatment of reading problems by tokens. The deficiencies and obscurities of these studies are then explored in the light of the theoretical and practical problems already noted.

The attention of behaviour modifiers and therapists was directed relatively late towards academic problems and disabilities. Not until 1962 did two papers indicate the possibilities held by this area of research, and begin what has become a snowballing growth of interest in the field.

I. Early questions

1. Efficacy

In a review of the then small number of studies applying modern learning theory to the behavioural disorders of children, Rachman (1962) in Britain referred to his own attempted remediation of reading problems by operant techniques. He and Loewenstein had used a FR of 6:1; every correct reading response was signalled on a panel of six lights, and a buzzer and sweet were presented when all the lights had been lit. The subject, a 9½ year old boy
who had had previous unsuccessful teaching attempts, learned
235 new words in twenty 20-minute sessions. No data on
difficulty levels, standardized test results or long-term follow-up were available.

In the same year, Staats, Staats, Schutz and Wolf (1962) published in the U.S.A. their report of reinforcing early reading skills (picture matching, word matching, picture-word matching, simple word recognition) in pre-school children. The six four-year olds worked and attended well over eight 40-minute sessions, when reinforced with tokens leading to a "secret surprise" toy. Withdrawal of material reinforcement, and provision of social reinforcement only, were not effective in maintaining work by the children, who indulged in various escape behaviours. The authors pointed out several problems which, despite later studies, continue to tease experimenters. A measure of progress is needed which somehow reflects the strength of the texting behaviour (and presumably, the learner's increasing competence and skill), since counts of new words learned, even if the number should increase from session to session, do not fulfil this need. The question of which reinforcers are most effective is raised, and of what schedules of reinforcement are most effective. Finally, the experimenter has the technical difficulty of avoiding (ideally) contaminations of the response data caused by different
reading-related tasks presented in the course of reading training.

2. Apparatus

Staats and his colleagues began a research programme aimed at solving some of these difficulties. In 1964, Staats, Minke, Finley, Wolf and Brooks reported on an automated reading-training apparatus; three 4-year old children responded by pressing the correct button, in various pre-training discrimination and reading tasks. Reinforcement was continuous and took the form of tokens (marbles) which could be used either to activate a Universal Feeder and gain a penny, trinket or edible, or to accumulate towards selected toys for which 10, 35, 80 or 150 tokens were required, corresponding to their monetary value in cents. All social reinforcement was carefully avoided by the E; one subject ("a behaviour problem at home and ... excessively demanding of adult attention") (p.219), began to emit a variety of inappropriate non-reading responses including striking the E, and was dropped from the experiment after 12 sessions.

The remaining two subjects stayed for 40 sessions of 20 minutes each. One immediately responded to the tokens as strong reinforcers per se, and maintained a very consistent rate. The other showed a great deal more variability, both within and between sessions, and the suggestion of a "scallop" effect after reinforcement.
3. Reinforcement Schedules

Staats et al. decided that although reading behaviour "is actually a lengthy and complex chain of responses rather than a simple unitary response which occurs more or less instantaneously" (p.224), it is nonetheless appropriate to record responses cumulatively like any other free operant. Thus Staats, Finley, Minke and Wolf (1964) provided cumulative records of reading under different reinforcement schedules from four pre-kindergarten boys over thirty 20-minute sessions. Results were "generally as expected on the basis of previous work (e.g. Ferster and Skinner, 1957). However, the differences in rate produced by the different schedules were not as great as are obtainable in more basic studies" (p.146).

George (1970) used a similar reinforcement system and programmed reading materials, with 100 second grade children. He failed to discover statistically significant differences between no reinforcements, and FR or VR reinforcement.

4. Extra-laboratory Settings

Having established to his satisfaction that reading responses could be regarded for experimental purposes as ordinary operant behaviour, Staats' next contribution was to de-automate the procedure and move from 4-year olds in a laboratory to a 14-year old delinquent boy, to whom a
probation officer supplied remedial reading training using the reinforcement principles derived from the earlier work (Staats and Butterfield, 1965). During 4½ months, the boy maintained his rate of progress through the reading materials, which increased in difficulty by 1.4 grade levels. Performance on standardized reading tests rose from the grade 2 level to the 4.3 grade level. Meanwhile, conduct and achievement in other subjects improved at school and the subject received passing grades for the first time in his life. Staats comments on the possibility of a Hawthorne effect contributing to the results of the study, "that is, that the social reinforcement by the E and possible extra-experimental reinforcement" (p.941); but of course the extent of such an effect cannot be evaluated.

Queries about the relative role of social reinforcement were raised by Whitlock (1966). After 23 sessions with token reinforcers exchangeable for a variety of enjoyed activities, her first grade subject was able to join the normal reading class. Previously S had been reluctant to try, and was considered below normal in reading by his teacher. In a follow-up study Whitlock and Bushell (1967) used, without prior explanation to S, a counter which advanced by one digit with a loud click, for each reading response made by a first-grade girl. For the first 6 experimental sessions, the rate of sentences read per minute rose steadily, only to fall again when the counter was turned off for 6 sessions. This decrement
in performance was only temporarily alleviated by returning
the (token) counter advances. When backups (tangible
objects and valued activities) were introduced contingent
upon the counter-reading, rate increased, thereby
demonstrating control by a reversal of conditions.

McKerracher (1967) compared the combination of phonic
training plus 10 minutes of "supportive" therapy per session,
with a token system with loud buzzer for incorrect responses,
and the token system alone, in successive 3-month periods.
He used the same sweet-plus-praise backup for 6 signal lights
"on" as had Rachman (1962, op.cit.), but gradually stretched
the ratio of words read correctly to the signal light (token).
His subject advanced by one month in reading age during the
first treatment, four months during the second, and nine
during the third. In this study the buzzer was discontinued
because the child was reading more fluently, so its aid as a
teaching device is unclear.

Staats' later research took two new directions; he
became interested in extending the use of his techniques by
non-psychologists, and the use of his apparatus to include
all forms of complex learning in children, such as writing,
number concept learning, concept formation, complex
discrimination learning, and speech acquisition.

"The apparatus and reinforcement procedure ... 
has the same advantages for children, and
the study of complex learning, as the Skinner
box has for animals and the study of simple learning." (Staats, 1968, p.50).

Staats, Brewer and Gross (1970) published a monograph report on Staats' "long-extended research project on the cognitive learning of children" (p.1): included is a discussion about experimental controls to which later reference will be made.

Clark, Lachowicz and Wolf (1968) provided remedial training in basic subjects to Neighbourhood Youth Corps employees who were school dropouts. Five girls worked for points on the remedial materials in the mornings and other jobs in the afternoons; their N.Y.C. salary was the backup reinforcement contingent on academic behaviours. (The subjects appear to have increased the usual earnings by about 70c. a week.) Point contingencies were manipulated to ensure extra work in relatively more deficient areas. The subjects showed a median gain of 1.3 years on the California Achievement Test battery, compared with 0.2 years for a matched control group of N.Y.C. employees.

Remedial education for culturally deprived children was also the goal of Wolf, Giles and Hall (1968). Sixteen 5th- and 6th-graders attended daily special classes for a year, being reinforced with different coloured trading stamps for academic achievement in both the remedial and regular classes. Backups were a selection of trips, food and goods. Gains on the Stanford Achievement Test were double (1.5 years) for the experimental as compared with control subjects (0.8 years).
who had had no remedial training. In an attempt to
investigate the role of token reinforcement in the progress
made, Wolf et al. manipulated the number of points awarded
for reading (with 2 subjects) and for working on different
materials (11 subjects), and in each case found that
correlated behaviour changes occurred. "Shifts in point
values to zero produced immediate cessation of behaviour.
Lesser shifts produced intermediate and more variable changes.
Again, it is clear that the token reinforcement system
functioned as such" (p.60). It is interesting to note their
observation that a return of the point contingency to its
original level often was not followed by a corresponding drop
in reading behaviour. "Apparently, exposure to reading
changed the operant level of the reading behaviour" (p.60).

II. Later explorations

1. Non-psychologist administrators

It is evident that token reinforcement motivational
systems have wide applicability, including use by non-
psychological personnel under the supervision and/or guidance
of the behaviour modifier.

Staats, Minke, Goodwin and Landeen (1967) trained adult
volunteers and high school seniors to administer such a system
to 18 retarded, emotionally disturbed or culturally-deprived
children. Over 60 half-hour sessions, reading rate increased
while number of responses required per token almost quadrupled.
Unfortunately, a control group originally part of the study was contaminated when special reading programmes were instituted in the regular classrooms. No differences were found on standardized tests between experimental and control subjects either before or after the experiment, but "when the measurement of reading ability was controlled and directly observed, and when the reading stimuli were among those on which the experimental group had received training, the affect [sic] of the experimental treatment was evident" (p.295).

"Sub-professional therapy-technicians" administered a similar programme to 32 ghetto children who were problem learners, with similar results (Staats, Minke and Butts, 1970). Experimental subjects improved more than controls on the non-verbal section of the Lorge-Thorndike (which involved similar tasks to the training material), and on a 100-word sample of the training materials. The reading section of the Metropolitan Achievement Test tended to favour the experimental subjects, although the difference was not significant at the conventional level (p = .07). Staats et al. suggested that the 40.2 hours of training involved was not enough to remedy chronic academic backwardness.

Ryback and Staats (1970) took this progression a step further by using the parents of four subjects as "behaviour therapy-technicians". One of the subjects was mentally retarded, one "an emotional problem", one a "learning
disability with minimal cerebral dysfunction", and the last a slow learner with a heart ailment. Pre- and post-test comparisons indicated an overall increase in reading ability with at least 53% of new words being remembered 10-15 days after training.

Brown (1972) found some evidence that adolescent tutors in a token reading programme themselves improved on standardized reading tests and on teacher's ratings of department. Eighth-grade students acted as behavioural engineers for fourth-grade students in a study by Willis, Morris and Crowder (1972). Green plastic chips and praise were given for correct sentences, and red chips for errors. Chips were counted and recorded at the end of the daily 30-minute sessions, and after 45 days could be exchanged for prizes. A mean reading gain of 1.2 years was achieved by the students which was a significantly faster rate of improvement than that obtained during the years before treatment, for each subject. Although the introduction of tangible rewards was associated with an increase in reading rates, inspection of individual graphs showed that 5 out of the 10 subjects consequently showed no change or a drop in improvement rate. To explain this finding, Willis et al. suggest that, "for these students, working with bright, enthusiastic, sincere eighth-grade students produced near maximum performance. The need for an individual prescription
of motivating conditions as well as academic material is indicated" (p.69).

2. Kind of Reinforcement

Haring and Hauck (1969) at first provided only feedback on correctness of responses, for four severely reading-disabled boys of at least average intelligence who were working on programmed remedial materials with sophisticated apparatus. Over eleven 65-minute sessions, the number of correct responses declined, while the incidence of behaviour incompatible with reading rose until very little reading was in fact occurring. At this point a counter (of number of correct responses) was added to the apparatus for each subject, with the fascinating result of an increase in correct responses and decrease in inappropriate behaviour for all subjects. These trends were continued for three subjects but reversed for one, when the counter reading was designated as "points", exchangeable for edibles, trinkets and more expensive goods. When the reinforcement schedule was changed from continuous to VR (stretching from VR2 - VR25), reading rates declined and non-reading behaviour increased for all subjects, although not quite to baseline levels.

Overall changes in instructional reading levels ranged from 1½ - 4 years, after five months of instruction, while a constant error rate of 5% was maintained. Teachers commented spontaneously on marked improvements in other areas of
classroom behaviour and achievement. The authors conclude that far more important than concern for possible biological etiology of reading disability, is an awareness on the teacher's part of how the problem can be dealt with by refining instructional and motivational procedures, such that the child's increasing success with reading establishes associated stimuli as secondary reinforcers.

Kortas (1970) found that, whereas the percentage of correct responses to a remedial reading programme increased contingently with tokens, the total number of responses decreased (i.e. speed dropped and accuracy improved). Disruptive and attending behaviours were affected in the predicted directions only on initial introduction of tokens, but not on withdrawal or re-instatement of the token system.

Glynn (1970) found that experimenter- and self-determined rates of token reinforcement were equivalent, and superior to chance-determined and no-token reinforcement. Subjects were four classes of ninth grade girls learning history and geography. There was a significant improvement from the first to the last baselines for the first two groups but not the latter two; the first introduction of the token system had much greater differential effect than the second. It also appeared that chance-determined reinforcement interfered with later capacity to benefit from self-determined reinforcement; the implication is that the experience of random, noncontingent reinforcement actually harms the individual's capacity for self-control.
Hamner (1969) compared the results of social and tangible reinforcement for reading in 14 underprivileged children, and found that, although more correct responses were made with tangible reinforcement, children varied widely in relative responsiveness. Loiry (1970) found that tangible reinforcement produced a higher reading performance in 40 retarded readers, than did knowledge of results alone. A non-verbal "punishment" (a noxious sound) had no effect, either alone or in combination with reward. For the delinquent subjects used by Miller (1971), tokens without backup reinforcers were initially effective in increasing academic performance, but did not maintain this effectiveness when re-instated later (in an ABAB design).

A within-subject comparison of rewarding with money or not was carried out by Camp and van Doorninck (1971). Sixty-six disadvantaged retarded readers were tested on four vocabulary lists, two without and two with token reinforcement with monetary backups. No differences were found dependent on the reinforcement variable, although there was some suggestion (p < .10) that rewards may have a greater effect when material is more difficult.

Kindergarten children of lower and middle socioeconomic status learned to recognize six words under three different reinforcement conditions, knowledge of results only, praise and smiles added, or M & M candy for correct responses and a
buzzer for incorrect responses added, in a study by Pikulski (1970). For the 24 middle-class children of both sexes, and the lower-class girls, the most effective reinforcer was social, with material reinforcement and straight knowledge of results equally inferior. Most errors were made with material reinforcement: "it appeared that the child's preoccupation with the candy, deciding whether to eat it or not, etc., was extremely distracting and contributed to a poorer performance" (p.521).

III. Some non-token successes

Clark and Walberg (1968) used verbal rewards, of which 110 under-achieving 10-13 year olds kept records by marking squared paper. After six sessions during which the teachers had been asked to distribute at least some praise to every child every day, five of them were secretly asked to double or triple the amount of praise given, while the other four teachers were asked to "keep up the good work". The increased frequency of social reinforcement was significantly associated with higher scores on a standardized reading test. The authors commented on how "revolutionary" the idea that reinforcement enhances learning seemed to the teachers and children in the experiment (p.309).

Lovitt, Guppy and Blattner (1969) added a free-time contingency, and a group contingency which was never in fact
won, to traditional procedures in teaching spelling and found a majority improvement. Again differential responsiveness to the "reinforcement" was noted. A similar Premackian system was implemented by McIntire, Davis and Pumroy (1970). Teachers supplied a list of all assignments completed since the last session, and the points earned. Depending on whether the total was higher, lower or the same as the last score, the subject was then permitted various activities (e.g. teaching machines only for a performance decrement, complete freedom of choice for a 10-point improvement). For the fifth grade but not the sixth grade children, counting or not counting spelling or mathematics papers towards points earned was effective in altering classroom and homework performance. On the assumption that "the breakdown in efficient classroom performance (reading) was related to tense behaviours commonly defined as anxiety" (p.184), Muller and Madsen (1970) compared desensitization with placebo procedures. The experimental subjects received 20 half-hour sessions comprising 15 minutes muscular relaxation training and 15 minutes presentation of a graded hierarchy of items relating to reading, culminating in in vivo practice for increasing amounts of time (from 10 seconds in the fourteenth session to 120 seconds in the twentieth) of reading aloud to the rest of the group. The placebo group were read stories for equivalent amounts of contact time, and had exactly the same practice in reading
aloud. Children's Manifest Anxiety Scale scores and self-reports of reading anxiety were significantly lower for both the groups after treatment, compared with matched no-contact controls. None of the three groups changed in reading achievement. The authors attribute the unexpected efficacy of the "placebo" treatment in reducing anxiety, to modelling of relaxed responses by the experimenter while reading aloud.

A significant improvement in achievement scores was obtained after desensitization of high test-anxiety by Katahn, Strenger and Cherry (1966). Unfortunately, the effects of the desensitization training are confused with those of counselling, which included suggestions and advice on how to develop solutions to problems of studying and taking tests. The students invariably felt that the counselling was of paramount importance: "it is impossible to determine which aspect of the programme was, in fact, more important" (p.548).

Lastly, a reading training programme is described in which the only reinforcer provided was a play period after each session. Beginning from the psycho-analytic premise that "learning to read, can in fact be treated as part of the conflict-free ego sphere and as a secondary automatism", Cameron, Borst, Fifer, LaVigne and Smith (1972, p.273) developed an individualized remedial programme of successive approximations to the final desired product (successful reading). A year's schoolwork could, this way, be completed
in six to seven weeks, with three 45-minute sessions per week; follow-up information was not given. The child's response to the programmed material was monitored constantly and any area of stress or conflict corrected by specific teaching. Strain on the child's part, in other words, was taken as a signal of the need to modify the teaching programme.

IV. Review of studies on token reinforcement for reading

As is evident from the foregoing, attempts to modify reading behaviour particularly by using token systems have flourished in the last ten years. Staats' early work has had a stimulating effect and been extended in many directions. It is also clear, however, that fundamental questions remain unanswered and in some cases apparently unrecognized. The main areas of vagueness are the effects of punishment, the necessity for material incentives, and the general adequacy of experimental design.

1. Punishment

It is interesting to note the extreme importance attributed by Staats to the avoidance of all negative consequences for wrong reading responses. Thus Staats, Minke, Goodwin and Landeen (1968, op.cit.) found it necessary to replace one of their therapy-technicians, and supplement instructions to the others, on noticing the effects of urging and implied criticism. Remarks such as
"You can do better than that", "Try harder", and "You know that before" (p. 297) were found to cause hostility and increased disturbance in one subject and brought him to the point of withdrawing from the programme.

Similarly, Ryback and Staats (1970) stress the necessity of ensuring the absence of any negative social reinforcement. They had difficulty with a mother who corrected reading mistakes with an annoyed, disappointed vocal intonation.

Of the other surveyed studies using an operant approach to remedy reading problems, none includes comments about negative social reinforcement and only three refer to using a supposedly "noxious" consequence of errors (a buzzer). McKerracher discontinued the buzzer when the subject began to read more fluently, Loric found no effect of the buzzer either alone or in combination with reward, and Pikulski included a buzzer for incorrect responses with the material reinforcement of candy for correct ones. Willis et al. signalled the occurrence of incorrect responses with a different coloured chip, and the withdrawal of the tutor's attention to the other member of the remedial pair.

Altogether, it can be seen that the effect of aversive consequences for inaccurate reading has not been systematically investigated. With low intensities of
supposedly noxious stimulation, the question of confounding with feedback of results of course arises (see Chapter VI). The sensitivity of subjects to negative social reinforcement noted by Staats et al. is, however, a provocative finding and suggests that classically-conditioned fear can be an important element of the learning process, in the failing student.

2. Material reinforcement

Individual differences in responsiveness to different forms of reinforcer can often be postulated to account for experimental findings. At present, however, there are few reliable guidelines to the controlling power of different contingencies, except those derived from a behavioural analysis of the person concerned (e.g. Kanfer and Saslow, 1965). Neither theory nor the experimental literature makes clear whether, or when, strong positive reinforcers of a material kind are necessary to counteract previous failure and/or missed learning opportunities, and when the mere provision of non-aversive training trials is sufficient to desensitize the reader's learned anxiety.

Staats et al. found originally (1962) that 4-year olds would not work at reading training for social reinforcement alone (over from ½ to 4 sessions), but worked well for tokens backed up with a novel toy (for up to 7½ sessions). It may be relevant to note how session length in Staats' published research with pre-school children has decreased: from 40 minutes in 1962 to 2.93 minutes average for the first 20
sessions in 1970 (Staats, Brewer and Gross, 1970). Staats' reliance on material reinforcers could possibly be unnecessary with the much shorter sessions, but he has not provided information on this point.

Few later investigators have made detailed comparisons between material reinforcement and its absence, where feedback of results and orderliness of learning materials were held constant. Also, the effect of withdrawing material reinforcement with a reversal design would a priori be expected to differ from the effect of never presenting it, where for example, matched groups were kept ignorant of the treatment received by their fellows.

Whitlock and Bushell (1967) and Haring and Hauck (1969) both found that the initial introduction of a counter (to increase feedback but not represent "points") was effective in improving performance. Willis et al. (1972) found that half their subjects were affected badly or not at all by the introduction of tangible rewards, and many other studies (e.g. Hamner, 1969; Haring and Hauck, op.cit.; Kuypers, Becker and O'Leary, 1969; Martin, Brukholder, Rosenthal Tharp and Thorne, 1968; O'Leary, Becker, Evans and Sandergas, 1969; Rickard, Clements and Willis, 1970) refer directly to the extreme variability in responsiveness to token system both across and within individuals. Finally, Pikulski (1970) found social reinforcement superior to material reinforcement which was associated with the highest
error rates, and Hewett et al. (1969) found that task-attention improved when token reinforcement was withdrawn, and that reading achievement was unaffected either way.

There is not enough evidence to decide the exact importance of material reinforcement, from the published experiments on token-management of academic behaviours. Clues from verbal-learning and discrimination-learning studies will be sought later (Chapter VI).

3. Experimental design

In many of the studies cited in this chapter, valiant efforts at measurement were rendered difficult to interpret by a weakness of experimental design, rather than the inherent elusiveness of reading quantification. This weakness comprised the failure to provide convincing control data.

In referring to his series of studies on token reinforcement of reading, Staats (1970, p.9) stated: "One of the strengths of such a developing research series is that each study does not stand alone and that much flexibility and innovation are possible. Each study need not include all of the informal and formal controls required of the series". Presumably it is on this basis that Staats has not included control procedures for tokens since his earliest work. There he used response-rate data to compare tokens with social reinforcement (1962, 1964) and with no reinforcement (1964), all in within-subject contexts. The effectiveness of tokens
having been established in maintaining and even accelerating work in young children faced with a new learning task in extremely long sessions, it was taken for granted in later experiments. As has been shown, results of other workers have not unequivocally confirmed the initial finding.

Several studies have claimed to demonstrate the usefulness of token systems without taking account of the special teaching and attention involved in their administration. It seems rash to conclude that tokens are superior teaching aids, when the only comparison made is with untaught subjects (as in Clark, Lachowicz and Wolf, 1968; Staats, Goodwin, Minke and Landeen, 1967; Staats, Minke and Butts, 1970). Staats and Butterfield (1965) and Willis et al. (1972) made systematic use of the subject's past history as a control measure, while Ryback and Staats (1970) claimed that siblings of subjects could be regarded as control subjects. Only Wolf, Giles and Hall (1968) have provided both kinds of control data which seem desirable; they compared remedial with non-remedial subjects and, within the former group, used specific contingency manipulations to show the relative value of tokens.

Another need demonstrated by the theoretical discussion of Chapter II was for attention to be paid to the extent to which effects of a token programme generalizes beyond the therapeutic situation. It has already been
noted that attempts to measure out-of-experiment reading frequency are practically impossible; however, permanence of therapeutic effectiveness could be assessed by follow-up tests.

Detailed long-term follow-up tests with adequate description of post-experimental conditions are never reported. Indeed, there is usually a failure to report follow-up examinations of any kind, however cursory, if we discount Ryback's (1970) "long term" retention test after 10 to 15 days.

Summary

Uncertainties remain about how token systems might most efficiently be set up and their effects most comprehensively assessed. Answers to these queries have relevance not only to theoretical understanding of behaviour modification processes, but also to understanding of the reading (and thereby verbal learning) process.

Where tokens are found to be an effective aid to reading training, the reason for this is not automatically clear. The possibility discussed in Chapter III, that much of their value lies in forcing increased systematization of the teaching material and of knowledge of results, remains.

The relative importance assigned to the various questions and criticisms raised in the last two chapters, will depend
partly on the theoretical bias of the reader. It seems fair to conclude, however, that although there is substantial evidence for the usefulness of token systems in improving reading achievement, definitive answers on specific points are rare. One could at least put up a case for the argument that token systems have proved to be of very uncertain value in attempts to produce irreversible increments in reading behaviour, and that their superiority to other modes of reinforcement (e.g. information on progress, positive social events, etc.) has not been clearly shown. Token systems must not only work, but must work better than other techniques, even when used by other than their originators, if they are to be worth implementing.
Chapter V. Reinforcement and human learning

Up to this point an operant theoretical context has been adopted in discussing the modification of children's behaviour, and particular attention has been paid to the reinforcers which control the acquisition of cognitive skills. When token reinforcement is made contingent on progress in reading, several experimenters have reported improvements in reading ability. Even laying aside for the moment methodological questions such as the measurement of skill and the assessment of change and its permanence, it has become evident that the operation of a token system may be more subtle than is frequently implied.

This chapter widens the range of behaviours studied, in order to take advantage of the insights collected by laboratory experimenters about how human learning is affected by different consequences contingent on the response. The two major issues explored are those arising from the earlier discussion, i.e. the effects of different modalities of reinforcement (material, social, and information), and the effects of aversive consequences.

Verbal learning, concept learning and probability learning, are often regarded as epitomizing those processes which are exclusively human, and there is an extensive experimental literature on these topics, recently summarized by Cofer, Millward and Postman in separate chapters of
Woodworth and Schlosberg's Experimental Psychology" (Kling, Riggs, et al., 1972). Paired-associate learning in particular has been seen as a suitable and convenient vehicle for the study of verbal behaviour. The generalisability of the findings from this area of experimental psychology to extra-laboratory learning situations can, however, be over-estimated.

The greatest proportion of research generated by Skinner's book "Verbal Behaviour" (1957) has involved the effects of reinforcing a preselected class of verbal responses (such as "plural nouns"), and the controversy over whether response-patterns can in fact be altered without the subject's awareness of what or how he is "learning" (e.g. Greenspoon, 1962; Holz and Azrin, 1966; Krasner, 1958; Spielberger, 1965). Recently studies have begun to appear in which less artificial speech-modification is achieved for therapeutic purposes by operant methods (e.g. Bennett and Ling, 1972; Goodkin, 1969; Johnston and Johnston, 1972).

Kintsch (1970) considers the two variables of amount of reinforcement and delay of reinforcement in his examination of whether it is appropriate to analyse verbal learning in terms of conditioning. He shows that whereas any delay of reinforcement impairs learning in the Pavlovian conditioning situation, delay of reinforcement in verbal learning may be irrelevant; it may have a deleterious effect (due to over-loading short-term memory), or it may even be facilitatory,
all depending on how the subject can best process the relevant information. In the case of magnitude of reinforcement, it is unexceptional to find that this variable does not necessarily correlate positively with performance level (e.g. Berman, 1970). Kintsch comes to the conclusion that it was premature of Thorndike to reject an interpretation of reinforcement in informational terms.

One important issue here seems to be concerned with methodology. As Meyer (1951) had found with monkeys doing discrimination reversals, the range of rewards being presented is an important influence upon the effect of any individual reward. This conclusion was supported by Cuvo and Witryol (1971) who decided that within-subject designs show differential incentive variance, and between-subject designs consistently fail to do so. They accordingly used a within-subjects design and found superior recall of words associated with the higher of two monetary incentives.

The distinctions between expectations from within-subject and between-subjects studies, and between a social or material reinforcer and a verbal information-giving event such as "Right"/"Wrong" (often referred to as a reinforcer), must be kept in mind when reviewing reinforcement effects on human learning: "reinforcing properties do not lie in the stimulus, but in its effect upon behaviour" (Reese, 1966, p.32).
I. **Type of reinforcement**

Both behaviour modifiers and learning theorists are interested in how and when contingent stimuli influence the efficiency of new learning, which includes behaviour change as well as symbolic associations. The studies reported in this section investigated (a) positive reinforcement alone, and (b) tangible and social reinforcers and information. The tasks varied, but a high proportion of discrimination experiments have been included in the belief that discrimination skills are a relevant ingredient of reading.

1. **Positive reinforcement**

"Your job is to work for points", Verplanck's subjects were told in 1956 (p.71). They did, making various motor responses such as tapping their chins, or verbal responses with a certain theme, while they were conditioned, shifted to intermittent reinforcement, taught simple discriminations, and extinguished. The process usually took up to an hour, and subjects showed varying degrees of "awareness" which did not seem to correlate highly with changes in rate of responding. Verplanck found it important that the subject should have before him a record of his points, but "backups" were not apparently regarded as a necessary feature.

A further demonstration of the reinforcing value of feedback alone, was provided by Miller and Estes (1961). Matched groups of 9-year old boys at different schools who
were reinforced with either 1c. or 50c. for correct
discriminations (as well as a light flash) made an equally
greater number of errors compared with those who received
the light flash alone. The experimenters suggest "that
preoccupation with the money (counting, gloating, worrying)
constituted a sort of interpolated task" (p.503).

Fazio, Lohr and Screven (1971) provided non-compulsory,
"spelling-game" tasks in a recreation room for educationally
disadvantaged boys. No extrinsic reinforcers were provided,
although an assistant said "correct" or "incorrect" after
each response, and the subject in each pair with the more
correct responses was declared the "winner". Despite
competition from pool, pingpong and television in the same
room, a median of 40% of the boys present each session
participated in the task over 9 months. Automation of the
procedure appeared to retard progressive weakening of
reinforcer strength.

In a very different project, Harley (1958) compared
paired-associate learning with and without monetary incentive,
within lists of pairs. The subjects, who were college
students, learned best and made most use of learning
strategies if incentive-cueing occurred during (rather than
after) response to the material. Pihl and Greenspoon (1969)
found that amount of associated monetary reinforcement (1, 2,
5 or 10 cents) affected the relative reinforcement value
of a verbal stimulus.
2. **Tangible versus social reinforcers**

Children of four different age groups (pre-kindergarten and Grades 3, 6, and 9) pressed knobs in an experiment by Nickell and Travers (1963). Consequences were either "good", a marble, "good" plus marble, or no reinforcement. Subjects were promised a toy at the end of the session; the authors unfortunately do not report whether or not the subject expected to retain the marbles afterwards (although presumably not). The physical reinforcement was found to be more effective than the verbal at all age levels and for both sexes, with the exception of the youngest girls. The greater novelty of mechanical marble-delivery compared with an approving comment was postulated as an explanation of these results.

Another motor task (key-pressing) was used with kindergarten children by Brown (1971) to investigate effects of tangible reinforcement (candy), social reinforcement (a fixed sequence of approving comments), the alternation of these two conditions, and subsequent non-reinforcement for each group. Density of reinforcement (FR = 10 or 20) made no difference, but sex differences were found in that girls increased their performance more over trials than did boys. The highest response rate was obtained when tangible reinforcement alternated with social reinforcement; the author suggested that this may have been due to both
intrinsich and conditioned reinforcing properties of the candy. When reinforcement was discontinued, groups previously receiving social reinforcement remained above baseline, but those previously receiving only tangible reinforcement dropped significantly below and gave behavioural evidence of frustration (e.g. shouting, pounding the key, etc.). Brown accordingly emphasized the need for contingent social rewards as a part of tangible-reinforcement treatment designs.

Similar results were obtained by Blair (1970). Low-achieving third-grade boys performed better at an unspecified learning task with tangible reinforcement, and normal-achievers with social and feedback reinforcement. When all reinforcement was withdrawn, those who had received verbal reinforcement again performed longer and at a higher rate than those who had received tangible reinforcement.

Smith, Brethower and Cabot (1969) provided praise, money, a work-break, and graphical representation of progress, in different studies with retarded readers working on a programmed literacy curriculum. Monetary and informational consequences were found to be equally effective in maintaining or accelerating work rates, and superior to praise and/or no feedback.

Finally, a therapeutic behaviour change was the goal for Clement, Fazzone and Goldstein (1970). Successive approximations of desirable social behaviours in withdrawn
boys were rewarded by verbal reinforcement with or without tokens leading to material backups. Two further groups were controls for (a) therapist-contact and (b) therapist-contact and group-contact; the experienced child-therapist who administered the treatment to all had an anti-behavioural orientation. Although the total therapeutic impact of the programme was described as "modest", contingent material reinforcement produced the most behavioural change, and the group members remained better adjusted than the others 1-2 years after termination of therapy. Having a therapist present caused more changes for both better and worse than not doing so, but "group treatment which includes token reinforcement for social approach behaviours is a better therapy than that which fails to use the tangible rewards" (p.423). This conclusion, of course, is not incompatible with an informational interpretation of the operation of "rewards".

3. Tangible versus social reinforcers and social class

Several researchers have proceeded on the assumption that responsiveness to material reinforcement may be a function of reinforcement history specifically mediated by socio-economic variables, possibly including child-rearing practices as well as absolute wealth.

Performance on a discrimination learning task was poorest when the high social-class subjects were promised a
bag of candy "after they had made the light go on enough times" (Terrell, 1958, p.233). Immediate contingent candy, immediate tokens for future candy, and no consequence other than the light flash, all produced equal levels of performance. A sample of rural children, however, performed worse in the light-only condition. "Good performance, irrespective of incentives", was thought to be immensely important to the first sample but not the second (p.235).

Spence and Segner (1967) varied social class (lower and middle), kind of reinforcer ("Right"/"Wrong" versus candy and a raucous sound), and reinforcement combination (reward only, punishment only, and reward plus punishment). The only treatment effect was inferior performance by children of both social classes under the reward-only reinforcement condition, for which "some kind of distractibility phenomenon" was hypothesized (p.36). Spence followed this lead in 1970 and found further evidence. Basically, material reinforcers or tokens with backups produced poorer performance in discrimination learning than did feedback alone. She concluded that skepticism is the only appropriate response to generalizations about the lesser susceptibility of working-class children to symbolic reinforcers. "Even within a given socio-economic group, the relative effects of the two kinds of reinforcers on performance, as the present investigation illustrates, appear to be highly dependent on the specific procedures employed" (p.110).
Later investigations by Spence (1971, 1972) along the same lines, have given similar lack of support to any over-riding generalizations about the poorer learning of lower-class children under symbolic compared with material rewards.

The effect of plastic cows was contrasted with that of positive remarks such as "very good", for control of both learning and choice behaviour in a study by Witryol, Lowden, Fagan and Bergen (1968). Instructions emphasized either skill ("remember to try hard and play the game right") or luck ("It doesn't really matter what you pick. It's just a matter of luck") (p.8). The children, from grades 1, 3 and 5, distributed their choices fairly evenly over the two reinforcers during 80 trials, although the older ones tended to prefer verbal incentives (perhaps being more quickly satiated with plastic cows). The authors make further interpretations of their data, concluding "that boys and high SES subjects were motivated by internal locus of control, while girls and medium SES subjects were more responsive to conformity and social approval needs in their achievement behaviour" (p.24).

Swingle and Coady (1969) compared the effects of money, "good", both of these, and neither, in a lever-pressing task with middle and lower class children of 6, 10 and 15 years of age. The younger children showed no differential responsibility to the reinforcers, but by 15 the middle class children performed faster than the lower class subjects under every
reinforcement condition, showing a special sensitivity to verbal incentive, but none between money and no reinforcement. The lower class children on the other hand with age became increasingly slower performers except where material reinforcement could be earned.

An extra dimension was added to this research question when Shores (1969) compared normal and disabled learners on a size discrimination task, and found that middle class normal achievers and lower class slow learners tended to respond better when tangible reinforcement was not added to confirmation of results, while the opposite was true for the middle class slow learners and lower class normal achievers. In other words "children with school learning problems seem to be affected differently from normal learning children by the available reinforcers" (p.644).

Norton, Versteeg and Rogers (1970) combined cue and motivational aspects of verbal reinforcement ("Yes, right" versus "No, wrong"), with or without candy. Their middle class 5 year olds performed better than those in a Head Start class, and there was a tendency for verbal reinforcement alone to produce better discriminations. Once again candy was seen as a distracting influence.

This view received yet further support in a careful recent experiment by McCullers and Martin (1971), who began as few others have done, by obtaining from the subject population measures of subjective desirability of different
typical incentive objects. Using this information, they explored the possibility that "incentives" which are valued lowly may simply act as a form of feedback, while highly valued objects motivate in addition. However, verbal feedback ("Yes") and "most preferred" incentive conditions were responsible for better discrimination learning than were "least preferred" incentives. Further, two different groups of children receiving high-value incentives for either correct or incorrect responses performed equally well.

There is thus a consistency in the results of experiments using a variety of material and social reinforcers, with children of different age, social status, and learning skill. This consistency lies in the demonstrable inadequacy of conceptualizing positive reinforcement as a homogeneous variable with a sole, incentive function. Information, and interference or distraction, seem to be the major alternative functions carried out by the awarding of incentives.

II. Aversive consequences

Thornkike's modification in 1932 of his original strong version of the Law of Effect failed to cope with conflicting evidence about punishment effects, which is still being gathered. Although punishment is often found to decrease response-probability or increase its latency, sometimes the effect is only temporary, sometimes absent, and sometimes the response is paradoxically strengthened (Church, 1963). In
1964 Solomon concluded (p. 252) that "Our laboratory knowledge of the effects of punishment on instrumental and emotional behaviour is still rudimentary ..."

Dunham (1971) complained that "there has been a lack of interaction between punishment theory and punishment data" (p. 59). Criticizing the assignment of all experimental time not spent performing a selected operant to the "nonresponse" category, he advocates the use of a multiple-response baseline procedure. Here several measurable responses are observed which fill the experimental time; the effect on the total behavioural picture, of aversive stimulation contingent on one response, can then be quantified. Dunham presents data (pp. 63-64) supporting his view that the probability of that response "which is most frequently associated with shock onset and/or predicts the onset of shock within a shorter time than other responses" will fall below baseline, and the converse for the response which is associated with and/or predicts the absence of shock onset.

In a recent theoretical analysis Fowler (1971) points out that there has been a general failure to distinguish the suppressive functions of punishment from its role as a signal of non-reward. The cue or signalling functions of punishment, however, are not distinctive; what is characteristic of punishment as a separate process from others is its function in eliciting avoidance, emotional and escape behaviours. The temporal relation between aversive stimulus and response,
the similarity or associated cues, and the intensity of the aversive stimulus, will all determine the suppressive influence of punishment (Fowler, 1971, p.597).

Although a variety of *prima facie* noxious stimuli have been employed in animal research, electric shock has been by far the most popular, for not only are its characteristics easy to measure and control, but its use at sufficient intensity is patently "noxious" (Church, 1972, pp.703-4). In research using child subjects, no such ready solution is feasible.

Marshall (1965) reviewed the literature on the effect of punishment on children, where "punishment" included criticism, failure, "wrong", and the repossess of material reinforcers. She found that by far the greatest proportion of studies show that punishment, either alone or in combination with reward, is superior to no punishment or reward alone.

As the mild degrees of punishment permissible with child subjects appear to aid rather than hinder learning, perhaps due to increased probability of non-punished responses and thus behavioural variability, Marshall sees the main theoretical issue as this conflict between the motivating, instigative role of punishment and the informational role; or whether perhaps both concepts may be necessary. She then re-examined the available research and concluded that "negative reinforcement of specific responses has an information and therefore beneficial effect, whereas the effect of negatively
reinforcing the situation depends upon the action of other factors" (p.32).

1. Effects of "Right" and "Wrong"

Buchwald (1962) using a modified paired-associates learning task obtained essentially similar results to those which led Thorndike to believe that, while "Right" strengthened the preceding association, "Wrong" had little or no effect in the opposite direction. When the subject had four alternative responses, "Right" strengthened learning more than did the control (experimenter silent) and "Wrong" had no effect; however with two response-alternatives, "Wrong" weakened learning more than did the control, and "Right" had no effect. These contradictory results were followed up by Mosberg (1970). To test whether a consideration of (a) the information value of "Wrong", and (b) the memory requirements of the task, had explanatory power, he manipulated the variables of reinforcement combination ("Right"/"Wrong", "Right"/Nothing, Nothing/"Wrong"), list length (5 or 10 word-pairs), and number of response alternatives (3, 5 or 8 to each stimulus word). His college student subjects learned best with "Right"/"Wrong" and least well with Nothing/"Wrong". List length and number of response alternatives affected results as would be predicted in terms of memory load and the information value of "Wrong".
Discrimination learning proceeded faster in first-grade subjects under "Right"/"Wrong" and Nothing/"Wrong" conditions, than with "Right"/Nothing, in a study by Rothberg and Harris (1972). They found none of the debilitating emotional effects of "Wrong" which Mosberg had hypothesized to be influential, and were forced to conclude that the nature of the specific task (they used three different ones) is an important variable; no "single simplistic interpration" (p.284) is possible.

Siegel and Van Cara (1971) found that incidental learning was significantly better for subjects trained under "Wrong"/Blank conditions, than under "Right"/Blank or "Right"/"Wrong". This finding tends to support the hypothesis of behavioural variability being greater under punishment conditions.

2. Effects of other positive and negative reinforcement

Schere (1969) compared the effects on paired-associate learning using pictures, of reward, reward-punishment, punishment, and no reinforcement; the reinforcement was giving and taking chocolate bars. For intellectually gifted boys, punishment was significantly more effective than reward; for educable retarded boys, reward or no reinforcement produced better learning than did punishment or reward-punishment. These results are consistent with the proposal that reinforcement effectiveness is correlated with prior expectations of success. The gifted children may be
intrinsically motivated and distracted by extrinsic rewards, while the retarded may strive to avoid punishment.

Money accumulated following correct responses (reward condition) and was withdrawn for errors in the punishment condition, in Powell and Hake's (1971) study with high-school boys performing matching tasks. In a series of four experiments they showed that (a) reward produced a superiority in accuracy over negative reinforcement, but this superiority was transient; (b) behaviour maintained by negative reinforcement was most sensitive to changes in experimental variables (e.g. became more inaccurate when the task difficulty increased); (c) positive reinforcement produced an observing response (button pressing on FR50) with higher rate and shorter latency than did negative reinforcement. The superiority of combined positive and negative reinforcement compared with reward only, and the poor performances under punishment only, are taken to suggest that punishment alone is not the critical factor influencing efficiency.

Cool (1972) found that groups of 1st and 3rd grade subjects who were informed about reinforcement conditions, did not perform differentially on a brightness discrimination-learning task, according to reinforcement combination (reward, punishment, or reward-punishment), or type of reinforcement (verbal reward and punishment, candy award and deduction, or candy reward plus tone punishment).
Two studies using token reinforcement as reward and a loud noise for punishment with discrimination learning, were published in 1971. Schlichter and Ratliff found that the performance of delinquent boys was controlled better by reward, and the performance of non-delinquents by punishment, a result fitted by the same expectancy-theory explanation as that of Schere (1969). Neither group showed evidence of learning in the combined reward-punishment condition, an odd result which the authors attribute to distraction of subjects from the task by too much information. Kindergarten children made tactile discriminations for Witte and Grossman (1971). Those receiving reward-punishment and punishment alone, performed better than the reward alone group. "Number of orienting responses" (not defined by the authors) varied across groups and trial blocks, and this is interpreted as providing support for an attentional rather than motivational explanation of the results of punishment, particularly as response-force (the motivational measure) did not differ between treatments.

3. Conclusion

It is evident from the diversity of results in this sample of the literature, that several parameters affect what happens when a supposedly aversive consequence is introduced. One might, however, query exactly how aversive the stimuli used as punishment were. It seems likely that
"Wrong", a loud noise, or having some of the money or sweets which have been presented beforehand withdrawn, are at most mildly aversive and certainly not comparable with the painful electric shock usually used in animal studies. In addition, the lack of published replications of experiments makes it difficult to interpret different results; too many variables are changed simultaneously, even in a "standard" learning task such as paired-associate learning. On the whole, information-load analyses of how weak negative reinforcement operates appear to hold promise, if qualified by considerations of past reinforcement history and possible emotional or motivational factors.

III. Summary

The subject of the present chapter has been experiments on positive and negative consequences for human learning where these have been material, verbal and informational in form, and mainly with children of various developmental and social characteristics. Many of these questions have apparently been regarded as solved, or have at least been shelved in the interests of making an immediate practical contribution to the personal and social crisis of reading retardation, by applied researchers (see Chapter IV).

In looking at the effects of presumed punishers, including saying "Right"/"Wrong", repossessing candy or sounding a buzzer, again and again not incentive but
informational effects were suggested. The situation is similar in reviewing the use of positive reinforcers, used alone or in attempts to explore the differential effects of verbal versus tangible rewards or social class of the recipients. It is painfully apparent that no valid generalizations can be made about what the relative effect of a particular consequence will be, for subjects of given maturity and reinforcement history, faced with a learning task. Predictions must take into account not only the incentive-value to the child of a consequent event, its role as a source of feedback, and as a source of distraction from the task in hand, but how the study has been designed and what conditions are really being opposed (particularly the question of what range of reinforcers is provided or known to the subject).

An informational interpretation of how "reinforcement" affects human learning has appeal, particularly with the mostly low-intensity reinforcers surveyed; however, it seems reckless of Kintsch to claim that "the law of effects is wrong" (1970, p.42). One of the main dimensions of complexity must lie in the effort to predict, as Athey requires (1971, p.98), "what kind of reinforcements work with what kinds of groups, and why". Other issues, such as upon what literal reinforcers must be contingent (e.g. an accurate response or attending to the lesson?), and what is the role, if any, of reinforcers labelled as "intrinsic", and how their supposedly desirable influence can be augmented, have not yet been considered, despite their importance.
Chapter VI. Other approaches to reading retardation and its remediation

Pre-dating and contemporaneous with the marked interest in reading difficulty as a suitable target for "behaviour modification" techniques, is the interest of teachers, medical practitioners, and educational psychologists in this problem. There is a wide range of literature from these points of view, and the approaches of these three groups have tended to overlap and stimulate each other. This chapter focusses on how the problem of reading retardation has been conceived from the standpoint of assessing its epidemiology, exploring its causes and effects (which are often very difficult to separate), and prescribing suitable treatments.

It should be noted at the outset that treatments are not usually assessed with the rigour characteristic of experimental endeavours, and that the proportion of children who can be expected to benefit from remedial help seems to be at least as dependent upon the age and determination of the child, and the child-teacher relationship (all of which might be summarized as motivational factors) as upon the particular remedial method used.

An abundance of practical advice to the reading teacher is available in print. He or she can learn the details of how to choose and present training tasks (from e.g. Broom,
Duncan, Emig and Stueber, 1951; Dolch, 1951; Durr, 1967; Frase, 1966; Harris, 1961; Moyle, 1968; Radford, 1960; Schonell, 1942, 1950, 1961; Tansley, 1967; Ulmer, 1969; Warderberg, 1963), or how to use specific techniques such as the initial (phonetic) teaching alphabet (Downing, 1965; Warburton and Southgate, 1969), and individualized approach (Veatch, 1959), words in colour (Gattegno and Hinman, 1966), or an emphasis on the intrinsic rewards of knowledge (Bruner, 1966). When things go wrong and some children of adequate intelligence fail to learn to read, numerous remedial prescriptions have similarly been made, from changing the slant of the paper so that a left-handed writer can see the whole word (Ancevich and Payne, 1961), to changing the social order in order to equalize educational opportunity (Eisenberg, 1966).

1. **Epidemiology of reading retardation**

Definitions and opinions of the phenomenon originating from the three different disciplines of education, medicine and psychology, converge and will be dealt with in turn. Incidence surveys by those concerned primarily with education, tend to rely on the mass administration of more-or-less standardized tests. Ullmann (1971) refers to these as "norm-referenced" measures, as opposed to the "criterion-referenced" measures which describe an individual's performance unrelated to that of others. After distinguishing
between underachievement, retardation, illiteracy, and functional illiteracy, he recommends a pragmatic selection of measures according to the purpose intended.

Barker, Fee and Sturrock (1967) gave the Burt Graded Vocabulary Reading Test to all Dundee school-children of between 8½-9½ years, and found that 1% of those with average intelligence (IQ 90+) were virtually unable to read at all. Vernon (1957) quoted U.K. Ministry of Education statistics to the effect that 1.4% of 15-year olds and 4.2% of 11-year olds could read less well than the average 7-year old (in a country where school attendance begins at 6), and that 4.3% of 15-year olds and 20.3% of 11-year olds were functionally illiterate, i.e. read less well than the average 4th grader. By 1965, although the proportion of those who were 2 years retarded in reading was similar (15-20%), the total non-readers had dropped to 1% of the 11-year old population.

Witty (1949) estimated that one man in seven in America was functionally illiterate, and Williams in 1970 put the figure at 10%. According to Ullmann (1971), 6.1% of Americans 25 years old or over, and 4.8% of 14 years or older, have completed fewer than five years of schooling. Meier (1971), however, found that about 15% of the second-graders he surveyed in 110 classrooms and 8 States, had some form of communication deficit which was hampering
reading achievement. Although local statistics are sparse, Lasscock (1965) found that one third of South Australian pupils on entering secondary school, had a reading standard below average for beginning Grade VII, the previous grade.

"Fewer" reading difficulties are experienced in Czechoslovakia and Spanish-speaking countries (where spelling is phonetic) than in English-speaking countries (Matejcek, 1964; Money, 1962, p.22). Makita (1968) quoted an incidence of only 0.98% in Japan. It does seem clear that a proportion like 10% of children, fail to read as well as they are expected to (whether expectations are realistic and well-founded is, of course, a separate question). It is not known how far this figure could be reduced by improvements in the educational system such as smaller classes, better equipment and better teacher-training to spot and help early difficulties. Other individuals might be suffering from missed or inconsistent schooling, pre-occupying family troubles, or too little encouragement from home. It seems reasonable that yet others should find reading difficult due to characteristics of their neural function which, for example, place them at the low end of a continuum of requisite information-processing skills.

2. Characteristics of retarded readers

Applebee (1971) argues for a distinction between extrinsic causes of reading retardation, such as home and
cultural variables and other "social" factors, and intrinsic or individual-specific causes, which determine the individual's relative standing within a homogeneous cultural group.

It is with such "intrinsic" factors that medically-oriented researchers have primarily been concerned. Where intelligence, sensory capacities, educational opportunity and emotional adjustment are all normal so far as can be determined, various manifestations of the syndrome "dyslexia" (to use the most common of several terms) are often held responsible for inability to read.

The relationship between reading disability and emotional disturbance is a thorny one. Commonly (e.g. Bannatyne, 1966; Rabinovitch, 1962; Silver, 1971), a category of non-dyslexic underachievers in reading is distinguished, where extreme anxiety, depression or hostility, or psychosis, has a primary role. At the same time, however, it is generally acknowledged that a common adjunct of the neurological dysfunction which can cause reading retardation, is personality maladjustment directly attributable to both failure at reading and to the characteristic clumsiness, hyperactivity, distractibility, and tendency to perseveration (Silver, 1971).

The dyslexic child, according to Critchley (1966), Money (1962, 1966), Reid (1968) and Vernon (1957, 1965),
may or may not have a history of minor cerebral damage before, during or shortly after birth; he may be clumsy and ill-co-ordinated or may excel at non-verbal tasks; he is sometimes impulsive and sometimes sluggish. Many have mixed lateral dominance, or affected relatives, and male sufferers out-number female by at least 2:1 (with ratios of up to 9:1 being cited).

The kind of neurological dysfunction involved can commonly be detected only on examination by means of its hypothesized result: the reading difficulty. In fact whether neurological impairments have a primary, or indeed any, role, is difficult to ascertain when emotional and educational factors are obviously contributing to the difficulty (Bryant, 1967). "The causes of reading disability can never be pinpointed accurately. They are complex, often obscure, and always interrelated" (Roswell and Natchez, 1964, p.7).

Psychological research on the characteristics of retarded readers has similarly failed to provide conclusive answers. Lyle and Goyen (1968) found that retarded readers were poorer than normal controls at recognizing letters, lines and word shapes when presented tachistoscopically. They regard their finding that the two groups of 7-year olds differed more than the two of 9-year olds, as support for Critchley's "developmental lag" theory of aetiology. Goyen (1969) concluded from a series of experiments that
reading failure is related to a slowness to register the
essential details of complex forms, while Lyle (1970) found
in a retrospective study that early speech defects were the
best of several possible birth and developmental predictors
of later reading retardation. Lyle and Goyen (1969)
reported that retarded readers would generally be found to
be retarded in all the basic primary-level school subjects.

To determine the effect of extrinsic incentives, Goyen
and Lyle (1971a) provided good and poor readers with either
money and knowledge of results, or neither of these
consequences, on a tachistoscopic shape recognition task.
Finding no measurable effect of the incentives, they then
switched to a learning task (visual-visual paired
associates). Incentives in this experiment (Goyen and
Lyle, 1971b) consisted of money, information on results and
verbal encouragement for one group, and none of these for
the other. Not surprisingly, the incentive group performed
better; however, no differences were found between good and
poor readers, which suggests that visual association was not
an area of particular difficulty for the poor readers.

Forty-six percent of the variance on 9 supposedly
relevant tests was attributable to a "neurological integrity"
factor in retarded readers, in a survey by Lovell and Gorton
(1968). They are careful to discount any implication of
direct causality, recognizing the chicken-and-egg sort of
problem stressed by Vernon (1957, p.7): "again and again we shall encounter the difficulty, if not impossibility, of distinguishing what is the cause and what is the effect of the reading disability". Lytton (1968) provides a further illustration. Following up boys who had benefitted and not benefitted from remedial lessons, he found that the latter showed less perseverance on a reading task, a higher degree of disruptive anxiety, and generalized psycho-physical vulnerability. Lytton also found a tendency for familial incidence of reading difficulty, which he and others have taken to imply the possibility of a pre-disposing neural substrate. It seems important not to overlook, however, the potential of modelling and other environmental effects, to produce the equivalent outcome.

In 20 subjects from a remedial reading clinic compared with 20 controls, Hunter and Johnson (1971) found greater familial incidence of dyslexia, a higher percentage of left-hand dominance, and lower scores on WISC Information, Vocabulary, Digit Span, Arithmetic, Similarities and Coding subtests in the clinic subjects. Although Black's (1971) sample of 100 working-class children, referred to a hospital reading retardation clinic, scored significantly lower than "normal" on the WISC (mean I.Q. 90.7), within the sample there was no evidence of correlation between intelligence and degree of reading retardation.
3. Treatment suggestions

There is widespread agreement with Schonell's (1961, p.233) opinion that "... disability in reading creates a sense of frustration and inferiority unparalleled by any other insufficiency or shortcoming in the individual's life."

The damaging effect of reading failure on self-esteem is attested by most writers on the subject. Reich (1967) describes the poor reader as a discouraged and frustrated, yet easily bored child. The problem of engaging the child's interest and co-operation to apply his energies yet again is thus paramount. Most relevant in the present context is this motivational aspect of the procedure, rather than the technical details of training, whether by hypnosis (Hunsaker, 1970), psycho-linguistics (Harborth, 1968), or fostering "neurological unity" (Delacato, 1959).

Intensive diagnostic measures are usually advocated before lessons begin, not only so that the teacher has a clear idea of the pupil's individual strengths and weaknesses (Johnson and Myklebust, 1967), but so that progress can be continuously assessed (Harris, 1961). The teacher is likely to rely not only on standardized tests, which represent an artificial selection of overall ability (Plessas, 1967), but on more informal day-to-day measures such as graphs of progress, lists of new words learned, rate measures on graded tasks, and so on. Some, if not all, of this information will be shared with the pupil, and
is indeed expected to aid in the recapture of his attention (e.g. Witty's 1949 report on successful 8 weeks literacy course for Army draftees; and Harris, 1961). Feedback then, is seen as a major source of reinforcement for the learning process.

To ensure that the reinforcement so gained is positive, a high degree of success should be guaranteed, by beginning training with material below the pupil's frustration level and maintaining a continuing low error rate (e.g. Gray, Baker and Stancyk, 1969). Rowell (1967) found improved attitudes as well as performance after a remedial programme.

As many pleasant associations as possible should be developed to print. The teacher should be sympathetic and permissive, give praise for effort rather than for performance, transform drill into games as far as possible, and use high-interest, unfamiliar reading materials (Francis-Williams, 1970). Social rewards should also be exploited by providing for public recognition of successful progress and, if necessary, discouraging reinforcement by attention of the reading disability.

The importance of the teacher's having, and presumably communicating to the child, optimistic expectations, is stressed by Hart and Richardson (1970): "Without this, success is unlikely. With it, method can take second place" (p.7).

Reliable and detailed information on the effectiveness of competing remedial methods is extremely scarce. Morrow
(1970) noted that psychodynamic insight counselling has given no evidence of efficiency in the treatment of academic under-achievement, and that by themselves, "enrichment" of the learning environment or the use of programmed instructional materials are inadequate (because in the latter case reinforcers are too weak and not contingent upon the verified occurrence of desired academic behaviours). Hart and Richardson (1970) use Early's (1963) classification of backward readers by treatment required rather than by aetiology; they agree with Vernon (1957, p.185) that the severest cases of reading disability (i.e. those children with gross discrepancies between reading achievement and intellectual capacity, often labelled "dyslexic") have a very dubious prognosis whatever extensive special help they receive. On the other hand milder cases, due to low verbal intelligence, minor sensory defects, bilingualism, poor teaching or interruptions in schooling, usually respond well to appropriate remedial help. (The way that "dyslexia" describes resistance to treatment is undoubtedly a major cause of the retention of the concept). Maginnis (1971) found no evidence for a ceiling effect such that the poorest readers would make the most progress with remedial help; he also noted that "different students gained most when different methods of computing gain were used" (p.322).
4. Implications for the behaviour modification approach

It can be seen that the traditional educational procedures for correcting "reading disability" have involved the following:

(i) extra training opportunities, in the form of special remedial help;

(ii) the use of feedback on progress as a motivating device, and reliance on "natural" or intrinsic rewards for learning;

(iii) social reinforcers (the teacher's approval and encouragement) made contingent upon effort rather than upon results;

(iv) careful avoidance of further failure and criticism.

The practical unavailability of (i) above, due to lack of trained and accessible personnel, has been one reason for the enthusiasm for reading-modification efforts, particularly where success is claimed using remedial tutors or "therapy-technicians" not much more expert than their pupils (see Chapter IV, Section 2).

Reliance upon "intrinsic" rewards for reading proficiency, and the use of information on progress as a spur, obviously succeed with the majority of beginning readers. The addition to this arsenal of material incentives by the behaviour modifier, should not cause interest and feedback to be overlooked as possible reinforcers. Indeed
as concluded from Chapters II and IV, the relative importance of material and inherent consequences of reading activity is by no means clear.

The third point above bears upon the question raised earlier (Chapter II) of what is being rewarded in a token system for reading. It is not the associative link between a visual stimulus and a sound (or a word, or a meaning, etc.), but the learner's attention and continued effort as the programme unfolds, which the rewards aim to maintain and/or strengthen.

The teacher's avoidance of all aversive stimulation during re-training is congruent with the possibility of anxiety-mediated response-decrement discussed in Chapter II. Any criticism or reproof by the teacher, or even any further experience of frustration and failure by the child, could reactivate the conditioned emotional response learned to print and thus impede or prevent any progress in remedial lessons.

Research attention to reading retardation continues to increase, as documented by Otto, Barrett, Smith, Dulin and Johnson (1972). Considerable frustration has, however, been expressed during the last few years over the unsatisfactory pay-offs generated by much expenditure of effort in this field. Wiener and Cromer (1967) pointed out that reading difficulty is not adequately distinguished from overall language
difficulty, and that the level of sophistication and specificity of current conceptualizations of the causes of difficulty remained disappointing. Williams (1970) characterized the area as one of diminishing returns for research, and concluded that more attention should now be given to motivational aspects such as how the child learns to focus attention, delay gratification, persist at a task, and develop interest in achievement and scholarly activities.

It must be admitted that overall success has been slight, despite what Applebee accurately refers to as "the intermittent efforts of many and the dedicated and continuing efforts of a few" (1971, p.91), in predicting or in curing reading disability, or in specifying the antecedents in particular cases.

Summary

Assessments of how widespread reading disability is in the community will depend on the effects of extrinsic factors such as the relative affluence of particular groups, tests used, and the standards set for adequacy (e.g. 'literate' or 'reading level within a year of estimated Mental Age'). One of the most widely-discussed hypotheses concerning individual-specific causes of difficulty in reading is that of a subtle neurological defect ("dyslexia"). Should such a syndrome ever be isolated satisfactorily, one might expect
a reduction in the confusions which presently exists between it and intellectual and emotional handicaps.

Teachers and psychologists with first-hand remedial-reading experience, tend to emphasize the discouraged, apprehensive condition of the backward reader. Their suggestions for successful therapy therefore all focus on ways to motivate the sufferer to participate co-operatively in the retraining programme. Evidence of improving performance, the intrinsic interest of the reading material, and the social reinforcers dispensed for effort, are the means by which the remedial teacher expects to produce and strengthen reading behaviour; punishment is avoided in any form.

While these traditional approaches to remedial reading instruction have valuable contributions to make to an operant attack on the problem, there are limitations of usefulness due to the uneven quality of research, non-empirical foundations for theorizing, and the lack of firm comparative data on success rates and individual prognosis.

The prediction and control of reading behaviour crystallizes many of the difficulties felt by available theoretical approaches, including the operant approach, in dealing with behaviour which is both private and consequential. A description of the author's experimental work follows, in which reference will be made to the studies already discussed, and others.
Chapter VII. Fact-finding

First-hand acquaintance with the behaving organism is often advocated by proponents of the operant viewpoint. Not only is observation sharpened, but serendipitous findings may be made (e.g. Skinner 1961 p.76-100). In the present work, preliminary experience seemed desirable before formal experimentation was begun, particularly in the following two areas:

a) the best organization of form and content of a remedial reading lesson, and
b) an examination of local conditions, with regard to incidence of reading retardation, and suitable measuring instruments for reading achievement.

This Chapter describes efforts made to fill these lacunae.

Study I: The conduct of remedial lessons

Although there is an extensive literature reviewed in Chapter VII on how to plan and carry out remedial reading lessons, specific guidance on how to incorporate a token reinforcement system is rare, and practical experience for the E in mechanics and techniques was regarded as necessary.

At the time of the work reported (mid 1970), Staats, van Mondfrans and Minke's 1967 manual gave the most explicit details on the operation of a token reinforcement reading system. The child is first presented in the Individual Word Phase, with every new word contained in the lesson, on separate cards. The cards are re-presented until errorless word recognition occurs. The Oral Reading Phase
follows, with successive presentations of paragraphs of reading material, again until it is read without errors. The child then reads the passage silently (watched closely by the instructional technician), and afterwards answers comprehension questions, either by writing or by choosing from several alternatives. Tokens of different colours and values are dispensed throughout for correct responses, and back-up by goods chosen by the subject. Graphs and charts of progress and reinforcement earned are kept, and "substantial positive comment and approval" (p.15) delivered by the instructor: "Regardless of Subject's progress during the session, he should be made to feel that he has performed well", (p.15, italics added). Also, "any verbal or non-verbal means of indicating disapproval of failure" are to be avoided (p.16).

As far as can be ascertained from their description of the procedure, Staats et al. do not attempt to follow up specific deficiencies with for example, further examples of the phonic group causing difficulty, or special exercises or drill. Their approach would undoubtedly be classified as "coaching" rather than remedial treatment, by many educators (e.g. Tansley 1967 p.84), and this may be one reason why improvement on standardized tests has usually been negligible or considerably less than improvement on actual samples of the teaching materials, in their research findings.

The reason of course for choosing such an approach is to standardize lessons exactly so that inter-subject comparisons, and rate changes in the same subject over time, become possible.
Loss of transfer of training, which may result when children are not instructed in word analysis, seems to be an automatic risk in that case.

Study I involved E giving 20 individual remedial reading lessons, to two boys whose reading backwardness was of grave concern to their parents. Staats' technique was employed with modifications, mainly to increase the flexibility of lesson content so that phonic training could be included. The various Phases of instruction were not rigorously adhered to; changes such as the introduction of response-cost contingencies, and "stretching the ratio" by increasing token-prices of backup goods, were introduced for exploratory purposes.

**METHOD**

The subjects came to lessons at separate times, twice a week after school. Materials used will be described in greater detail in Chapter VIII; they included phonic exercises and high-interest story material, all of suitable difficulty for the child's beginning standard.

Before each reading task, the subject was told how many "points" could be earned, or about other contingencies such as point-loss for specified errors. E kept tally of these tokens, which at the end of each lesson could be exchanged for goods, including objects requested by the subject and books, or time on or off the next lesson. For the first 10 lessons point value was effectively 1/5¢, thereafter it was decreased to 0.1¢.
Assessment of progress was via the Neale Analysis of Reading Ability (Neale 1958), which provides scores on Rate, Accuracy, and Comprehension of reading. School reports were also used, parents were interviewed and encouraged to make telephone contact as desired, and the subjects were required to write an "essay" about the point system near the end of their lessons.

RESULTS AND DISCUSSION

Results are best presented separately for the two subjects, who differed greatly in both the nature of their difficulty and in their response to the treatment outlined.

a) Subject C

C was an 11 year old boy of average intelligence (WISC IQ 95), who was 2½ years behind in reading. He was passive, inhibited and "too quiet and well-behaved" according to his father. He had very little understanding of the phonic correspondence between letters and sounds, and never, as far as was known, read anything voluntarily. Previous remedial efforts by a specialist teacher and by his grandmother had proved unhelpful and had been strongly resisted by C.

C was awarded points for effort rather than for success at first in order to induce "trying again" behaviour. Later, deductions were made for errors such as confusing p, d, and b, and m and n. Throughout the lessons this subject was compliant although inexpressive. He spent most of the $4.54 worth of points he earned on sweets. Reading Rate showed a very large gain after 20 one-hour lessons, and significant progress had also been made in Accuracy and Comprehension on the Neale parallel forms A and B.
Overall word recognition gain however, was not as impressive as has been reported by other researchers. C's school report showed an improved grade for "oral expression" and the teacher commented that he was showing more self-confidence; whether this was directly or indirectly associated with his participation in the study (of which the teacher was unaware) is of course unknown.

Qualitative aspects of this case were of great interest. C's parents were surprised when he continued to express eagerness to attend lessons, and when they discovered from E that the time was all spent working at reading and not in evasive tactics. After six sessions, the mother reported that, for the first time in his life, C had begun to read out incidental bits of print such as packet labels, street signs and television commercials. After twelve sessions he spontaneously read to his parents from a book, which further increased their enthusiasm.

C's "essay" about the token system expressed mainly his wish for more points, and his enjoyment of the reader. Overall, this child's co-operating behaviour seemed to be well controlled by the reinforcement system, with some evidence of beneficial effects on reading itself. Although no contingency reversal was attempted, the past history of remedial efforts may be regarded as some control for the present manipulations (as Yates 1970 suggests is reasonable in N = 1 studies).

C's family situation led E to speculate about possible
environmental determinants of his disability. It appeared that C's younger sister dominated many activities, and obtained greater attention within the household, by her asthma attacks and need for speech therapy. Without positive means of competing with her, and rejected somewhat by his "tough" police-detective father, C may have in fact been subtly rewarded for his worrying difficulty. In particular, his mother displayed great sympathy and continually identified herself as having had the same problems, adding that they had been solved by marrying.

b) **Subject M**

M was an 8 year old boy regarded as underachieving academically in relation to his ability (WISC IQ 118). He was the fourth of five children, three of whom had received psychiatric help for emotional disturbance. Family tensions were high: the father had deserted and mother was bitter and constantly tired. M was taking tranquillizing and anti-depressant medication; this and chronic ear infections may have contributed to his poor concentration.

M liked reading but was impulsive and inaccurate. A form of response-cost was therefore introduced where errors were penalized by point deductions from an initial lump sum. M earned a total of $3.72 worth of goods during his 20 45-minute lessons; his "purchases" included three books and ten minutes of extra lesson-time. The main gain on Neale tests was a large increase in Comprehension of read material, with a two year decrease in Rate.

At the end of the school year M was graded as worse in every
school subject except reading and composition, than he had been at the midyear assessment. His teacher, who was unaware of his extra lessons, told mother that M was reading more carefully, understanding more, and making more effort to analyse words phonically rather than guess them.

Although M was generally cooperative in the lessons, he was less reliably controlled by the token system than was C. An example was his refusal despite high point-payoff contingency, to write the requested essay or original sentences about everyday topics such as "A birthday party" (which may incidentally have been a traumatic subject for him). In addition, he expressed that he had formed an emotional attachment to the E which made the lessons (but presumably not "reading better") positively reinforcing.

CONCLUSION

Study I indicated both the success and the limitations of token reinforcement, and highlighted practical questions such as how to obtain baselines and assess change, in reading.

The work was meant as a pilot study in the sense that it aimed to provide information about the application of token reinforcement to reading; this goal was achieved. The formal experiments which followed subsequently are described in Chapters VIII-XII.

Study II: Incidence and Measurement of reading retardation in South Australia

As noted in Chapter VI, the definition and measurement of reading retardation is too complex for there
to be generally-accepted norms of occurrence. Findings will and do vary according to the social class and age of subjects, the measures used (whether tied to school or national averages), and also according to whether one includes obvious cases of inadequate teaching, physical or mental disability, prohibitive emotional disturbance, and so on. To gain some information about the South Australian situation was one aim of Study II. In the process, data on the appropriateness of the widely-used Schonell Word Recognition Test (WRT) was also obtained and evaluated.

The other aim was to explore the possible local usefulness of the Gates-McKillop Reading Diagnostic Tests. This collection of individually-administered tests is intended by Gates and McKillop (1962), to provide the basis for "discovering causes of reading deficiency, in terms of the pupil's unique handicaps" (Manual of Directions 1962, p.2). They believe that

"Study of the child's answers, the nature of his errors, and the patterns of his responses should become, with experience, more revealing and significant for diagnostic and remedial purposes than reliance upon numerical scores and rating systems alone".

Tables are provided, converting raw scores into grade score norms, or into ratings of ability as Normal, Low or Very Low, but no details on standardization population, reliability or validity, are given.

Study II surveyed scores of South Australian school children on subtests V - 1, VI - 1, and VIII - 1. These involved
pronouncing nonsense words, recognizing the visual form of a heard nonsense word, and spelling, respectively. The latter two tests were given en masse by means of the school P.A. system; the first individually.

**Reasons for choice of tests**

a) Schonell WRT

Routinely used by the S.A. Education Department Psychology Branch and widely administered within schools for regular assessments of progress, this test has the reputation of being a valid and sensitive indicator of reading prowess and progress for children of this State. Details of original standardization are surprisingly not given by its originators (Schonell and Schonell 1960).

In brief, the child is required to read single words of increasing difficulty until he fails 10 consecutively. At each approximate grade level there is a sample of 10 words, and the test ranges altogether from first grade to third year high school. It takes only a few minutes to administer, is unambiguous in scoring, and the final score is readily converted to a Reading Age convenient for comparisons with Chronological Age, Mental Age, etc.

b) Gates-McKillop nonsense words tests

While there is a variety of word recognition tests available, performance on them will necessarily be affected by prior experiential factors such as exposure to the stimulus-words in the natural, out-of-school environment, specific coaching, and previous administrations of the same
test. One might expect that such factors would be eliminated, and degree of mastery of reading skill demonstrated, if verbal stimuli are presented which are sure to be unfamiliar (because they are not "real", i.e. denote nothing), but which obey the usual regularities of structure found in meaningful words.

Thus the choice of nonsense word tests was hoped to provide a source of information which was relatively more "culture fair", although it is realized that the whole concept of phonic games and nonsense words could itself be "culture-biassed".

c) Gates-McKillop Spelling

It is not uncommon for persons to be able to read, yet not to spell to acceptable standards of accuracy. The reverse situation however does not occur (except possibly in rare brain-damage conditions). Spelling in fact may constitute a severer test than reading of the learned correspondence between a letter or group of letters, and a sound.

The Gates-McKillop Spelling Test was in given in preference to the Schonell (Schonell and Schonell 1960) primarily because, being shorter, the whole test could be given to all the subjects.

METHOD

The Headmaster of a primary school in the Adelaide metropolitan area, who was known to be particularly interested in reading and sympathetic to research thereupon, was contacted and his cooperation obtained for the study to be described.
1. Subjects and materials

The pupils of the primary school in question were representative of the middle socio-economic level in this community, with a few fathers at either extreme of the occupational status hierarchy. Relatively few children (only about 80 in the entire school), had non-English-speaking homes and thus, language difficulties.

One way in which pupils may have differed from a random sample of middle-class Adelaide public-school children, was in the enthusiasm and interest of their Headmaster as regards reading. This man had been very active in grouping the children according to ability for reading lessons (from the honoured and lightly supervised "library" group to the small numbers and intensive instructor-student ratio of the poorest readers), and in organizing volunteer help from mothers to act as reading aides and work with individual children several times a week.

All the children from Grades III - VII inclusive (N = 453) were used to obtain local norms for the Schonell WRT and the Gates-McKillop subtest VI - 1: marking the correct alternative out of 4, of a visual-verbal stimulus to correspond with an auditorally-presented word. As an example, on the eighth trial, the subjects heard "fingle", and had to mark the corresponding word out of "findle vimble fingle fackle". This test will hereafter be referred to as Choosing Nonsense Words (CNW).
2. Procedure and Instructions

a) Schonell WRT

The teachers of each class had within the previous fortnight administered this test individually to their pupils. The standard recommended procedure had been followed, and scores obtained were made available to E.

b) Choosing Nonsense Words test (CNW)

The Headmaster introduced E over the public address system, which was received in the appropriate classrooms. After telling the subjects that they would need a pencil and ruler, and asking them to mark their sheet of the 20 rows of alternatives with their name and grade, E said "Let's look at the lists of words on the paper in front of you. You will see that these are not real words, but made-up words. There are four in every row, and 20 rows. I'm going to say the number of the row, and then a word. I want you to find the word that I said, and put a line underneath it. Be sure to listen very carefully and put your line under just the word I say. If you're not sure which word I said, have a guess and put your line under the one that looks nearest" (with suitable repetition of salient points).

The 20 verbal stimuli were then presented slowly, with one repetition each, being preceded by the relevant row-number.

c) Gates-McKillop Spelling (subtest VIII - 1)

By means of the P.A. system again, E gave the following instructions: "Some of these spelling words will be easy and
some hard. I don't expect everyone to be able to spell all
the words, but I would like you all to try at each one,
if you have any idea of how it might go. Now let's begin.
I'll use each word in a sentence, and then say it again."

The 40 words were then given, alone and embedded in
simple sentences, to all Grades III - V (N = 266).

d) Reading Nonsense Words (RNW): subtest V - 1

After introducing herself to the subject, E said
"Here is a list of nonsense, or made-up words, like the ones
we did on Friday. Can you read them to me? Have a try.
Remember, they are not real words. See how many you
can work out". 30 seconds was allowed for each item.
The scoring of this test was not as clearcut as the other
two, as several of the "words" were ambiguous, i.e. a case
could be made by analogy, for pronouncing them in different
ways (e.g. prible, thasp, clow). Any reasonable attempt
was given credit: e.g. 'clow' could rhyme with 'blow'
or with 'cow', but nothing else, and the initial letter
had to be said as k. Subjects were one class of Grade III
children (N=40).

RESULTS

a) Schonell WRT

Fig. 1 shows the average Chronological Age (C.A.), and
Reading Age (R.A.) derived from this test, for boys and girls
in Grades III - VII, i.e. throughout the primary school.
The correspondence between CA and RA overall is clear, as are
the slight but consistent sex differences. Although girls
Fig. 1: Schonell R.A., C.A., and grade level (N=453).
are about two months younger on average than boys in each class, mean reading performance as measured on this test is superior. The RA - CA difference is 7 - 12 months for the girls, 4 - 6 months for boys.

A possible explanation for these differences lies in the larger proportion of severely retarded readers amongst the boys. If the criterion of "reading retardation" is set at a 2 year CA - RA discrepancy, as is the case when the S.A. Education Department makes decisions about the provision of remedial help, the number of affected children at this school is 24 boys and 11 girls. If the cut off point is set at 18 months instead of two years, there are 36 boys and 16 girls. Thus the commonly quoted ratio of two boys to one girl seriously retarded reader, is supported.

The higher proportion of male sufferers may also account for the age difference within classes, relatively more boys being required to repeat a year's work.

The overall proportion of backward readers is 7.7% with the two year criterion and 11.5% with the more lenient one.

b) Gates-McKillop subtests CNW, Spelling and RNW

78% of Grade VII children, and 75.8% of Grade VI, obtained the maximum scores on the CNW test. The ceiling was therefore judged to be too low for children of that age. Fig. 2 shows how the average scores on the three tests varied with grade level, and the corresponding American norms. Without more information about the American sample, the
Fig. 2: Gates-McKillop subtest scores by grade-level, and American norms (N=266)

Test Score

Grade

III

IV

V

n=79

n=95

n=92

- CNW test
- Spelling
- Amer. norms

RNW test: S.A. mean •
Amer. norm ★
significance of the South Australian children's better performance on Spelling and RNW, and poorer performance on the CNW test, is difficult to assess. It does appear however, that the published norms could not be applied with confidence to the local population.

c) Correlations between scores on different tests

Table I shows Pearson product moment correlations between scores on Schonell WRT, and the three Gates-McKillop subtests.

Table I: Correlation between scores on four different reading-related tests

<table>
<thead>
<tr>
<th>Grade</th>
<th>n.</th>
<th>WRT/Spell</th>
<th>WRT/CNW</th>
<th>WRT/RNW</th>
<th>Spell/CNW</th>
<th>Spell/RNW</th>
<th>RNW/CNW</th>
</tr>
</thead>
<tbody>
<tr>
<td>III</td>
<td>girls</td>
<td>21</td>
<td>.80</td>
<td>.42 (p&lt;.05)</td>
<td>.58</td>
<td>.33 *</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td>boys</td>
<td>19</td>
<td>.86</td>
<td>.44 (p&lt;.05)</td>
<td>.74</td>
<td>.34 *</td>
<td>.70</td>
</tr>
<tr>
<td>III</td>
<td>girls</td>
<td>21</td>
<td>.81</td>
<td>.53 (p&lt;.01)</td>
<td>.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>boys</td>
<td>18</td>
<td>.65</td>
<td>.13 *</td>
<td>.29 *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>girls</td>
<td>43</td>
<td>.86</td>
<td>.45</td>
<td>.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>boys</td>
<td>52</td>
<td>.84</td>
<td>.43</td>
<td>.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>girls</td>
<td>40</td>
<td>.81</td>
<td>.60</td>
<td>.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>boys</td>
<td>52</td>
<td>.87</td>
<td>.42</td>
<td>.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>266</td>
<td>.88 (p&lt;.001)</td>
<td>.54 (p&lt;.001)</td>
<td>.67 (p&lt;.001)</td>
<td>.55 (p&lt;.001)</td>
<td>.69 (p&lt;.001)</td>
<td>.17 *</td>
</tr>
</tbody>
</table>

Correlations unless otherwise marked are significant at p < .005 (one-tailed test).

Asterisks signify non-significant r's.
There is, as expected, a strong relationship between word
recognition and spelling at all ages and for both sexes.
The other correlations indicate the overlapping nature of
the skills tested. Somewhat unexpected was the weakness
of the relationship between ability to pronounce, and to
discriminate the visual form of, nonsense words. Reading
real and nonsense words however were closely related skills.

DISCUSSION

1. Incidence of reading retardation

Inasmuch as score on the Schonell WRT appears to
increase steadily with age in an unselected population,
\( (\text{Pearson } r = .61, p < .001) \) it seems a suitable test of
reading achievement for the purposes of this research. No
data was collected on reliability. The Headmaster felt
that with a three month interval between trials, and with
the non-correction method of administration of the test,
scores show little change due solely to practice. A
split-half analysis seems inappropriate when the number
of responses sampled is already so small. Validity data
is equally difficult to produce. In this school, the
reading grade awarded at the middle and end of each year
was based on Schonell WRT performance, so an independent
skill-assessment was not available.

The incidence of backwardness seems comparable to the
only other known survey conducted within this school-system.
Berndt (1972) found that 14% of 465 children, who had from
intake had well-trained teachers using eclectic methods, were rated as "below average in reading" by their teachers at the end of Grade III. Of these, half were in fact, "slow learners", whose word recognition was equal to their comprehension on the Neale Analysis of Reading Ability. The remainder, 7.1% of the total population, appeared to have literally a "specific learning disability". Their IQ on the Wachsler Intelligence Scale for Children ranged from 92 - 129 (mean 108), and Schonell Reading Age was on average one year below CA. Five times as many boys as girls were affected.

The present study did not separate out the backward readers from the generally backward pupils. In addition, no control for quality or consistency of teaching was possible. The age range (8 - 12 years) was a further difference from Berndt's study (5 - 8 year olds).

2. Comments on usefulness of the Gates-McKillop subtests

Without much more information on both the original norm group, and the responses of a more representative sample of S.A. children, it would clearly be hazardous to use the Gates-McKillop subtests for other than its original purpose: the individualized diagnosis of difficulties, and by other than its intended user: a teacher with much experience not only of reading problems, but with the test itself.

Some of the changes which would appear desirable should the test be used as a screening device, or for a full-scale
validation, include re-arranging the form of the CNW subtest so that the correct answers are distributed more evenly across columns, acknowledging the low ceiling (about 9 years) of both nonsense words tests, and making the RNW stimuli less ambiguous. The Spelling subtest appears to be a useful measure throughout the primary school.

An extensive standardization programme for Gates-McKillop subtests would be required before this test could be more widely used. However Study II indicated that the Schonell WRT was worthwhile as a means of assessing the reading standard of available subjects.

Study III: Frequency of Print Behaviour

As noted in Chapter IV, very little is known about the natural frequency of reading responses when the organism is unrestrained. Also unknown is the relationship if any, between competence and frequency of reading. Perhaps the child who scores high on a word recognition test is one who has read widely beyond the requirements of the teachers. Perhaps good readers enjoy reading and therefore do more of it, and bad readers avoid print due to its association with aversive stimulation such as experience or apprehension of failure.

Study III aimed:
1. to gather information on amount of voluntary reading by primary school children.
2. to relate this if possible to some measure of reading ability.
METHOD

1. Materials

A questionnaire was devised, as follows:

Q1. What do you like to do in your spare time?
   (maximum of three activities to be noted. If child does
   not include reading, ask Q2).

Q2. Do you ever decide to do some reading?

Q3. How do you feel about reading? Which one of these
   would be true for you:
   
   - enjoy reading greatly
   - quite like reading
   - don't like reading very much
   - hate reading (Child to mark one.)

Q4. When do you do most reading (that nobody tells you to)?

Q5. What do you read?

Q6. On an average weekend, how much time would you spend
   reading?

The last question was the vital one, the others being
largely intended to induce thought about attitudes to, and
frequency of, reading.

The measure of reading ability chosen, was a revised
version of Gates-McKillop RNW test. The principles of
revision were

a) to make the test harder and therefore suitable for children
   older than Grade III. This was done by introducing a greater
   variety of phonetic combinations, including some less
   frequent ones.
b) to remove duplication of phonic groups.

c) to reduce the number of possible pronunciations of each item, as far as possible.

d) to order the items roughly according to difficulty.

The final list used was:

<table>
<thead>
<tr>
<th>trage</th>
<th>pharp</th>
</tr>
</thead>
<tbody>
<tr>
<td>cled *</td>
<td>grebble</td>
</tr>
<tr>
<td>swink</td>
<td>ceivit</td>
</tr>
<tr>
<td>twock</td>
<td>clow *</td>
</tr>
<tr>
<td>grumb</td>
<td>stupe</td>
</tr>
<tr>
<td>frable *</td>
<td>lawk</td>
</tr>
<tr>
<td>dween *</td>
<td>wrell *</td>
</tr>
<tr>
<td>slidge *</td>
<td>thew</td>
</tr>
<tr>
<td>troab</td>
<td>dright</td>
</tr>
<tr>
<td>plurk</td>
<td>floy *</td>
</tr>
</tbody>
</table>

(Asterisks indicate items taken from the Gates-McKillop subtest V - I.)

This test had a much higher ceiling than alternative measures such as "letter-sounds given correctly out of 26", or "correct letter-sounds given per second", which had been considered.

The items were all equally unfamiliar to all the subjects. The test was therefore free from the effects of specific teaching: it measured an aspect of reading assumed to be an important component of the skill, namely, the ability to analyse novel visual stimuli in terms of learned regularities of written English (i.e. to generalize stimulus control).
2. Subjects

One whole class of middle-socio-economic status Grade V children were used initially as subjects.

3. Procedure

The subjects were seen individually. After asking name and age, E began immediately on the questionnaire. Every effort was made not to provide any feedback, verbal or non-verbal, as to the rarity or acceptability of the answers obtained, in the interests of minimising experimenter-bias effects. The nonsense words were then presented, subjects being encouraged to guess if in doubt.

RESULTS

Fig. 3 shows the pattern of answers to the question "On an average weekend, how much time would you spend reading?", and number correct on the nonsense words test. It can be seen that there appears to be little relationship between the two measures, and certainly not a significant positive correlation between time spent in voluntary reading, and word-analysis skill. Means are respectively 1.8 hours, and 14.9 correct pronunciations.

DISCUSSION

The insensitivity of the natural frequency measure, i.e. the child's estimate of "time spent reading on an average weekend", was a serious but unavoidable weakness of Study III. From the answers to Q4 it is evident that reading is very often a fill-in activity: something to resort to when nothing else exciting is going on, or when sick or waiting to
Fig. 3: Reading frequency and competence
\((N = 36)\)

Reading on an average weekend (hours)
fall asleep. Presumably much of the stimulatory role formerly held by books and even comics has been usurped by television. Time spent reading is therefore likely to fluctuate quite widely depending on extraneous factors, as well as on whether the current library book is regarded as "good" or not.

A further important factor in the influence of reading practice on reading competence, is what the children read. The level of difficulty of reading material must be assumed to affect skill-facilitation. The child who reads very easy material, or indeed the one who reads more difficult books but "skips the hard words", is unlikely to benefit with increased accuracy. They may on the other hand, if receiving intrinsic reinforcement from reading as an activity, be more likely to pay attention and make effort in reading lessons at school. The difficulty of the material, the child's comprehension of it, and his enjoyment of the activity, are therefore all pieces of information needed to make a meaningful interpretation of the present data - but all require a more elaborate and individualized design to elucidate their role.

Children's reading-time estimates, (and none others are easily available or likely to be more reliable), may suffer from over-dependence on recent memory, or from distortion in the direction perceived as approved by the examiner. That the E was not in this study regarded as neutral in attitude to reading, was likely in view of other reading-
related activities being pursued by her at the same school.

On Q3, 21 of the 36 subjects said that they "quite like" reading, and a further 13 said that they "enjoy reading greatly". A continuous scale rather than discrete categories might have increased the scatter usefully.

Those who rated their attitude to reading most favourably, spent an estimated average of 38 minutes extra in reading per weekend, compared with the mildly positive category; average words read correctly were 15.1 and 14.5. Of the two boys who marked that they "don't like reading very much", one got all the nonsense words right and the other 15, and both read for half hour per weekend.

The relation between achievement in reading real and nonsense words was found to be quite close for the Grade III children of Study II, \(r = .67, p < .001\). Any extension of Study III would have to include an assessment of conventional word recognition ability. However such an extension seems more suited to the long-term collection of formal baseline data, perhaps enlisting suitably briefed parents as observers, and perhaps in a clinical situation, rather than to the survey approach first attempted.

CONCLUSIONS

A pilot study and two preliminary surveys have been reported, in which first-hand experience was provided for the E, and data gathered on measuring instruments of reading skill and behaviour, which were appropriate to the local environment.
Norms for the Gates-McKillop tests were found to be inapplicable to the local subject population. The Schonell Word Recognition Test, however, proved a useful assessment tool for reading skill in this group, and the rate of reading retardation discovered was similar to that commonly reported elsewhere (see Chapter VI).

An attempt was made to explore the relationship between the child's voluntary reading behaviour, and skill in pronouncing artificial "words". Refinements were, however, clearly necessary in the measurement of the former variable in particular; such refinements seem likely to require the active and vigilant cooperation of parents.
Chapter VIII. Experiment I. Effects on retarded readers' progress of extra teaching, tokens, kind of backup reinforcement, and response-cost for errors.

AIMS

The studies reviewed in Chapters III, IV, V and VI, did not provide unequivocal answers to questions such as the nature of the optimal reinforcer for complex human learning, especially as exemplified by the reading task, the effect of systematic penalties for incorrect responses, whether the main effect of tokens is informational or motivational, and how tokens compare in efficacy with non-token approaches.

Using technical information gained from the preliminary studies, Expt. I aimed to investigate such questions. The specific goals were to obtain information on
a) tokens vs no tokens in the remedial reading lesson.

b) comparative effects of response-cost for errors, and no response-cost.

c) comparative effects of "consumable and manipulatable" backup reinforcers, and informative but materially worthless consequences.

General considerations

Just as the attainment of literacy spans several years, so attempts to treat failures of the process must span more than a few lessons. The behaviour requiring alteration must be re-shaped by repeated trials, (in contrast for example with an overt motor behaviour such as disruptiveness, where the alternative is already part of the total repertoire and merely requires strengthening).
Regular or protracted access to primary-school children is only possible with the co-operation of educational authorities, unless the researcher is willing to compete with family and leisure activity for the small amount of waking time left after school each day. A few highly motivated parents might agree to this, but even then advantage may be lost by not seeing their children in the usual academic environment.

Headmasters in practice determine, in the writer's State of residence, whether and when their pupils will be made available as research subjects. Naturally a prime consideration with them is, not only the advancement of human knowledge, but the welfare of individuals in their charge. Thus they usually ask, especially where a long-term programme is contemplated, whether or not it will benefit the children concerned. (The phrase "used as guinea pigs" seems to have a high probability in preliminary negotiations). The researcher therefore is under considerable practical as well as ethical pressure to assume a remedial teaching role rather than perhaps concentrating on behaviours which, although easily defined and measured may be largely irrelevant to improved academic performance. Such a commitment to permanent change places restrictions on the usefulness of the ABAB design, for a start.

Further deviations from laboratory ideals are necessary when dealing with a complex multi-phasic cognitive skill. Unfortunately, the "task" cannot be held constant across subjects: reading is a collection of heterogeneous abilities,
in any one of which a "disabled reader" may be inadequate. Some backward readers do not know even the regular sound-equivalences of the letters; others read accurately but so slowly that they lose the thread of meaning; others "read" fluently but accuracy and comprehension are both poor. Several consequences follow from this:

(a) on any single measure of skill, some retarded readers may be near the ceiling for their age, while others have much greater room for improvement.

(b) each subject will progress through work of "graded" difficulty at individually determined pace (for pedagogical efficiency), and is likely to begin at a different point in any series of steps towards skilled performance.

(c) the difficulty of the task is likely to vary unevenly over sessions, for a given subject

(d) the difficulty of the task will vary across subjects even of the same overall "reading-score", for any given passage or exercise.

(e) the "teacher" must manipulate difficulty so that it remains at a roughly constant level for every subject, despite the occurrence of learning in particular areas.

Other practical difficulties of "real-world" research, into reading, have been discussed elsewhere (i.e. in Chapter IV). The solutions attempted to these problems in the present study will be described where appropriate.
METHOD

1. Design:

The experimental design was a fixed effects 2 x 2 analysis of variance, with two control groups. Each subject received 20 bi-weekly individual reading lessons.

a) Variables

(1) Response-cost

Two experimental groups had tokens deducted from those they had earned, for errors (the D. groups). The ND groups never had tokens deducted.

(2) Kind of reinforcement

Two experimental groups (the B groups) could exchange their tokens for a selection of sweets and small toys (the "back-ups") at the end of each lesson, at a rate of 10 tokens approximately equal to 1č worth of goods. The NB groups had coloured paper stars pasted to a sheet of paper at a rate of 1 per 15 tokens. The subject could choose the colour of the star, and could, at the completion of the lesson-series, choose whether or not to retain the page of stars.

Tokens took the form of "points" accumulated on a counter in front of the subject, and awarded or deducted by E pressing telegraph keys (see figure 4). A technical diagram of the apparatus will be found in Appendix A.

(3) Token system vs none. One control group (TC's)
Figure 4: Apparatus and experimental situation
received teaching from the E, including moderate social reinforcement and feedback on results, but no tokens were ever delivered.

(4) Remedial lessons vs. no remedial lessons. A second control group (NTC's) received no teaching from the experimenter.

b) Experimenter

All the work to be described was carried out by this author, who had worked clinically with children beforehand but had not had formal teaching experience.

Adequate monitoring of subject responses necessitated individual sessions.

2. Procedure

(a) Subject selection

To avoid unnecessary confusion between reading difficulty and language difficulty, subjects were sought who were all native English-speakers. As for practical reasons the subjects all had to come from the same school, a middle-class primary school was selected, and the Headmistress approached.

It was hoped to minimize the effect of variability in intelligence by excluding subjects who could be regarded as below average in intelligence. Also, it was regarded as desirable to exclude any cases of gross neurological defect, and the possibility of "specific dyslexia", due to delayed maturation of other causes. For these two reasons children past the first few years of schooling were
chosen as suitable. The expectation was that significant intellectual retardation would by the age of 8 years be apparent to the teachers, and that also by that age any minor cerebral dysfunction or immaturity would have been overcome or compensated for. The subjects could then all be regarded as physically able to learn.

The teachers of Grades IV, V and VI, at the school selected, were asked to provide lists of children whom they thought would benefit from remedial reading lessons. It was thought that teacher-selection of the original subject-pool would decrease the likelihood of their individual subversion of, or refusal to co-operate with, the planned programme. Such obstacles must be regarded as having a probability greater than zero: it is understandable that teachers, many of whom feel under-paid and under-valued, should resent investigation of their apparent failures by "ivory-tower" interlopers.

The children suggested as candidates for remedial reading lessons by the teachers (37 cases) were screened, and final selections made, as follows:

a) the Schonell Word Recognition Test (Schonell & Schonell 1960), was administered to all. This test had been found in Study II (see Chap.VII) to reflect fairly the reading competence of South Australian schoolchildren; it is in fact the measure relied upon by the State Education Department.

As the potential subjects were mostly of middle and upper-
middle socio-economic status (according to a classification of Adelaide suburbs contained in the Karmel report 1971 p.603), and because reading standards are known to vary with this factor, no rigid criterion of a CA-RA difference was maintained for "reading retardation". The teacher's definition of the child as needing help, was accepted in preference to some arbitrary degree of discrepancy. Children were excluded however if RA exceeded CA regardless of how high Mental Age had been judged to be, relative to CA (4 cases).

b) Children known to have non-English speaking homes, or difficulty in speaking or understanding English, were eliminated (3 cases).

c) An attempt was made to screen for intelligence using the Ravens Coloured Progressive Matrices, in a group administration. Apart from the fact of undoubted cheating, the results were judged unreliable because of the age of the children. The Otis IQ test taken in a group by Grade V each year at this school was thought unsuitable due to bias of results by reading skill. Thus efforts to obtain valid estimates of intelligence without administering individual tests such as the WISC, were dropped. The students' confidential record cards were however examined for details of past progress. Six potential subjects who had either failed Grade I or II without this being attributed to illnesses, or who had scored extremely poorly on either group or individual intelligence tests,
were allocated to a "Reserve" corps for use only in emergency. In fact two of these children later became subjects, replacing others whose parents moved interstate after their 6th and 13th lessons respectively. The replacement in each case matched the original subject as closely as possible in age and degree of reading retardation, sex, and grade level, in that order of priority.

Twenty-four children remained after the selection procedures just described. They were allocated to groups of four on the following principles.

(1) distribution of the 7 girls as evenly as possible. Particularly in all measurements of verbal ability, girls are generally superior to boys at this age (Tyler 1956, p.249), so what control was available for sex differences, was used.

(2) grade level: modal composition was 1 each Grade IV & Grade VI, 2 Grade V children. Only one group was different, consisting of 3 Grade V and 1 Grade VI children. The group deviant for sex (i.e. with 2 girls instead of 1) was not also the group deviant in grade-composition.

(3) the three children whose reading achievement fell most conspicuously below attainment in other subjects were put into different groups.

With a small number per group, matching on variables relevant to the effect of experimental procedures seemed a wise course. CA, RA and CA-RA discrepancy were all
matched across groups, and 1 x 6 analyses of variance on these measures showed no significant differences (F's < 1).

The 6 groups were then assigned to treatments thus:

(1) one group containing one of a pair of sisters, had to be the untaught control group (NTC) because of the high risk of result-contamination by communication and comparison between subjects, of differing treatments.

(2) The three groups containing the children who gave some evidence of "specific learning disability" (see (3) above) were all made experimental groups.

(3) of the two groups remaining as candidates for the taught controls (TC's), that one was selected which contained the brightest child. The purpose was to weight group-allocation against the hypothesis that tokens would aid learning.

(4) Random number tables were then used to assign the remaining groups to the four token-experimental conditions:
   - No (extrinsic) backups, No deductions (NBND)
   - No backups, Deductions (NBD)
   - Backups, No deductions (BND)
   - Backups, Deductions (BD)

b) Lesson programme

Subjects came to the school bookroom for two lessons a week on either Mondays and Thursdays, or Tuesdays and Fridays, always at the same period of the day. The experimental space was a small concrete-floored room lined with shelves of texts and stationery, (see Fig.4). A child-sized table, two chairs, and radiator were made available; lighting and
ventilation were good.

Lessons lasted the conventional 40 minutes, therefore one subject was seen in each of the seven periods of the school day, for four days each week for the first 10-week batch of 14 children, two days a week for the second 10-week batch of six. Missed lessons due to absences were made up; if subjects failed to arrive for the lesson, they were reminded by the E after five minutes.

The possibility of diurnal variations in learning efficiency was recognised, but little reliable information is available. It was hoped to minimize any such effects by keeping the time-of-day constant for each lesson a subject had, and by comparing him with his own record. Measures such as seeing all subjects only before the lunch break, would have increased the running time of the experiment to such an extent that natural variations in the school year such as nearness of long holidays, might well have outweighed the possible diurnal variations.

c) Form of lessons

The materials used in training sessions were the Essential Read and Spell Series (Schonell & Schonell 1968) and the Dolphin readers (Taylor 1965). The former consist of graded reading tasks, each "lesson" including phonically-similar new words, a short passage built around them, and related exercises. Examples of this material appear in Appendix B. Both sets of material were chosen on the basis of recommendations in Hart & Richardson (1970).
Half the lesson was spent working on a Schonell unit, and
half in the subject's reading aloud from a self-chosen story
in the Dolphin series of approp. difficulty. Here there
was some vocabulary extension and phonic drill, but the
main purpose was for subjects to read expressively, with
understanding, and hopefully with enjoyment. The E corrected errors,
supplied words and meanings, asked questions, joined in
dialogues etc. as appropriate.

Point of beginning and rate of progress through the
graded material (and promotion from one level of difficulty
to the next), were determined according to the individual
subjects' original capacity and response to the extra
lessons. E aimed to keep the error rate constant at
about 5% of the Schonell passage. Every effort was made
to standardize social interactions and reinforcers and
maintain them at medium intensity. Praise was dispensed
for effort and successive approximations to desirable
levels of performance and where relevant, was associated
temporally with token-award and -exchange. Emotional
involvement with the subjects was avoided as far as was
consistent with a warm encouraging relationship, and with
the fact that a few of the children suffered obvious
conflicts and anxieties.

Feedback on results and progress was similarly standardized.
Each subject in the five taught groups had a graph on which
number of errors, and time taken, were recorded each
lesson for the Schonell passage. The "time" measure was
a compound of actual time taken plus five seconds extra for
every error. This was not altogether a satisfactory
measure, as it took no account of the variation in passage-
length, but had the great advantage of being immediately
calculable. Praise and where appropriate tokens,
were awarded for downward movements of the graph (i.e.
indicating greater fluency). Subjects were encouraged
to set themselves specific short-term goals, e.g. "to
get the graph below the two minute line", or "to keep
the errors less than five", as well as aiming to improve
performance on the regular spelling revision tests,
finish a reader in n more lessons, etc.

d) The token reinforcement system

Lessons were the same in form for all five taught
groups, and their content was determined by constant
principles as already described; however, the TC group
received no points or other "contrived" reinforcers and
the point-apparatus (keys, transformer and counter), was
not visible during their sessions. Instructions for the
four experimental groups were as follows:
a) (for all): "It's your job to get points. See how many
points you can get, each lesson. I'll tell you how many
you can get for each thing."
b) (for D groups, add): "You lose points for getting
things wrong. I'll tell you how many and for what. You
have to see how many you can collect, each lesson."
c) (for B groups, add): "At the end of each lesson you'll
be able to use the points you've earned, to buy things with. Here are the sorts of things you'll be able to buy with your points, and here is a list showing how many points you need for each thing." (Demonstration of backups collection, "price list", etc. as shown in Fig. 5. The S's could request items to be added to the backup collection, and could also "save up" points from lesson to lesson for more expensive items.)

d) (for NB groups, instead of c), say): "At the end of each lesson you'll get stars according to how many points you've earned. We'll keep a page with your stars pasted on it, here with your graph. You can take it away with you at the end of the lessons if you like." (NB: the stars for each lesson were not pasted in separate rows, in order to avoid endowing them with extra informational value.)

In the first session with each S, E said, "I don't discuss your lessons with anyone else, and I'd rather that you didn't either; they're just our business - O.K.?" As will be shown later, this request was apparently sufficient to preclude detailed comparisons of lesson conditions between subjects, and ensure a remarkable degree of inter-group ignorance.

e) Measurement

(i) Pre-treatment measures

Schonell Word Recognition test and Dominion word list (Ontario College of Education 1947): 100 words including most phonic groups, less steep difficulty gradient
Figure 5: Backups and stars.
than the Schonell.

Fluency measure (time taken and number of errors) for the first passage in Schonell reader of appropriate difficulty according to Schonell RA.

Teacher's ratings of child's classroom behaviour, interest in reading, and reading ability, relative to both the child's classmates and his performance in other academic areas.

(ii) Post-treatment measures

The above measures were repeated exactly, and the teachers made additional ratings, of willingness to attend lessons, and of any changes in the child's reading, spelling, other subjects, attitude to schoolwork, or attendance.

Subject questionnaires were completed on attitudes to lessons (by semantic differential), to reading (direct multiple-choice question), and knowledge of other experimental conditions. (See Appendix B for copies of rating scales and questionnaires. Scales for the semantic differential were chosen from Osgood, Suci & Tannenbaum (1957)).

"New word" tests were given on all the words from the Schonell lesson list which were unknown at first presentation in lessons 10 - 19.

(iii) Long-term follow-up

At the end of the school year, i.e. 3 - 6 months after the last lessons, the Schonell Word Recognition Test was re-administered to all subjects. Rate of improvement could then be calculated.
RESULTS

1. Attendance and co-operation

All subjects attended all scheduled lessons except when absent through illness, and none were unco-operative during lessons. No systematic records were kept of tardiness: some subjects were routinely reminded by a predecessor in the next classroom; others were reminded or delayed by teachers; sometimes the activity they left was of extraordinary interest.

2. Pretreatment homogeneity

The six groups did not differ initially in RA ($F = .089$), Dominion score ($F = .388$), or Fluency (words per second with five second allowance for errors $F = .694$, correct words per second $F = 1.062$). As previously noted groups were also matched for degree of reading retardation of the subjects.

3. Treatments received

To see whether all taught groups had been required to work equivalently hard, the distance progressed through the Essential Read and Spell books was expressed for each subject in years of Reading Age. A $1 \times 5$ AOV showed that no favouritism in this respect had occurred ($F = .168$).

Among the groups receiving token reinforcement, Deduction groups received significantly fewer points on average per lesson than did ND groups ($F = 6.4041, p < .03$). The average difference amounted to 5.5 points per lesson, or an accumulation of 11 cents worth of goods, or 7 stars, over the whole 20 lessons.
Average subject "earnings" over the lesson-series were $1.35 (worth) for Backup groups and 90 stars for the NB groups. Half the NB subjects decided at the close of treatment to keep their page of stars; they belonged evenly to D & ND groups.

That there had been few if any breaches of confidence about treatments undergone, was ascertained from Question 3 of the Post-treatment Subject Questionnaire. Although subjects claimed to know others having special lessons, the "differences" they cited referred to work-books rather than reinforcers.

4. Teaching effects

To compare the No Teaching Controls with all other groups, Dunnett's procedure for comparing means with a control was applied (Winer 1970 p.89).

On the Schonell WRT, the NTC's improved significantly less (p<05) over the period of treatment, than did the two best taught groups (see section 6). On the other word recognition measure, the Dominion test, pre-post gains were again significantly smaller for NTC's than for the best two, and very nearly the best three (critical interval 7.50 instead of needed 7.75), of the taught groups.

The NTC's did not differ significantly from the taught groups on overall Schonell gains, i.e. from before treatment to the long-term followup 6 - 9 months later, or for the period between the end of treatment and the longterm FU.

Fluency changes on the same passage pre- and post - the treatment time similarly failed to reach significance (NTC mean change 1.25 words/second, mean change of taught groups
5. **Token effects**

Dunnett's test was again used, to compare the Token Control group (TC's) with the four other taught groups, to which the token system had been applied.

Not once did differences reach statistical significance. Further, the TC's were far from being always the least improved group. On the word recognition measures they equalled the best token groups; on semantic differential ratings they were intermediate in their judgments of lesson pleasantness and potency; on fluency measures, the New Word test and a count of cumulative errors, their performance was the poorest. (Details of means etc. will be found in Appendix C).

6. **Response-cost**

The deduction of points from amount previously earned, contingent upon inaccurate responses on a variable schedule, affected word recognition achievement. 2 x 2 analyses of variance showed that on the Schonell WRT, groups which had received No Deductions (the ND groups) improved more (F = 12.5, p < .005) than did Deduction groups, over the course of the treatment. On the Dominion test, D groups improved more (F = 8.9, p < .01) than did the ND groups. The difference between these two tests lay in conditions of administration. The Schonell WRT was used to screen and match subjects, therefore in the post-treatment assessment was again given under No-token conditions.
(after apparatus ostentatiously switched off, and suitable remarks).

By contrast, both pre- and post-treatment administrations of Dominion test occurred under token conditions (i.e. contingent award, and deductions or not).

7. Kind-of-reinforcement effects

Material backups (B groups) vs paper stars (NB groups) affected the subjects semantic differential ratings of how pleasant ("sweet" and "happy") the lessons were: lessons were rated as less pleasant by the B groups ($F = 6.72, p < .03$).

A similar but non-significant trend was found for ratings of potency, ("large" and "strong": $F = 2.33, p = .15$).

Although the reinforcement variable did not influence any measure of progress during the treatment, there was a suggestion that it may have had a "delayed effect": B groups improved more per month than did NB groups, from pre-treatment assessment to long term follow-up on the Schonell WRT, ($F = 3.73, p = .07$). Consistent with this finding was the relatively greater (but not statistically significant) improvement of B groups compared with NB groups, in the period between end of treatment and follow-up (mean gain of B groups 1.13 words per month, against .13 words per month for NB groups).

8. Teacher's ratings of improvement

The five teachers who had pupils in the programme (from 1 - 8 children each), differed subjectively along personality and attitudinal dimensions. They were not told which children were receiving which treatments until
the study had finished (although it was obvious in the case of
the NTC's), but no check was possible on how much they
actually knew, which might have biased their ratings
unpredictably: for instance objections are often expressed
to "bribing" children.

Because of the small subject numbers and the use of discrete
categories in the teachers' rating scales (Appendix B),
statistical analysis was not considered appropriate.
Tables 2 & 3 summarize the pre-post differences obtained,
and Table 4 presents data from the post-treatment only
ratings.

Table 2: Changes in teacher's ratings of subject's
disruptiveness, interest and ability in reading,
during the treatment period, (algebraic sums for
each group).

<table>
<thead>
<tr>
<th>Group</th>
<th>Classroom behaviour</th>
<th>Interest in reading</th>
<th>Ability in reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDNB</td>
<td>+1</td>
<td>+1</td>
<td>+2</td>
</tr>
<tr>
<td>BND</td>
<td>0</td>
<td>0</td>
<td>+1</td>
</tr>
<tr>
<td>DNB</td>
<td>0</td>
<td>+1</td>
<td>0</td>
</tr>
<tr>
<td>BD</td>
<td>0</td>
<td>-2</td>
<td>-1</td>
</tr>
<tr>
<td>TC</td>
<td>0</td>
<td>+2</td>
<td>+1</td>
</tr>
<tr>
<td>NTC</td>
<td>0</td>
<td>+2</td>
<td>+3</td>
</tr>
</tbody>
</table>
Table 3: Treatment effects over all 6 groups, as rated by teachers.

<table>
<thead>
<tr>
<th></th>
<th>Class behaviour</th>
<th>Interest in reading</th>
<th>Ability in reading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average or better</td>
<td>disruptive</td>
<td>none</td>
</tr>
<tr>
<td>Pre</td>
<td>18</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Post</td>
<td>19</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4: Post-treatment rated improvements in other areas of behaviour.

<table>
<thead>
<tr>
<th>Gp</th>
<th>NDNB</th>
<th>BND</th>
<th>DNB</th>
<th>BD</th>
<th>TC</th>
<th>NTC</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Spelling</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Other subjects</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Attitude to school-work</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Attendance</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Willingness to attend extra lessons</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>11</td>
</tr>
</tbody>
</table>
Table 5 shows change in subject's affective attitude towards reading as elicited by the direct question "How do you feel about reading?"

Table 5: Subjects' attitudes to reading (rated by subjects).

<table>
<thead>
<tr>
<th></th>
<th>NBND</th>
<th>BND</th>
<th>DNB</th>
<th>BD</th>
<th>TC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
</tr>
<tr>
<td>Negative</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Neutral</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Positive</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

A generalized positive effect was thus attributed to inclusion in the study by both teachers and participants. The possibility of Hawthorne, cognitive dissonance and expectation biases is so strong (and the ratings so unrelated to objectively-measured changes), that limited validity is felt to inhere in this data.

DISCUSSION

1. Summary of statistical conclusions

   (a) Word recognition was influenced for the better by response-cost for errors while the token system was operating; when it stopped those who had not been penalized for errors made fewer errors than those who had.

   (b) Kind of backup reinforcement influenced attitudes, those who had material backups finding the lessons less pleasant than did those who had received "only" stars.
There was a suggestion of a delayed effect of material reinforcement, improving word recognition after treatment stopped.

(c) There was no evidence that tokens aided the learning process when compared with a no-token procedure of equal informational and social characteristics.

(d) 20 remedial lessons produced improvement in word recognition while treatment was being given (but not over a longer period of 6 - 9 months), compared with not having remedial lessons, for about half the subjects depending on the measure used.

(e) There was no evidence for co-variation in different measures of reading skill, and no evidence for the susceptibility of fluency to influence by external manipulations.

(f) Teachers and subjects demonstrated a generalized and mildly positive response to the programme.

2. Interpretations

The relative inefficacy of the token system contrasts with the favourable results usually reported in the literature (e.g. Chapter IV). The fact that the subjects received exactly comparable lessons, feedback and social reinforcers, and that they were unaware of the extrinsic reinforcers and apparatus—novelty provided for other groups, may contain at least a partial explanation. Children from a middle-class background could be expected to be under considerable pressure to achieve academically; possibly
then they need no further motivation than supportive individual lessons with systematized knowledge of results. O'Leary & Drabman (1971) have commented on the paucity of published negative evidence on token efficiency, but quote several examples themselves. Perhaps a myth, or at least an incomplete picture, has been fostered by editorial bias.

It is conceivable that the token groups gained in some ways which were then nullified by other aspects of the procedure. This conjecture would be supported by the poorer performance of the TC's on all "in-lesson" measures (which never reached statistically reliable levels), despite their good showing on the pre- and post-treatment assessments.

Indeed, further statistical analysis of the data for all 24 subjects shows that the different measures of reading ability which were used, are not highly related. While the two word recognition measures correlate highly (Pearson r = .775, p = .001), the relationship between the Dominion score and the continuous Fluency measure, is significant and negative (r = -.511, p = .006). This unwelcome complication highlights the difficulty in obtaining reliable measures of reading competence for more than one specialized aspect at a time.

A further possibility would be the negation of positive token-dependent effects by the commonly-postulated "distraction". While no attempt was made to quantify how much attention token-group subjects paid to their point-totals, it is the writer's impression that continuous
monitoring dropped out after a very small number or fraction of lessons. Certainly subjects did not appear to watch the counter with the intensity of concentration noted in studies by other experimenters, and deemed by them sufficient to interfere significantly with learning. Occasional comments did show however that e.g. total of points earned in the last lesson, was accurately recollected.

Finally, one might retreat into the hypothesis that the attempts made to "weight against the hypothesis" (of token efficacy) by assigning the token groups to the first part of the day, putting apparent cases of "specific reading difficulty" into other groups, etc., worked too well. Of those advanced, the first explanation still seems to be most plausible although of course there is no reason why all of the suggested factors should not have played some part in the final result.

The dramatic finding that effect of response cost is dependent on whether the system is still in force or not, is at first puzzling. The only model which seems even remotely appropriate is the "compensatory increase": a re-surgence of formerly punished responses after punishment ceases, reported by Azrin & Holz (1966). However, to equate the suppression of an inaccurate reading response "punished" by token deduction, with e.g. a decrease in speed or frequency of a free operant followed by electric shock, appears to place a considerable strain on the generality of the original result.
If such a parallel seems farfetched, an alternative or perhaps supplementary explanation can be framed in terms of an initial and transient emotional response to the response-cost contingency, of over-arousal (or possibly resentment) which might depress performance on the first Dominion test and thus artificially inflate pre-post Dominion changes. The clear reciprocity of performances under the two conditions however (D's better while token system in force, ND's better immediately afterwards), cannot be adequately explained by this supposition.

As seen from Chapter V, we cannot effectively predict how contingently saying "No", or "Wrong", or taking away a token or material object, or doing nothing, will affect the frequency of errors in a given child working at a given task. The distinction previously stressed between the informational and motivational roles of presumed "reinforcers" seems to be highly relevant here.

The next three experiments performed all aimed to replicate and explore further the response-cost results (see Chapters IX and X).

The failure of material backups to facilitate learning more than did the "value-less" stars, is not so difficult to understand. There is experimental support (see Chapter V, pt. II (c), for the contention that middle-class children often are unaffected or adversely affected by the provision of extrinsic reinforcers. In successive 4-week
periods of teaching fractions, Richter (1972) found that monetary rewards gave a consistent but not statistically significant advantage, measured by pre-post improvement, over verbal praise.

Additionally, queries can be raised about the subjective values to the subjects in the present experiment, of the reinforcers provided. Efforts had been made to ascertain at least the availability of these reinforcers in the natural environment. In fact the B group subjects received an average of 11¢ pocket money per week (range 0-40¢). Money for lunches and stationery items similar to those included in the backup collection, may have been provided separately from pocket money by parents; the objects wanted by these children may have been far too expensive to be included in the programme; some children had considerable savings; for a number of reasons the material backups may not have had the functional reinforcement power expected. Last but not least important of these reasons could be the widely-accepted idea in our culture that children should learn for love of learning, and that "bribery" demeans the receiver.

The last point opens up the issue of "intrinsic" reinforcers, and how the addition of extrinsic rewards will affect an ongoing activity maintained by conditioned or intangible consequences. Expt. VI was designed to investigate such questions (see Chapter XII).

The rating of lessons as more "bitter" and "sad" by B
group subjects seems to support a formulation including perceptions and emotional responses to the reward-for-learning situation, as highly relevant variables. In similar vein, the stars may have had potent subjective significance to the NB subjects. Most of them associated the award of stars for success with the earliest grades of school, and although this might have given the stars a "babyish" meaning, it appears that the opportunity to actually gain stars for themselves was novel and valuable to several of the NB subjects, (e.g. one remarked "we had them in Grade I, but I only got about three stars. I kept getting things wrong all the time. This is the first time I've had more than 20 stars").

Regarding the suggestion of a delayed influence of material reinforcement (B groups out-performed NB groups on long-term word-recognition improvement), a pertinent study has been carried out by Crano & Messe (1970). College students were asked to write essays in favour of undergraduates being drafted into the army, either shortly or some time before stating their contrary opinion on the issue, and with payment of either 50¢ or $5 for the essay. When attitude measurement was immediate, those paid less changed their attitudes more (in line with predictions from cognitive dissonance theory); however when measurement was delayed, the reinforcement effect was stronger and those paid more changed their attitudes
more. Long-term effects of material reinforcement were investigated further in Expt. V (Chapter XI).

3. Conclusions

Expt. I was a long-term exploration of the effects of four variables (teaching, tokens, response cost and kind of reinforcement) in remedial reading training. While earlier findings on token systems applied to reading training were on the whole not supported, results were not inexplicable, and raised several interesting questions for subsequent follow-up.
Chapter IX. Experiments II & III. Follow-ups of the response-cost effect with multiple choice materials.

Experiments II & III, followed up the finding from Experiment I of more accurate reading under response-cost conditions followed by less accurate reading when the token system was no longer operational. Within-subject comparisons of performance under different conditions were made, as well as between-group analyses. The main interest was to discover whether the response-cost effect, resembling as it did the post-punishment increase in rate of negatively reinforced responses found in animal studies, would also occur in other human cognitive tasks, or whether there was some other feature of Experiment I (such as long duration) which was necessary for the appearance of the effect.

A new task was therefore chosen, on which subjects could provide manageable amounts of objective data over a shorter time-span. As it was desired to contrast performance under different conditions within the same subjects, learning effects with practice were required to be minimized. Multiple-choice comprehension exercises were selected on the grounds that, while it is a cognitive task involving reading, it was not likely to demonstrate large transitions in performance over time, unless influenced by extrinsic variables.

Experiment II

AIM: to replicate the Experiment I result response-cost effects, in a within-subjects design, where contact with each subject was limited to one session.
METHOD

1. Design

Four groups of 10 subjects each worked on multiple-choice comprehension material which was divided into three blocks. The treatments received by each of the four groups were as follows: A represents point award for correct responses only, and B represents concurrent point deductions for errors:

(a) A - B - A
(b) B - A - B
(c) A - A - A (a control group for block difficulty)
(d) A - B - A with backups (points acted as tokens exchangeable later for money and sweets).

2. Procedure

a) Subjects

Grade V children from a middle-class primary school (equivalent but not the same as that used in Experiment I) were the subjects. They were selected for "average" reading ability by their teachers. The sexes were evenly represented in each group of 10, and the subjects allocated at random to the groups. The E saw subjects one at a time, in an unused work area, during school hours.

b) Materials

Passages were selected from the Schonell Essential Read and Spell Book III, which is appropriate in difficulty for RA 9 - 10 years, i.e. includes the "average" Grade V level of achievement. Originally 15 passages were chosen and turned into multiple-choice comprehension material by
adding two alternatives to suitable original nouns and verbs (mainly). An example of the material facing the subject would be "One day in London I went for a trip by (camel, launch, balloon), down the River (Thames, Amazon, Murray)". Preliminary runs indicated that five passages per block would take too long, so only the first three of each section were used (Materials in Appendix D). The subjects were required to read the passages silently and underline "the best, most sensible word out of the ones in the bracket". Nursery rhymes were used as demonstration and warmup exercises.

As instructions for children need to be very concrete and repetitious, details are relegated to Appendix E.

The three blocks of three passages contained 55, 53 and 56 choice-points respectively. The position of the correct alternative within the bracket was randomized but distributed evenly overall.

It was hypothesized that this task would accentuate the effect of point-award and deduction on a performance which was already at a stable level (or "plateau"), rather than one where new learning was involved. Correct choices did not necessarily facilitate subsequent performance, except presumably by improving the subject's grasp of the passage content. As there were three different passages in each block, independence of responses from one choice-point to another largely prevailed.

c) Apparatus

The same telegraph key and counter apparatus was used as
for Experiment I, with the addition of a buzzer which provided a distinctive auditory stimulus associated with point-deduction, compared with the click for point-award. The buzzer was not loud, and was intended to have informative rather than aversive properties.

d) Reinforcers

As Experiment I had shown no performance differences between subjects receiving material reinforcers and those receiving symbolic rewards (stars) or even no additional consequence other than feedback (TC group), the provision of information was expected to act as a sufficient reinforcer to sustain the work behaviour, at least over the half hour or so envisaged as the experimental period for each subject.

The "points" informed and were therefore seen as motivators. The subjects were told that the E wanted "to see how many points you can get", and that "you have to see how many points you can get altogether". They were tested individually and at the end of the session asked not to discuss the procedure with their friends until after the E had left the school. The A and B conditions were equal in feedback characteristics: silence or a buzz and point deduction both conveying the fact of an incorrect choice, but not the correct alternative. In terms of the studies discussed in Chapter V, A was a "Right"/Nothing condition, while B represented "Right"/Wrong", all non-verbally.

It was thus assumed that any differences in performance
under the A and B conditions would represent the effect of induced motivational differences only.

3. Predictions

If Experiment I findings were replicated, response cost would lead to superior performance until it was removed, when a performance decrement would follow. It was predicted therefore that on both rate of correct choices and percent correct overall, the B condition would be associated with better performance than that under a subsequent A condition.

Backups added to the second A-B-A group were expected to have no influence on performance, repeating the Expt. I finding.

RESULTS

For 12 of the 40 subjects, the alternative end-point of "10 minutes per block" rather than "3 passages" had to be applied, in the interests of not removing children from the classroom for periods too much in excess of what the teachers had been led to expect. Total number of correct choices therefore became an inappropriate measure, and the data was analysed in terms of time per correct choice, and percentage of correct choices, by subjects and treatments. Figs. 6 & 7 show group averages on these two measures under different conditions.

The A-A-A subjects can be regarded as controls for passage difficulty and any fatigue or practice effects. In fact it can be seen that there were substantial differences between blocks in average rate of correct responding when
Fig. 6: Performance rate across blocks, under different conditions

Rate (seconds per correct choice)

Group
- A-A-A
- A-B-A
- B-A-B
- A-B-A with backups

Block I  Block II  Block III
Fig. 7: Percentage of correct choices across blocks, under different conditions.
conditions remained constant.

A $1 \times 4$ analysis of variance having shown no difference in initial rate for the four groups, ($F = .237$), performance on Blocks II and III was expressed as a difference from Block I performance for each subject. These Block II – I and III – I change scores were then analysed to see whether A/B conditions had any significant effects either within subjects or between groups. Dunnett's test was applied to test for differences between the control and other groups.

The whole above procedure was repeated using the "percent correct" data. Group averages appear in Tables 6 & 7.

Table 6. Mean performance rate (seconds per correct choice)

<table>
<thead>
<tr>
<th>Group</th>
<th>Block I</th>
<th>II-I difference</th>
<th>III-I difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-A-A</td>
<td>14.55</td>
<td>1.90</td>
<td>.91</td>
</tr>
<tr>
<td>A-B-A</td>
<td>13.29</td>
<td>6.06</td>
<td>3.42</td>
</tr>
<tr>
<td>B-A-B</td>
<td>14.75</td>
<td>2.09</td>
<td>-.24</td>
</tr>
<tr>
<td>A-B-A with backups</td>
<td>12.93</td>
<td>3.92</td>
<td>.63</td>
</tr>
<tr>
<td>$1 \times 4 F$ (3,36)</td>
<td>.237</td>
<td>1.9584</td>
<td>1.5156</td>
</tr>
</tbody>
</table>
Table 7. Mean percent correct choices (of subject's total)

<table>
<thead>
<tr>
<th>Group</th>
<th>Block I</th>
<th>II-I difference</th>
<th>III-I difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>77.46</td>
<td>-3.53</td>
<td>2.84</td>
</tr>
<tr>
<td>ABA</td>
<td>80.19</td>
<td>-6.85</td>
<td>-1.12</td>
</tr>
<tr>
<td>BAB</td>
<td>77.67</td>
<td>-6.99</td>
<td>5.57</td>
</tr>
<tr>
<td>ABA with backups</td>
<td>76.00</td>
<td>-8.06</td>
<td>3.71</td>
</tr>
<tr>
<td>1 x 4 F (3,36)</td>
<td>.4242</td>
<td>.9756</td>
<td>.6387</td>
</tr>
</tbody>
</table>

Thus it can be seen that none of the subjects' performances were consistently influenced by the treatment received, and no groups differed from the controls (A-A-A) on either measure. The prediction that the response-cost effect of Expt. I would be repeated was not upheld; the prediction of "no difference" between ABA groups with and without material backups, was sustained.

DISCUSSION

Further evidence was found that information alone can maintain cognitive behaviour, and that additional extrinsic consequences do not enhance performance.

Subjects in Experiment II underwent alternations of A & B conditions (equivalent to Experiment I No Deductions and Deductions), and the response-cost effect was not replicated. The next experiment therefore, constituted a more exact repetition of Experiment I conditions: tests without any
points, before and after a series of trials all under the same (A or B) conditions.

**Expt. III**

**AIM:** to attain a closer replication of Expt. I conditions than was provided by Expt. II, with the same goal of exploring the response-cost effect further.

**METHOD**

1. **Design**

   A more exact parallel with Expt. I conditions was instituted. Before and after a series of multiple-choice comprehension exercises, with either point-award (A) or point-award and response cost (B) throughout, subjects completed test exercises without the point system operating. The design was thus \( O_1 - A_1 - A_2 - A_3 - O_2 \) vs. \( O_1 - B_1 - B_2 - B_3 - O_2 \). In line with the earlier results it was predicted that the exercises performed under B would contain fewer errors than those under A, but that the \( O_2 - O_1 \) difference would be greater for subjects who had practised under A conditions. Apart from the two groups already described, one further group of 10 subjects did the same work without ever receiving points, and a fourth group of retarded readers (of similar RA but CA 2 years greater) underwent the \( O_1 - B - O_2 \) procedure, in order to see whether retarded readers differ in their response to "punishment" for errors, from normal readers.
2. Procedure
   a) Subjects
   Forty pupils of a third middle-class primary school, who were average Grade V readers and balanced for sex distribution within groups, were the subjects. They were seen individually in a disused classroom during school time.
   b) Materials (see Appendix D)
   The task was again multiple-choice comprehension exercises, but slight changes were made from the Experiment II material.
   Frequency of different numbers of errors made was plotted for the nine passages of Expt. II, and five chosen which had similar difficulty profiles. The first and last of these were the no-point tests, and the intervening three were used for training under A or B experimental treatments.
   c) Instructions
   Apparatus was the same as for Expt. II and instructions again stressed point-collecting as the subject's task. The counter was introduced and apparatus switched on (which lit a red light) after the first test passage, and the reverse (with suitable comments) before the last passage.

RESULTS
All subjects carried out the same task. Data consisted of the number of errors made in, and the time taken to complete, the three sections: pretest passage without tokens (0), three training passages (under A or B conditions of point award without or with deductions for errors), and
post-test passage without tokens ($O_2$).

Figs. 8 & 9 show graphically how number of errors made, and performance rate (in correct choices per minute) failed to vary between groups according to treatment. Tables 8 & 9 give group averages and analysis of variance results. Again, no group differed significantly from the controls (who never received "points").

Table 8: Mean number of errors made

<table>
<thead>
<tr>
<th>Group</th>
<th>$O_1$</th>
<th>$O_2$</th>
<th>$O_1 - O_2$ difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-A-0</td>
<td>6.4</td>
<td>4.7</td>
<td>1.7</td>
</tr>
<tr>
<td>0-B-0</td>
<td>6.6</td>
<td>4.1</td>
<td>2.5</td>
</tr>
<tr>
<td>retarded</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-B-0</td>
<td>6.2</td>
<td>4.9</td>
<td>1.3</td>
</tr>
<tr>
<td>0-O-0</td>
<td>6.8</td>
<td>6.0</td>
<td>.8</td>
</tr>
<tr>
<td>$1 \times 4 F$</td>
<td>.1044</td>
<td>1.0126</td>
<td>.513</td>
</tr>
</tbody>
</table>

Table 9: Mean performance rates (in correct choices per minute).

<table>
<thead>
<tr>
<th>Group</th>
<th>$O_1$</th>
<th>$O_2$</th>
<th>$O_2 - O_1$ difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-A-0</td>
<td>3.67</td>
<td>4.50</td>
<td>.84</td>
</tr>
<tr>
<td>0-B-0</td>
<td>4.03</td>
<td>5.12</td>
<td>1.08</td>
</tr>
<tr>
<td>retarded</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-B-0</td>
<td>3.75</td>
<td>4.49</td>
<td>.74</td>
</tr>
<tr>
<td>0-O-0</td>
<td>3.40</td>
<td>4.06</td>
<td>.66</td>
</tr>
<tr>
<td>$1 \times 4 F$</td>
<td>.604</td>
<td>.988</td>
<td>.2292</td>
</tr>
</tbody>
</table>
Fig. 8: Mean errors made under different conditions, by 4 groups.

No Deductions (A) — —
Deductions (B) — —
Retarded B's — —
Controls — —

Errors out of 18 choices

Pre-test (O₁)  Experimental  Post-test (O₂)
Fig. 9: Performance rate under different conditions, by 4 groups.
DISCUSSION

Experimental results did not show the predicted pattern, i.e. better performance with Deductions for errors than without, but worse performance afterwards, relative to subjects who had not had Deductions.

The differences between Expt. I and Expts. II and III which might be relevant to the failed replication of the response-cost effect, are as follows:

1) time span. While Expt. I subjects were exposed to the reinforcement conditions for 20 40-minute sessions, subjects in Expts. II and III were so exposed for maximum periods of 30 minutes and 12.5 minutes respectively. It seems plausible to suspect that the effects of the response-cost contingency might follow a time-dependent course and that therefore long-term results may only be replicable in similarly long-term experiments.

2) task. The subjects in Expt. I were assumed to be acquiring discriminative operants during their sessions; those of Expt. II and III were assumed to be performing a response not very susceptible to momentary changes in efficiency. While the Expt. I task which produced the response-cost effect involved word recognition, the response required in Expts. II and III was a demonstration of comprehension of textual material, albeit probably compounded with word recognition and even short-term memory.

Two other differences between Expt. I and Expts. II and III were not believed relevant to the present results:
a) use of average rather than retarded readers. The group of 2 years-retarded readers (of same RA but 2 years older) which was included in Expt. III performed indistinguishably from the other groups.

b) failure to provide material or tangible consequences for point-totals. The "with backups" group in Expt. II, like the backup groups in Expt. I, showed no performance benefit from the extra contingency.

The failure to find a significant difference between the no-point controls and the groups which did receive feedback on progress, suggested that the task was less than ideal for the purposes of the experiments. The role of information would have been reduced in importance where, as noted, successive responses were virtually independent of each other. Further, over the short periods of time involved for each subject in these experiments, it seems feasible that performance could be maintained by the well-drilled compliance, and the responsiveness to noncontingent adult approval, of middle-class children.

The next attempt to recapture, and explore further the Expt. I response-cost effect, accordingly required quite different behaviour from the subjects.
Chapter X. Experiment IV. Response-cost with a paired associate learning task.

INTRODUCTION

Between the pre- and post-treatment assessments of Experiment I, the subjects had been receiving training which was assumed to aid word recognition efficiency. Thus the tokens with or without deductions had been applied for performance on a learning task, with an expected shifting baseline. In Expts. II and III however, performance of the task-response was expected to be in a relatively stable state. Whether in fact this was true, and practice alone had little positive effect, cannot be gauged from the results obtained, despite the use of control groups, due to the uneven difficulty of the passages selected for comprehension exercises. It is not clear indeed whether difficulty differences were associated with variable word-familiarity or theme-familiarity.

For Expt. IV a return to a more specifically learning-oriented task was chosen. In this way Expt. I conditions were supposed to be more exactly repeated.

AIM: The basic aim of Expt. IV was to repeat Expt. III using paired-associate learning (PAL) rather than multiple-choice comprehension exercises, as the task.

METHOD

1. Design

The design was slightly different from Expt. III in that instead of the 0 - A - 0 vs 0 - B - 0 analysis, subjects in the present study learned two lists under A or B
conditions and then no-token learning was tested on a third list.

A again represents the No Deductions, "Right" or reward only situation, and B the Deductions, "Right"/"Wrong" or reward plus punishment condition. The design can then be characterized as $A_1 - A_2 - 0$ vs $B_1 - B_2 - 0$. A control group worked at the same task but never received any points, in order to gain token-free information on list difficulty.

2. Procedure

   a) Subjects

   Subjects were again middle-class Grade V children in groups balanced for sex, of average reading achievement, and allocated randomly to groups. The 25 subjects of each group were seen individually in a familiar empty room.

   b) Apparatus

   The same apparatus as for Expt. III was employed and subjects were again told that their task was to get points. The B group subjects were given a "bank balance" of 20 points before beginning learning trials.

   c) Materials

   Three lists of 8 pairs of verbal stimuli were constructed, using principles exemplified and enumerated in Battig and Spera (1962), Haagen (1949), Runquist (1966), Runquist and Schulz (1959), Schulz and Runquist (1960), Thune and Underwood (1943), and Underwood and Schulz (1960).
Materials actually used

<table>
<thead>
<tr>
<th>List I</th>
<th>List II</th>
<th>List III</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat - 17</td>
<td>knife - Scotland</td>
<td>red - beyond</td>
</tr>
<tr>
<td>dog - 19</td>
<td>fork - China</td>
<td>yellow - for</td>
</tr>
<tr>
<td>mouse - 6</td>
<td>spoon - Italy</td>
<td>green - that</td>
</tr>
<tr>
<td>sheep - 9</td>
<td>plate - Russia</td>
<td>purple - about</td>
</tr>
<tr>
<td>house - 8</td>
<td>bowl - Holland</td>
<td>blue - how</td>
</tr>
<tr>
<td>cow - 11</td>
<td>glass - England</td>
<td>orange - with</td>
</tr>
<tr>
<td>pig - 14</td>
<td>cup - Greece</td>
<td>white - under</td>
</tr>
<tr>
<td>snake - 15</td>
<td>saucer - France</td>
<td>brown - either</td>
</tr>
</tbody>
</table>

Pilot runs provided opportunities to manipulate and equate list difficulty, and to make decisions about procedural details. E practised with a stop watch until timing of stimulus presentation and response period were standardized.

Each list was arranged into four different intra-list orders of stimulus-response pairs, such that no pair occurred in the same list-position more than once, and no given pair succeeded another more than once.

<table>
<thead>
<tr>
<th>Training trial</th>
<th>Order Test Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>
A card with a 3" x ½" slit was used to obscure all of the list except for one stimulus term, or the stimulus-response pair, as appropriate.

d) Detailed procedure and instructions

Each subject saw the same order of pair-orders in each list. After having the task explained, the subject was additionally given an idea of the population of stimulus and response terms, e.g. "In this list the first word in each pair is an animal, and the second word is a number between 1 and 20".

For the first presentation of the 8 pairs in each of the three lists, both terms were exposed together and the E said them (1 per second); an extra 2 seconds was then allowed before the next pair was exposed. On three subsequent presentations of the list (with items in a different order on successive trials), two seconds of stimulus-term exposure (which included E saying the word) were allowed for the subject to anticipate the correct response, which was then exposed with the stimulus for a further two seconds. These periods are as suggested by Runquist (1966). Correct anticipations were followed by point-award, and incorrect or no anticipation by point-deduction, with their associated clicks and buzz. All information was thus presented both visually and aurally.

Subjects had three anticipation trials on each list. They learned the first two lists under the experimental conditions (A or B) and before the third list was learned
the apparatus was turned off and pushed aside, and E said, "Now we'll see how you can do without getting any points".

(Verbatim instructions are in Appendix E.)

Data consisted of the number of correct anticipations made over three trials, for each list. Originally all subjects got the same list first and half of them got each of the other two lists next. However it was found that list difficulty was a significant variable confusing results; (a 2 x 2 analysis of variance of the first 40 subjects showed that list order had a greater influence than, and interacted with, treatment condition.

The order of lists which progressed from easiest to hardest was therefore chosen and a further 15 subjects run with this material under each treatment condition. In the data to be reported then, all subjects learned the lists in identical sequence.

3. Predictions

If the earlier response-cost effect were replicated, it would be expected in this experiment that:

(a) on the second list treatment B (Deductions) would be associated with better performance than would treatment A.

(b) on the third list those who had had treatment A would perform better than those who had had treatment B.

(c) the no-point controls would not perform differently from the experimental groups on either list, but would indicate how token-withdrawal interacted with increasing list-difficulty.
RESULTS

Mean number of correct anticipations by treatment are shown in Fig. 10.

List II performance was taken to represent the effects on learning of the treatment conditions. A 1 x 3 analysis of variance revealed no differences between groups. Individual change scores from lists II to III were similarly analysed with the same result. Table 10 gives the relevant group means.

Table 10. Mean number of correct anticipations of three groups on each PAL list

<table>
<thead>
<tr>
<th>Group</th>
<th>List I</th>
<th>List II</th>
<th>List III</th>
<th>List II - III difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-A-0</td>
<td>6.48</td>
<td>6.60</td>
<td>5.76</td>
<td>.84</td>
</tr>
<tr>
<td>B-B-0</td>
<td>6.44</td>
<td>7.32</td>
<td>5.80</td>
<td>1.52</td>
</tr>
<tr>
<td>O-O-O</td>
<td>6.70</td>
<td>6.90</td>
<td>4.90</td>
<td>2.0</td>
</tr>
</tbody>
</table>

DISCUSSION

The superior average performance of the Deduction (B) subjects while token reinforcement was still operating (on List II), although not statistically significant, concurs in direction with the prediction from the Expt. I result, as well as with findings such as those reported by Mosberg (1970), and Rothberg & Harris (1972).

In the present situation, the "tokens" signalled no accumulating tangible reward and, what has been suggested as at least equally relevant, had no unique information value.
Fig. 10: Mean P.A.L. performance across lists, under different conditions

![Graph showing mean correct anticipations across lists for different groups.](image)
Subjects saw and heard the correct response to each stimulus-term, two seconds after the latter appeared. This feedback on correctness of response was automatic and immediate; all that the points added was a visible total of correct responses for each list. Subjects however had no means of knowing the significance of their point totals (e.g. what percentage of their peers had attained a higher score). Therefore the new information obtained from the total of points earned was slight.

Two other possible explanations of the failure to replicate lie in (a) the task, and (b) the absolute length of exposure to the experimental contingencies.

(a) Task difference as an explanation of failure to replicate

The Expt. I task was a diverse, school-imitative one perceivable as directly related to academic achievement. None of this applied to Expt. IV, seen more readily as a contrived, "guinea pig" task. Notwithstanding this, the desire to achieve in accordance with the demands of authority figures in educational settings appears to have been sufficiently well inculcated in several subjects to cause overt signs of anxiety during the test session. Without physiological measures it is difficult to judge whether such anxiety was sufficient to depress performance, or whether it had any facilitating effect.

Apart from the relative artificiality of the tasks, the question of stage of learning may be important. Discriminative control was only just being established over the three trials of each PAL list. The average number of responses
correctly made in the time allowed, was only 6 out of the 24 possible on each list (three trials of 8 pairs per list). Typically subjects seemed to spend the first trial learning more about the population of response items. Although it was rare for subjects to produce a response not included in the original list, it was common for them to refuse to guess (as they had been asked to) on every trial, especially for the first run through each list. The fact that the response terms were derived from increasingly large and perhaps amorphous universes, viz. "numbers from 1 - 20", than "countries", then "preposition or joining words" may have contributed heavily to the relative difficulty of the lists. Some of the "learning" that was occurring therefore, was one step more fundamental than the reading-training situation of a 10 year old child: i.e. learning how two more or less well defined populations of stimuli (auditory and symbolic visual) correspond to each other.

Possibly, requiring learning of the PAL material to a criterion of mastery such as 7/8 correct on the last trial, would have enabled comparisons to be made between subjects at a similar stage on the route to $S^D$ control, instead of scattered all along the path. This would however have at least tripled the number of trials required for some subjects, and there is no compelling reason to believe that it would have materially altered the outcome of the experiment.

(b) **Time of exposure as an explanation of failure to replicate**

Expt. I subjects had 20 (19.8 exactly) 40-minute
sessions before being re-tested without tokens; Expt. IV
subjects had about 7 minutes of corresponding experimental
condition. It does not seem impossible that this temporal
difference may be germane to the difference in results.
Not necessarily time per se, but some gradual time-
mediated learning process, possibly of an emotional or
even social kind is the factor envisaged. As the Expt. I
subject discussed his lessons with his parents for example,
or began to read with more confidence and ability in class,
many extra-experimental sources of reinforcement may have
become operative and presumably, interacted with those
controlled by the experimenter.

Another and probably related time-dependent influence may
have been the way in which subjects responded emotionally to
the experimental situation. It is feasible that all
children when first taken from their group and familiar
teacher and classroom by a stranger, experience some
anxiety which will vary from uncertainty so slight as to be
pleasurable, to fairly severe panic, doubtless depending on
past consequences of such a separation and also on the degree
do security or attachment felt in the usual environment.
The child who has spent 10 minutes with the E, and the one
who has spent 20 lessons, will differ in their understanding
of what reinforcers the E controls, because of their
different opportunities to observe then.

Design considerations

The within-group variance in scores was very large; some
children were well on the way by the third trial to mastery of the 8 pairs, while others still seemed to be responding (if at all) in a more or less random manner, or using devices such as learning only one or two pairs altogether, or learning one or two response terms and repeating them to every stimulus until a match was obtained. Apart from the variability in subjects' learning efficiency, it is clear that the list-difficulty varied in an idiosyncratic manner for individual subjects. Fig. 10 shows a constant arrangement of lists: for every group List II was the easiest on average and List III the hardest. In fact a constant proportion of about 30% of subjects in every group scored highest (or equal highest) on List III, and only 12% of the two experimental groups and 20% of the controls produced exactly the group-mean shaped pattern of numbers of correct anticipations.

This unrepresentativeness of averages derived from groups is of course a classic argument against the use of groups to test hypotheses. As this issue has been discussed elsewhere (mainly in Chapter IV), it suffices now to point out that, where "learning" is a deliberate goal of the experimental procedure, the use of the intra-subject replication design has serious if not equal hazards. The use of within-individual change scores from List II to III in the present analysis was an attempted compromise. Percent of total correct anticipations which were made on each list, was also determined individually, with similar rationale, but gave no
further insights: the proportion of all correct responses which was made in the last list, averaged 30.2% for the No Deduction (A) group, 29.3% for the Deduction (B) group and 29.8% for the Controls.

Conclusion

The choice of a "pure" learning task in Expt. IV proved not to alter the negative results obtained with a performance-oriented task in Expt. II and III. Any replication of the Expt. I response-cost effect seems to require replication of the Expt. I task - and therefore of the Expt. I length. The acquisition of reading skill cannot be researched in one short session per subject, unless perhaps a highly individualized and limited "task" could be devised for each subject, with corresponding loss of generalizability of results.

The response-cost effect of Expt. I was too marked to be discarded as a statistical freak. That deductions aided learning compared with no deductions is, if not predictable, at least consistent with much other evidence from other researchers. The reversal in performance once the context is changed by token-elimination remains unexplained by the experiments described in this and the preceding chapter, which were finally concluded to be inadequate as replications of the Expt. I situation, by virtue or both task-substitution, and treatment brevity. Some questions of design were discussed.
Chapter XI. Experiment V. Effects of contingent and noncontingent material reinforcement on reading fluency and error rate.

INTRODUCTION

The effects of the provision of material reinforcement in Expt. I were (a) less favourable ratings on semantic differential of the pleasantness of the remedial lessons, and (b) a weak long-term effect of greater word recognition efficiency.

Expt. V was planned to follow up the latter result. Another question investigated was the relative effects of contingent and non-contingent or "free" positive reinforcement. If reinforcement is acting in a strictly operant sense to increase the probability of a given response under given conditions, it would of course have to occur contingently upon the emission of that response. If however the reward is being associated in a Pavlovian way with less specific responses, such as "coming to lessons" or "reading", it need presumably not be contingent on accurate reading responses. Again the impossibility of adequately monitoring reading frequency and accuracy in the nonexperimental environment is frustrating. Certainly in any clinical therapeutic endeavour a detailed analysis of the reading behaviour, requiring close co-operation with parents and probably teachers, would be essential.

There is in principle, no reason why the administration of a given appetitive event should not reinforce different responses simultaneously, or differ in effect in different
individuals. It is important in the applied situation to know what is the most efficient target behaviour to reward. A little evidence is available from the literature, of "noncontingent" reinforcement effects.

When Burchard (1967) awarded a fixed number of tokens, no longer contingent upon workshop and school performance, his delinquent subjects immediately reduced their output of these behaviours almost to zero. Re-instatement of contingent token award (in an ABA design) resulted in a return to first-phase performance levels. Unfortunately Burchard does not provide details of pre-token frequencies of the behaviour desired. He found that making response-cost contingent or noncontingent had less clearcut effects on unacceptable behaviours.

Tyler and Brown (1968), also working with delinquent boys, found that performance on a current-affairs test was better when reinforced with contingent tokens (redeemable for canteen items and privileges), than when a "straight salary" was paid. This finding held for both between- and within-subject comparisons. It is worth noting however, that the difference although statistically significant was too small to be of much educational significance. The authors comment that previous efforts they had made had showed no improvement in academic performance with token reinforcement, "presumably because of inadequate controls, particularly with regard to the measurement of the criterion" (p.167),
and suggest also that

"It is doubtful that the tokens would have been this effective in a prosperous urban junior high school in which the youngsters were satiated on tangibles, enjoyed school, and were achieving 'success' in the middle-class culture"

(Tyler and Brown 1968, p.167).

Christophersen, Davis and Wolff (1970) failed to separate the effects of contingent and noncontingent reinforcement, from the variables of reading aloud or silently, and receiving feedback on the correctness of responses or not. Their third-grade subjects read faster but less accurately when they were provided with noncontingent money, no feedback, and read silently, than when they received contingent reward, information on response-correctness, and read aloud.

With kindergarten children learning the elements of handwriting, Mandelker, Brigham and Bushell (1970) found that the teacher's rate of social contact was consistently higher for those receiving contingent rather than non-contingent tokens. When both groups received contingent tokens, relative rates of teacher-contact were the same as when neither group got any tokens.

75% of adult subjects and 13% of child subjects successfully detected the noncontingency of a following event, and thereafter emitted fewer responses, in a study by Poresky (1970, 1971).
"Noncontingent reinforcement, which by its presentation of both response and reinforcement stimuli imparts more information to the organism, appears to be a more efficient form of extinction than nonreinforcement (which is perceivable as a very high ratio schedule) for detectors. For nondetectors the indication is that noncontingent reinforcement is no less efficient than nonreward. Noncontingent reinforcement may well be the most efficacious extinction procedure."

(Poresky 1971 p.692).

Other experimental results are conflicting. Whereas Rickard, Clements and Willis (1970) found that noncontingent tokens reliably reduced output of programmed maths materials to negligible levels for three of their five subjects, Cormier (1970) concluded that noncontingent praise from the teacher could increase the task-relevant behaviour of all students, whether selected as targets or not.

The question of how a fixed wage (noncontingent reinforcement) affects human motor performance, when compared with contingent money or points on a counter, was investigated by Weiner (1972). His five subjects pressed buttons on a FR 40 and each underwent every experimental condition eight times (over 8 days). Although without exception the subjects worked harder for money than for points alone, considerable behaviour was maintained for points only. The noncontingent wage produced intermediate rates of responding, even when more money could be earned by the wage than by response-contingent pennies.

The effect of making material reinforcement noncontingent, is functionally to separate the motivational and informational
aspects of reinforcement. As seen in Chapter V there has been a failure to resolve the relative roles of these two aspects of reinforcement in human learning. At the same time, the noncontingent subjects must not be misled about which responses were correct.

The experiment to be described manipulated reinforcement variables in the context of reading by backward readers. A group procedure was used so that each subject could have an appreciable number of lessons. In addition, subjects were expected to learn from each other, as well as from the modelling and error correcting of the E.

A mixture of modelling and "in vivo desensitization" was employed by Muller & Madsen (1970) in their unsuccessful attempt to improve word recognition by behaviour therapy techniques. Modelling the desired response is virtually the only way to teach aspects of reading aloud such as response to punctuation. It is more problematic for the model to objectify the discriminative control exerted over his verbal behaviour by the visual symbols, but slight exaggerations of syllabification for example, can give clues as to how word analysis proceeds.

AIMS

Expt. V was designed

1) to explore the effects of contingent and noncontingent material reinforcement upon reading performance of low-standard readers.

2) to investigate the long-term effects of the treatments.
received.

3) to use and evaluate teaching by modelling and practice, within a small group.

METHOD

1. Design

Apart from subjects receiving contingent and noncontingent reinforcement for reading, it was necessary to include controls for reinforcement and for teaching, in order to interpret results with confidence. Thus four groups were planned:

1) Contingent monetary reward for reading performance (including the fluency and the accuracy of the reading): Ct group.

2) Separate equivalent feedback on reading progress, and money given noncontingently: NCt group.

3) Equal feedback on progress but no money, i.e. taught controls for reinforcement: TC group.

4) No lessons and no money: NTC group.

Note: ideally a group or groups would have been included which received lessons but never any feedback; however this was considered neither ethical nor feasible.

2. Procedure

a) Subjects and subject selection

Grade V teachers at two middle-class primary schools within two miles of each other were asked to send for screening, children "whom you feel might benefit from some remedial help with reading, but who seem of average intelligence". As in Expt. I, children with Schonell RA
in excess of CA, or language difficulties, were excluded from
the subject-pool.

24 children remained eligible. The use of two schools
was recognised to be undesirable, especially as the groups
could not in practice be split between schools. The
following considerations however were important:
a) only by selecting subjects from two schools could even six
subjects per group be obtained (although it was originally hoped
to have at least 10).
b) the schools were equivalent in socioeconomic status of
their pupils' zone of residence (derived again from the
Karmel report on education in South Australia 1971); further
the two Headmasters were friends and had been so since their
own primary-school days. Therefore differences in style of
administering the school were expected to be small rather
than large. Additionally it could well be argued that
particular teachers might have a more direct influence in
augmenting or subverting the results of the experimental
programme than did the Headmaster, and this was a factor
totally outside possible control by the experimenter.
c) being forced to use two different schools at least solved
the problem of confidentiality and possible contamination of
results by subjects comparing notes on reinforcement
procedures: the two groups receiving monetary reinforcement
were geographically separated from the two control groups.

't' tests were performed to see whether subjects at the
two schools differed in age, word recognition or fluency
scores. In no case did the differences approach statistical significance: Table 11 shows the relevant means.

Table 11: Means on several measures, of subjects drawn from two schools.

<table>
<thead>
<tr>
<th>Measure</th>
<th>School H</th>
<th>School M</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.A.</td>
<td>10 years 4.3 months</td>
<td>10 years 2.6 months</td>
</tr>
<tr>
<td>Schonell WRT Score</td>
<td>40.4</td>
<td>40.6</td>
</tr>
<tr>
<td>Dominion WRT Score</td>
<td>83.4</td>
<td>84.1</td>
</tr>
</tbody>
</table>
| Fluency: Time to read standard passage (corrected for errors) | 2'45" | 2'35"
| Reading retardation: CA-RA difference | 13.9 months | 14.1 months |

Within schools, the subjects were then sorted randomly into two groups, and 1 x 4 analyses of variance were performed to see whether the groups differed before treatment on the foregoing variables (after one change had been made to improve sex-balancing within groups). Again the groups were all found to be matched.

The schools were allocated to experimental or control conditions, and the groups in each school to treatments, by
flipping a coin.

In order to equalize the absolute amounts of reinforcement earned by Ct. and NCT groups, members of these groups (and the others) were assigned a group-rank derived from the average of their group ranks on the Dominion test and the Fluency measure. Subjects of the same group-rank were then yoked, such that for example if No. 1 (the poorest reader) in Ct group earned 3 cents, No. 1 in NCT group would be told at the start of the lesson that today he would be given 3 cents at the end of the lesson. In the unusual event of a Ct group-member missing a lesson through illness, his NCT counterpart was awarded a sum equal to the average earning of the Ct group for that lesson.

The ranking of subjects within each group was also used to standardize order of performance within each lesson (see next section).

b) Lesson form

Each group met with E every school day for two weeks (i.e. 10 sessions). Each day the TC group was seen first, then (at the other school) the Ct then NCT groups. Sessions took place at the same time every day and all were held in the morning, to avoid the effects of possible after-lunch weariness.

The format of every lesson was the same. The target behaviour was accurate, fluent reading aloud by the children, and the materials used were Dolphin series readers (as used in Expt. I) of appropriate difficulty. After a brief resume
of the plot so far, the E modelled reading aloud for five minutes. After this each subject read aloud for two minutes. The order in which they read was the same in terms of group-rank for every group in any one lesson, but varied from lesson to lesson in a manner previously determined by using a table of random numbers, with the restriction that individuals should read first or last for the day approximately as often as each other.

The two minute period was timed with a stopwatch, and the number of errors made was noted. After the lesson, E counted how many words each subject had read, per lesson. While each subject was reading, every line read without error gained him a "mark" or point on the counter (using the same apparatus as before). Lines which included one or more errors received no mark. The total number of marks gained was then entered on the subject's graph after the two minutes was over.

Instructions, (which are given in full in Appendix E), emphasized the goal of increasing the number of marks earned (or making the graph rise), each lesson. If a subject hesitated while reading, he was allowed five seconds to work out or guess the word before it was supplied by E and counted as an error.

In order to decrease competitiveness and maintain effort as much as possible, it was stressed to subjects that they were attempting to better their own past record. In fact,
great interest was always shown by group members in how many marks were earned by their colleagues, but an individual's graph was not shown to subjects other than the one who had generated it, without a specific application being granted by the individual concerned.

c) Reinforcement procedure

The feedback described above was common to the three groups receiving lessons. Social reinforcement was similarly standardized: E gave praise for increases in fluency (e.g. "Very good (child's name), you did better today than you did yesterday"), and made encouraging comments such as "Perhaps tomorrow you can make it go up again" when fluency decreased.

Material reinforcement took the form of 1¢ pieces. The Ct group exchanged marks for cents in a 10 : 1 ratio; average total earnings was 17 cents. Although this amount was hardly princely, the great attention paid by subjects to their own and other subjects' earnings is felt to support the hypothesis that money acted as a reinforcer of high generality: it acted as a kind of super-token which could be hoarded or exchanged for chosen goods, but further, it seemed to act as a conditioned sign of approval and esteem.

As previously explained, the NCT groups were matched for amount of money received. E was alert to deny any connection between performance and pay-off, and carefully announced the day's "salary" to subjects before any reading took place.
d) Measurement

Pre-treatment assessment consisted, as described, of the Schonell WRT, the Dominion WRT, and rate of accurate reading on a standard passage. The number of words read both correctly and incorrectly in two minutes, provided a session-by-session measure of reading-aloud performance efficiency.

After the two weeks of experimentation, subjects were re-tested individually on the Dominion test and a second standard passage. At the long-term follow up four months later, the Schonell and Dominion WRT's were re-administered, and time and number of errors on a third standard passage were recorded. On each occasion subjects were seen in the same order and within a two day period.

The passages used to measure fluency were taken from the Schonell Essential Read and Spell Book III. Prior exposure to the Schonell WRT had varied amongst these subjects, so the main reliance for an estimate of word recognition was placed on the Dominion, with which the Schonell had been found in Expt. I to be highly correlated (Pearson r = .78, P < .001).

Using "lines read correctly" was not an ideal basis for reinforcement determination, as some variability in line-length was unavoidable. However, use of the more exact "words read correctly" would have entailed a time lag (while E counted) which seemed neither therapeutic nor necessary. It is recognised that words themselves are a crude stimulus unit to observe, as they also vary greatly,
not only in length but in complexity (e.g. number of syllables),
familiarity, frequency within a particular passage, and
perhaps other factors. However no other unit of reading
was as appropriate to the present situation; reinforcement
effects had been shown by other researchers using this
insensitive index.

Systematic notes were made by the E after each lesson
concerning the non-reading behaviour of group members, with
particular reference to the dimension of co-operative vs
disruptive behaviour. These observations however were not
quantifiable without independent reliability-checks
(impossible as E was the sole experimenter in all the work
described), so they remain anecdotal.

RESULTS
1) Word Recognition (see Fig. 11)

Individual change scores on the Dominion WRT from pre-
treatment to post-treatment testing, post-treatment to
follow-up, and pre-treatment to follow-up (i.e. overall),
were calculated and subjected to 1 x 4 analysis of variance
(see Table 12). There were no significant differences found
between groups (F's of .8963, 1.001 and .1110 respectively).
It is interesting to note how, although the groups receiving
extra lessons improved relatively more than the groups
receiving extra lessons improved relatively more than
untaught controls (NTC's), such that a significant effect
may have been demonstrated had lessons continued for longer,
this advantage disappeared during the follow-up period.
Fig. 11: Dominion Word Recognition Test scores over time

Graph showing the Dominion Word Recognition Test scores for different groups over time. The groups are labeled as NTC, TC, C, and NCt. The x-axis represents the time points: Pre-test, Post-test, and Follow-up. The y-axis represents the Dominion score (group mean) ranging from 70 to 100. The graph indicates an overall increase in scores over the follow-up period, with NTC and TC showing a slight decrease in the follow-up phase.
Performance levelled off or even declined during these 3½ months for the subjects who had received remedial help, and showed a relative spurt for those who had not.

A similar finding of no significant group differences occurred for Schonell WRT changes between pre-treatment and follow-up assessments, \( F = .5027 \). Details appear in Table 13.

Dunnett's test for comparing treatment groups with a control showed that the 10 group-remedial lessons had not influenced word recognition skill.

Table 12. Dominion WRT score change means.

<table>
<thead>
<tr>
<th>Group</th>
<th>Pre-post</th>
<th>Post- Follow up</th>
<th>Pre- follow up</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTC</td>
<td>1.3</td>
<td>3.0</td>
<td>4.3</td>
</tr>
<tr>
<td>TC</td>
<td>4.2</td>
<td>-0.8 *</td>
<td>3.4 *</td>
</tr>
<tr>
<td>Ct</td>
<td>2.8</td>
<td>0.7</td>
<td>3.5</td>
</tr>
<tr>
<td>NCT</td>
<td>4.3</td>
<td>-0.2</td>
<td>4.2</td>
</tr>
</tbody>
</table>

* One TC subject was lost from the long-term follow up due to protracted illness with glandular fever.
Table 13. Schonell WRT score means

<table>
<thead>
<tr>
<th>Group</th>
<th>Initial</th>
<th>Overall changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTC</td>
<td>39.2</td>
<td>4.3</td>
</tr>
<tr>
<td>TC</td>
<td>41.7</td>
<td>3.2 *</td>
</tr>
<tr>
<td>Ct</td>
<td>39.2</td>
<td>2.3</td>
</tr>
<tr>
<td>NCT</td>
<td>42</td>
<td>5.8</td>
</tr>
</tbody>
</table>

* One subject was lost from the long-term followup due to protracted illness with glandular fever.

2. Fluency

Fig. 12 shows group means for rate of correct reading responses on the standard passages before, just after, and 3½ months after treatment. Again individual change scores were analysed (see Table 14) but no significant differences found between groups.

Table 14. Fluency change means (correct words per second)

<table>
<thead>
<tr>
<th>Group</th>
<th>Pre-Post</th>
<th>Post- follow up</th>
<th>Pre- Follow up</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTC</td>
<td>.10</td>
<td>.17</td>
<td>.28</td>
</tr>
<tr>
<td>TC</td>
<td>.06</td>
<td>.15</td>
<td>.19</td>
</tr>
<tr>
<td>CT</td>
<td>.02</td>
<td>.11</td>
<td>.12</td>
</tr>
<tr>
<td>NCT</td>
<td>.04</td>
<td>.21</td>
<td>.24</td>
</tr>
</tbody>
</table>
Fig. 12: Mean fluency over time (in correct words per second on standard passages)
The number of words read by each subject during his/her daily two minutes of reading aloud, constitutes another fluency measure. Words read correctly, errors, and total words read, were all summed over the 10 lessons for each subject (see Table 15). 1 x 3 analyses of variance showed no significant differences between groups (F's all < 1).

Table 15. Mean Total Words read over 10 lessons.

<table>
<thead>
<tr>
<th>Group</th>
<th>Words read Correctly</th>
<th>Errors made</th>
<th>Total words read</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC</td>
<td>1661.2</td>
<td>54.3</td>
<td>1715.5</td>
</tr>
<tr>
<td>Ct</td>
<td>1811.3</td>
<td>62.0</td>
<td>1873.3</td>
</tr>
<tr>
<td>NCT</td>
<td>1854.7</td>
<td>56.5</td>
<td>1911.2</td>
</tr>
</tbody>
</table>

Fig. 13 charts fluency lesson by lesson for each group. All groups read the same material; all finished the first reader in three lessons; the third reader, which was graded as more difficult, was begun in the eighth lesson for TC and Ct groups, and the seventh lesson for NCT's.

3. Reinforcement effects

There appeared to be a negative relationship between total monetary earnings and improvement on the Dominion WRT. As amount of reinforcement earned was a direct function of ability, (because of the application of an invariant criterion for token-award, viz. one line read without errors), and as initial ability was reflected in each individual's group-rank
Fig. 13: Words read correctly in two minutes each lesson (group means)

Group
TC
Ct
NCT

Mean number of correct words

Lessons
a ceiling effect was thought to be a possible explanation. In this case those who were poorest initially would have improved more than those who began relatively well. Accordingly correlations were computed between group-rank and Dominion change; Spearman's rho for groups Ct and NCT was -.81, for the TC group -.41, and +.2 for the NTC group. Although groups are too small for firm conclusions to be based on this finding (with p = .05 and N = 6, rho = .829), it does appear that the relation between low initial ability and greater increase in word recognition score, holds more strongly for the groups receiving material incentives. Within these two groups, less improvement was shown by those who earned more money (Spearman -.67).

4. **In-lesson behaviour**

Reactions to the experimental procedure varied distinctively across groups. The TC subjects concentrated on their task, indulged in little irrelevant talk, and did not interrupt or sabotage each other's efforts. The Ct subjects by contrast, were intensely competitive. They paid more attention to how many marks their colleagues were accumulating than to the reading in progress, and frequently disrupted proceedings by inappropriate talk or motor activity. Fig. 13 shows their relatively greater variability across lessons. The NCT group fell midway between the others in compliance with the official requirements of the lesson. They obviously found the noncontingent money difficult to understand, often asked e.g. "why do I only get that much
They quickly formulated hypotheses that their reinforcement was in fact dependent in some way on performance or behaviour, finding E's non-explanation ("That's just the way it is") very unsatisfactory.

The groups were drawn from different schools, which conceivably differed in "permissiveness" or "discipline oriented" staff-styles. However the fact that the characteristic group-response to the experimental task increased in both magnitude and uniqueness as the lesson progressed, makes it plausible that a genuine treatment effect was responsible.

DISCUSSION

This experiment made clear the value of control groups and long-term followup, without which quite misleading conclusions could have been drawn. The main conclusion from the experiment itself may be, that group designs have uncertain utility where the measurement problem is so complex. Although not attaining statistical reliability, trends were seen for fluency to remain stable as word recognition improved, during the course of remedial reading lessons, and for the reverse to happen in untaught subjects and after lessons were terminated.

A published precedent for failure to obtain benefits from remedial lessons compared with no remedial lessons, is provided by Camp and van Doorninck (1971). They found that after 14 hours of extra training, scores on a standardized test had not been affected. Samples of the teaching material
did however show significant gains in word recognition.

There is no way of discovering to what extent the extra attention involved in being pre- and post-tested, affected the untaught (NTC) subjects. The very fact of these children's being assessed formally by an outsider might have altered their teacher's behaviour towards them; teachers perceived favourable changes in NTC’s in Expt. I. (Such effects had been expected of course to be overshadowed by experimental manipulations.)

The failure of monetary reinforcement to produce more improvement than did feedback on progress alone, is consistent with all the earlier results. The "marks" had strongly reinforcing properties for all subjects, judged by the close attention which was paid to their daily graphing, and by the expressions of pleasure or disappointment often uttered. Other intangible reinforcers were verbal consequences emanating from E, status in the immediate group of peers, and interest in the actual reading content. The latter two sources of reinforcement were not experimentally controlled, and such control would have been in practice very difficult to implement and to ascertain. Intra-group interactions, and expressed enjoyment of the reading task, were apparently affected by experimental conditions. As noted, material rewards were associated with highly competitive behaviour, but not with any greater interest or pleasure in performing the task. The non-inclusion of an attitudinal measure is regrettable. Some measure of
anxiety may also have helped to determine the function of the reading aloud done by each subject. In particular, it would be of interest to know whether reinforced practice of the behaviour had any effect in desensitizing anxieties which might contribute to poor performance.

It seems premature to conclude that in this experiment, whether "reinforcement" was contingent on desired behaviour or not made no difference. A mixture of facilitative and adverse effects of each condition, may have obscured matters. An example would be the puzzlement and resentment occasioned by the noncontingent reward. Making the same "payment" for every subject in the group would probably have solved that problem, although the reinforcement equality between individual members of the two groups would have been foregone.

REPLICATION

Part of the original design of Expt. V, was an exact replication of the foregoing procedure using low socio-economic status subjects. It was hypothesized that the effect of money as a reinforcer might be quite different in an environment where deprivation is presumably stronger. The studies quoted in Chapt. II showing greater responsiveness to money of working-class compared with middle-class children, used simpler, less culturally-valued tasks than reading.

A large working-class school was selected and pre-testing of subjects undertaken. Although the incidence of reading
retardation was greater here (proportion of professional fathers 7.6 vs 33.4), it became quite apparent that again subjects would have to be obtained from two schools, especially if group-size was to be increased as hoped. Plans were proceeding when a prolonged petrol shortage, with consequent mobility reduction, forced their postponement. The need to transport 40 lb. of apparatus between schools cut out the option of public transport; in any case there was doubt about for how long buses would continue to operate.

By the time unrestricted travel was again possible, the follow-up period desired was not, due to approaching close of the school year. Additionally, dissatisfaction with aspects of Expt. V made the E reluctant to invest further effort in repeating that particular design. The planned replication was therefore abandoned, although the goal of comparing underprivileged and affluent children's responses to monetary reinforcement for reading achievement, still seemed worthwhile (see Chapt. XIII).
Chapter XII. Experiment VI. Effects of extrinsic and intrinsic backup reinforcement on fluency and error rate.

INTRODUCTION

The results of Expt. V indicated that feedback on progress is an effective motivator for staying-in-the-situation and complying with its demands, for reading-disabled children having remedial lessons. The relatively more disruptive behaviour of the subjects receiving material reinforcement, coupled with the significant tendency of corresponding Expt. I subjects to rate their lessons as more bitter and more sad, raise questions about how "reinforcing" a material consequence is, in this situation.

Attention is beginning to be paid to the concept of "intrinsic" reinforcement. White in 1959 spoke of a "feeling of efficacy" experienced when an organism interacts pleasurably with the environment ("competence"). Although he was mainly concerned to explain playful and exploratory behaviour in the context of "drive" theories of motivation, White's contribution can be seen as a fore-runner of modern interest in the relation between inwardly-determined and externally-bestowed rewards (and perhaps of some elements of the humanistic movement in psychology).

Bijou's comments on the desirability of "weaning" children from material reinforcers for academic performance, to those more natural to the normal classroom, were discussed in Chapter III. The whole issue of control and self-control as goals of psychological knowledge and intervention, has recently been the subject of lively debate.
The ethical question of how and whether to manipulate behaviour by contingent reinforcement, is however not the primary concern at present. Rather, Deci (1971, 1972 a, 1972 b) has raised the empirical question of what in fact happens when extrinsic reinforcers are provided for an ongoing activity which appears to be maintained by "intrinsic" motivation: i.e. when the behaver "receives no apparent rewards except the activity itself" (Deci 1971, p.105). Deci avoids the problem of whether the doer might have developed his intrinsic motivation through a long series of learning trials, possibly with decreasingly obvious and tangible reinforcers; he deals with the end product, be it learned or innate, and has performed some experiments on the effects of adding material incentives to apparently intrinsically motivated behaviour.

In an A-B-A vs A-A-A design, college students worked at a plastic puzzle or on the college newspaper, without (A), or with (B) payment for completed tasks. A third experiment replaced the monetary reward with verbal praise, for puzzle-workers. Results were suggestive rather than conclusive, and supported the hypotheses that money decreases, but praise and feedback increase, a person's prior intrinsic motivation to perform an activity. A more elaborate test of these propositions (Deci 1972 a) supported a "cognitive evaluation" interpretation: that the provision of money reinforcement changes perceived locus of control (to an external one), while verbal reinforcement and feedback
increase intrinsic motivation by increasing the person's sense of competence and self-determination.

"External controls may get others to act the way we want them to, but such controls absolve them of the feeling of responsibility for those acts. Controlling others seems to insure that others will not control themselves." (Deci 1972 b, p.92).

Interpretations of the (weak) phenomena just described can be made in terms of resistance to extinction instead of "intrinsic" reinforcers. Why verbal reinforcement should produce different responses than do money rewards however, is then difficult to explain, except possibly in terms of verbal reinforcers being more "natural", less novel, and less permanent and tangible, therefore the switch to nonreinforcement conditions being harder to distinguish (thus delaying extinction).

A dissonance reduction analysis has been made of somewhat similar findings by Freedman (1965). He found that mild threat more effectively prevented playing with a highly attractive toy (nearly 6 weeks later) than did a severer threat, if subjects were allowed time alone with the toy and resisted temptation. An implication is that

"inculcating moral values will be most successful if a minimal amount of justification of any kind is offered for the relevant behaviour... [The child] should be given just enough justification to cause him to obey in the presence of the justification; and then his acceptance of the value itself will be maximal".

(p.154).
Colvin (1972) awarded some children tokens for engaging in their preferred art activities, and others the tokens noncontingently. When the tokens were withdrawn after 4 weeks, work in and choice of the formerly preferred activities decreased below baseline for the former group, while the latter were not at either stage affected. Satiation and consequent lowering of natural probabilities of occurrence might however be regarded as likely to be relevant.

**AIMS OF EXPT. VI**

Expt. VI aimed to investigate the variables of material vs "intrinsic" and again contingent vs noncontingent material reinforcement. A "no extra reinforcement" condition was included in an intra-subject reversals design, in order to see whether material or intrinsic reinforcement actually changed reading performance, compared with their non-provision. It was hoped to gain further information on the two findings recalled at the beginning of this chapter: the disaffection and disruptiveness of materially reinforced subjects in Expts. I and V, by specifically comparing effects of material and other reinforcers.

**METHOD**

1. **Design**

A longitudinal, small - N experiment was planned, the shorter-term, matched group design having revealed its limitations clearly in prior work. The desirability of comparing effects of different reinforcers within subjects,
necessitated a lengthy involvement with each subject so that differential effects would have a chance to become apparent.

The two subjects were each given a half-hour remedial reading lesson every week-day for 7 weeks. The 1st, 4th and 7th weeks constituted Baseline conditions, when feedback and mild verbal reinforcements were provided for reading. Material reinforcers were added for one subject in weeks 2 and 3, and for the other in weeks 5 and 6; intrinsic reinforcement was added similarly but reversed in order:

<table>
<thead>
<tr>
<th></th>
<th>1 week</th>
<th>2 weeks</th>
<th>1 week</th>
<th>2 weeks</th>
<th>1 week</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject D</td>
<td>Base</td>
<td>Material</td>
<td>Base</td>
<td>Intrinsic</td>
<td>Base</td>
</tr>
<tr>
<td>Subject B</td>
<td>Base</td>
<td>Intrinsic</td>
<td>Base</td>
<td>Material</td>
<td>Base</td>
</tr>
</tbody>
</table>

The Material reinforcement condition consisted of alternate days of contingent and noncontingent reward. Thus the following comparisons could be made:

a) Base I vs Base II vs Base III: to show overall effects of remedial lessons
b) Material vs Intrinsic reinforcement for each subject
c) effects of Material followed by Intrinsic, compared with vice versa (but sequential effects compounded by subject variation)
d) contingent vs noncontingent material reinforcement for each subject.

Parametric statistical analyses were planned in accordance with Gentile, Roden & Klein (1972).
2. Procedure
   a) Subjects

   Teachers of Grade V at a middle-class primary school were asked to send children whom they thought would benefit from remedial reading lessons, for preliminary screening. From the 10 potential subjects thus obtained, two were selected for their similarity on the screening measures (see next section). The two boys chosen were about equally behind in accuracy and fluency of reading compared with comprehension, and they also claimed parallel attitudes towards, and frequency of, reading. They were allocated to treatment orders by flipping a coin.

b) Measures

   Potential subjects were administered
   (i) the Questionnaire on Voluntary Reading described in Chapter VII, Study III.
   (ii) the Neale Analysis of Reading Ability form A, the 2nd, 3rd and 4th stories.
   (iii) the Neale Supplementary Test I (giving the sounds of randomly-ordered lower-case letters); performance was timed and errors noted.
   (iv) the Dominion WRT (the Schonell was not used because of recent prior use by the teachers). The two selected subjects were in fact the lowest-scorers on this measure out of all the Grade V retarded-readers population of this school.

   Once lessons began, continuous (i.e. session-by-session) records were kept of the following:
(i) fluency of reading Schonell Essential Read and Spell passage, in words read correctly per second.

(ii) percentage of errors to total words read in Schonell passage.

(iii) "points" earned by the subjects. The rules for point-award were the same for both subjects. Daily totals were graphed in the subject's presence and verbal and other reinforcers were contingent upon them as will be described.

(iv) other in-lesson behaviours such as punctuality, spontaneous remarks by the subject relevant to the treatments undergone, and number of errors made on exercises and spelling.

Additional measures were taken throughout the course of lessons:

1) at the end of every week subjects rated "our lessons this week" on the four semantic differential scales used in Expt. I, viz. large-small, bitter-sweet, strong-weak, and happy-sad. They also ticked, as true for "How I feel about reading now", one of 5 affective statements.

2) at the end of each Base week, word recognition score was obtained on a 100-word sample of the teaching material. Although more frequent administration of this measure would have been theoretically desirable, in practice it was not thought pedagogically helpful for subjects to guess wrongly repeatedly without getting any feedback, even on which responses were correct and which not; to give such feedback however would have greatly reduced the usefulness of the test.
Post-treatment testing (carried out in the next schoolday after the end of the 7 weeks of lessons), consisted of a repeat of the screening measures, using Neale form B instead of A, plus a recognition test of all the words from the Schonell lists read incorrectly on first presentation, (same as the New Words test of Experiment I), and also a short structured interview:

What did you think of the lessons?
What did you like about them?
What did you dislike about them?
Which system did you like best? nothing )
               time ) for points?
or things )
Which system did you like least?
Which system made you read best?

Followup one month after the end of lessons involved the re-administration of the Dominion WRT. It was not possible unfortunately to have a longer followup period, as the end-of-year vacation was about to begin.

c) Lesson form and materials

A similar individualized format to Experiment I was employed. As indicated by preliminary tests, the reading standards of the two subjects were very close, and it was appropriate that both should work at the same level of difficulty and progress at the same rate through the formal exercises.
The first 15 minutes of each lesson were spent working on the Schonell Essential Read and Spell workbook. Both subjects began at the beginning of Book II and progressed through to the end of Book III (a coverage of 2 years equivalent in reading competence); first and last lessons appear in Appendix B. The new words would first be read, with points accumulating on the counter for correct responses, and correction and explanation or instruction for wrong ones. The story passage was then read and timed, with errors noted and a point awarded per line read faultlessly. Exercises (mainly word recognition drill and comprehension) were selected; subjects had equal opportunities to earn points here and on the spelling test which followed.

In the latter half of each lesson the Dolphin series of readers was again used. Although given a choice, both subjects in fact read the same two books at the first level of difficulty, corresponding to Schonell Book II (although they read the Dolphin books in different orders), and the same one at the next standard. The E read first for five minutes, modelling accuracy and a reasonable expressiveness (e.g. by paying attention to punctuation, a failure to do which often results in not only atonal but incomprehensible reading aloud by backward readers). Subjects were then asked three comprehension questions and received points for correct answers. A point was earned for every two lines read by the subjects, regardless of errors; however errors of any significance were corrected by E or the
subject was required to retry the word. Thus more accurate reading in fact gained more points, in the fixed times allowed, although the main reinforcement was for the behaviour "reading" rather than for "getting everything right". A further three comprehension questions were then asked, daily tally of points marked on graph, reinforcement delivered or arranged as appropriate, and subject reminded of next lesson time before leaving.

All lessons were to be held at the same time every day for both subjects. However each week lesson 3 and 4 had to be held in the afternoon instead of the morning to avoid a clash with the subjects' library and Art periods. Although the teachers were quite willing for the subjects to forego these activities in favour of reading lessons, the E thought that the subjects might not be, and accordingly avoided such a conflict on their behalf. It seems reasonable to expect that attention and progress within a reading lesson could be affected by strong reluctance and/or resentment about being there. The change mentioned did introduce the possibility of diurnal variations, but inspection of Fig. 14 suggests that if anything, time of week had a stronger influence than time of day.

d) Reinforcement procedures

Social reinforcement was standardized as far as possible. On the principle that adult attention can be a powerful strengthener of child behaviour, E reserved special warmth and interest for improvement rather than difficulties, which
were dealt with in a friendly but businesslike manner.

The point system of feedback, which was known from previous experiments to have considerable reinforcing power of its own, has already been described; it remains to discuss the two other more contrived reinforcement treatments which were being compared.

(i) Material reinforcement

The following typed instructions were initially read to the subject, and were continuously available to him throughout the two week Material reinforcement condition:

"Every 10 points is worth one cent

You may either take your earnings in 1¢ pieces,

or use them to buy items from this price list:

5¢ sweets
10¢ scotch tape, marking pens, tin whistles
20¢ candles, balloons
30¢ farm animals, books, racing-car, notepaper
If something you would like is not on the list, ask Mrs. Winefield about it."

Thus the backup array was essentially similar to that of Expt. I, except that the option of taking money rather than goods was included.

At the start of odd-numbered lessons, E said "Today you earn for today and for tomorrow" (contingent reinforcement). The corresponding reminder for noncontingent reinforcement was "Today you get what you earned yesterday: x cents worth".
It was expected that thus making explicit the operation of different treatments, and running them consecutively, would maximize sensitivity to induced performance effects. However on the basis of Expt. V results and the experimental literature, the contingent vs noncontingent effect was expected to be small enough not to obscure the major comparison between extrinsic and intrinsic rewards.

(ii) Intrinsic reinforcement

The relevant instructions read:

"Every 10 points is worth 30 seconds
You may use this time that you have earned
either to read to Mrs. Winfield from any book not used in our lessons,
or to cut the next day's lesson short by that amount of time."

The deliberate provision of "intrinsic" reinforcement posed some problems. In Deci's sense, intrinsic motivation is demonstrated when an activity is carried out for no other apparent reason that enjoyment derived from the activity. "Opportunity to read more" was therefore chosen as the intrinsic reinforcer, with the option of reading less to demonstrate negative intrinsic consequences. It was expected that allowing the subject to choose the material, would ensure high "automatic" reinforcement for reading in Skinner's sense, i.e. the pleasure of discovering "how to" or "what happened next" (see Chapter II).
The use of materials not part of the regular lessons was expected to minimize any major role of transfer from the extra practice gained, to lesson performance. The books actually chosen by the subjects were in fact too different from lesson materials: (some of preschool and some of adult difficulty) for much positive generalization to be effected to the lessons. Ensuring that it was the time spent reading, rather than time spent out of class, or time spent with E, which was the effective reinforcer, was difficult. During the extra reading time, E supplied unknown words or meanings but was careful to give no extra social reinforcement whatsoever – or to behave in a reinforcing way when the subject announced whether he would add to or subtract from his next lessons' length. No points were earnable for the extra reading.

RESULTS

Table 16 shows pre- and post-treatment assessments of the two subjects, and Table 17 the weekly attitudinal measures. Table 18 summarizes other measures.
Table 16. Changes over 7 weeks of remedial lessons

<table>
<thead>
<tr>
<th>Measure</th>
<th>Subject D Pre-</th>
<th>Subject D Post-</th>
<th>Subject B Pre-</th>
<th>Subject B Post-</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.A.</td>
<td>10-10</td>
<td>11-0</td>
<td>9-10</td>
<td>10-0</td>
</tr>
<tr>
<td>Dominion WRT</td>
<td>78</td>
<td>81</td>
<td>70</td>
<td>86</td>
</tr>
<tr>
<td>Neale error %</td>
<td>8.02</td>
<td>6.88</td>
<td>11.73</td>
<td>10.0</td>
</tr>
<tr>
<td>Neale comprehension correct,of 16</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Neale rate (correct words per second)</td>
<td>1.08</td>
<td>1.19</td>
<td>.57</td>
<td>.77</td>
</tr>
<tr>
<td>Letter familiarity: correct per sec.</td>
<td>.51</td>
<td>.67</td>
<td>.25</td>
<td>.27</td>
</tr>
<tr>
<td>Voluntary reading per weekend</td>
<td>1 hour</td>
<td>2 hours</td>
<td>½-1 hour</td>
<td>¼ hour</td>
</tr>
<tr>
<td>New Words % correct</td>
<td>--</td>
<td>83.02</td>
<td>--</td>
<td>71.43</td>
</tr>
</tbody>
</table>
Fluency and percent of errors, both measured on the Schonell workbook passage, are shown from lesson to lesson in Fig. 14.

Table 17: Weekly semantic differential ratings of "Our lessons this week", and of "how I feel about reading now".

<table>
<thead>
<tr>
<th></th>
<th>Subject D</th>
<th></th>
<th>Subject B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pleasant-ness %</td>
<td>Potency</td>
<td>*Att. to reading</td>
<td>Pleasant-ness %</td>
</tr>
<tr>
<td>Base I</td>
<td>97</td>
<td>96</td>
<td>c</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reinforcement I</td>
<td>Material reinfor.</td>
<td>Intrinsic rft.</td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>98</td>
<td>88</td>
<td>e</td>
<td>59</td>
</tr>
<tr>
<td>(b)</td>
<td>100</td>
<td>99</td>
<td>e</td>
<td>67</td>
</tr>
<tr>
<td>Base II</td>
<td>100</td>
<td>65</td>
<td>e</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reinforcement II</td>
<td>Intrinsic rft</td>
<td>Material rft</td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td>99</td>
<td>100</td>
<td>e</td>
<td>70</td>
</tr>
<tr>
<td>(b)</td>
<td>99</td>
<td>99</td>
<td>e</td>
<td>72</td>
</tr>
<tr>
<td>Base III</td>
<td>100</td>
<td>90</td>
<td>e</td>
<td>70</td>
</tr>
</tbody>
</table>

* Where a = I hate it, b = I don't like it, c = It's all right, d = I like it, e = I love it.
Table 18: Other measures of reading skill and attitude to lessons

<table>
<thead>
<tr>
<th>Measure</th>
<th>Subject D</th>
<th>Subject B</th>
</tr>
</thead>
<tbody>
<tr>
<td>100-word sample of teaching material end of Base I</td>
<td>77</td>
<td>89</td>
</tr>
<tr>
<td>end of Base II</td>
<td>84</td>
<td>91</td>
</tr>
<tr>
<td>end of Base III</td>
<td>85</td>
<td>92</td>
</tr>
<tr>
<td>Overall gain on Dominion WRT: pre-treatment to followup (3 months)</td>
<td>11%</td>
<td>14%</td>
</tr>
<tr>
<td>Lessons of which subject needed reminding</td>
<td>Lesson 19</td>
<td>Lessons 1-9, 11, 21, 26-28 (i.e. most Mondays)</td>
</tr>
<tr>
<td>Post-treatment interview</td>
<td>Material</td>
<td>Material</td>
</tr>
<tr>
<td>Procedure rated best liked</td>
<td>reinforce-</td>
<td>reinforce-</td>
</tr>
<tr>
<td>least liked</td>
<td>base</td>
<td>base</td>
</tr>
<tr>
<td>most help</td>
<td>Intrinsic</td>
<td>Intrinsic</td>
</tr>
<tr>
<td></td>
<td>reinforce-</td>
<td>reinforce-</td>
</tr>
<tr>
<td></td>
<td>ment</td>
<td>ment</td>
</tr>
</tbody>
</table>
Statistical analysis

As both subjects worked at the same material each lesson, their efforts are comparable on any given occasion, despite the gradually increasing difficulty of the material over all the lessons.

a) Base weeks

2 x 3 analyses of variance were performed on the fluency and error scores generated by the two subjects over the three weeks of baseline (i.e. "no extra reinforcement") condition.

Neither fluency ($F = 2.28$), nor percentage of errors ($F = 84$), showed differences over time; in other words when no other reinforcement conditions were superimposed, feedback and low-intensity praise for improvement sustained accuracy at the same level despite an increase in difficulty equivalent to two years Reading Age (from beginning Essential Read and Spell Book II, to end of Book III).

As is evident on Fig. 14, fluency throughout showed a clear separation between the two subjects which was highly significant statistically ($F = 31.0458, p \leq .001$, in 2 x 3 AOV). Accordingly, the remainder of the analysis dealt with the two subjects separately re fluency. Error scores however showed no such separation between subjects ($F = .3955$ in 2 x 3 AOV).

b) Reinforcement conditions

Material vs intrinsic reinforcement for each subject's fluency was calculated by 't' tests for unrelated samples.
Fig. 14: Fluency and error percentage under different conditions

![Graph showing fluency and error percentage]

Fluency (correct words per second)

Error percentage

Subject D
Subject B
Differences were not statistically significant:

Subject D, t = .5438; Subject B, t = 1.5224, (df = 18).

Error % showed a significantly higher rate of errors under material reinforcement, for both subjects:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>ms</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>0.3367</td>
<td>1</td>
<td>0.3367</td>
<td>1.995</td>
</tr>
<tr>
<td>Reinforcement</td>
<td>15.1659</td>
<td>1</td>
<td>15.1659</td>
<td>8.9840*</td>
</tr>
<tr>
<td>S x rft.</td>
<td>2.1577</td>
<td>1</td>
<td>2.1577</td>
<td>1.2782</td>
</tr>
<tr>
<td>Error</td>
<td>60.7714</td>
<td>36</td>
<td>1.6881</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>78.4317</td>
<td>39</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* p < .005

c) Contingent vs Noncontingent material reinforcement

No difference was found for either fluency (Subject D, t = .7580; Subject B t = .3493, d.f. = 8), or error % (F(1,16) = 1.1412), between contingent and non-contingent material reinforcement.

d) Semantic differential ratings of the lessons

As can be seen from Table 17, the subjects showed most variability in their ratings of the potency, rather than pleasantness, of the lessons. Subject D's estimate of how "strong" and "large" the lessons had been, fell off sharply when material reinforcement was discontinued, but was
reinstated when intrinsic reinforcement was introduced.
Subject B's estimate of potency increased throughout the period
of Intrinsic reinforcement, fell off thereafter and was not
restored by the introduction of material reinforcement.
e) Other measures of reading achievement

As with d), no statistical analysis can be applied to
results from these two children, whose treatments differed
in sequence. Tables 16 and 18 summarize the information
gathered, which indicates a reliable and generalized
improvement for both subjects.

DISCUSSION

The fact that the subjects covered work equivalent to two
years of progress in reading during their 7 weeks of lessons,
with an approximately constant error rate, shows that skill
was being progressively acquired. There is unfortunately
no obvious way to quantify the increase in difficulty of the
reading material, apart from in terms of Reading Age.
A word-length count was in fact made; the average number of
letters was obtained for a sample of words (all the words in
the first three lines of the Schonell passage) read in
each lesson, and as expected a fairly consistent increase
in length demonstrated. However this difference was small
(from 3.74 letters average in the first week to 4.42 in the
seventh), and was also felt not to reflect at all, the
increasing complexity (e.g. of syllabification) and
infrequency of words read.

Both subjects commented in the post-treatment interview
that lessons during the Base procedure (the least liked for both), seemed shorter. In fact, the same time was available for reading (and earning points) with every treatment, and a separate analysis of the strong-weak and large-small ratings failed to support the subjects' judgment. Indeed, post-hoc statements about the lessons afterwards seem to bear at most an oblique relationship to behaviour within lessons. A further instance was both subjects praising the "useful things" (like sticky tape, etc.) which could be earned in the Material reinforcement periods, when they had both spent most of their points on sweets.

The fact that both expressed highest preference for the condition of material reinforcement does not necessarily conflict with the less favourable semantic differential ratings given to material reinforcement in Expt. I. This latter is a more covert measure, and in addition, Expt. I subjects were not comparing different reinforcement procedures with each other as were the present subjects.

It is interesting to speculate on the significance of the fact that both subjects chose goods rather than money, in this situation where a choice was available, except when a few spare "cents-worth" of points remained. Possibly money represents payment in a rather direct way common to the outside commercial world, whereas goods are perceived more as presents, and therefore more denotative of approval or liking.
Both subjects always chose to read more, rather than less, during the Intrinsic reinforcement condition. As previously mentioned, the books they read were not likely to have provided much specific transfer to the lesson material, although the extra practice in reading aloud, with feedback on errors, may have been of generalized benefit.

The major finding of Expt. VI was the increase in proportion of mistakes made in reading, under conditions of material reinforcement. This, and the argumentativeness about correctness which appeared in this period in the more articulate subject (D), support an interpretation of the effects of tangible rewards employing the concept of over-arousal, i.e. a degree of motivation greater than that at which the reading task can be most efficiently performed. The children enjoyed getting things for reading, as shown by both spontaneous and elicited comments, but the effect on actual task behaviour was deleterious, possibly due to super-optimal levels of striving. Whether subjective tension varied under different procedures cannot be assessed directly from the data collected. Indirect attitudinal measures (see Table 17), showed no less positive feeling for the intrinsic reinforcement, but an attribution of greater effectiveness.

That added material reinforcement is associated with a higher error rate than is feedback alone, had been found by Miller and Estes (1961). The negative correlation between money earned and word recognition improvement shown
in Expt. V (Groups Ct & Nct: Spearman rho = .67) appears to be a congruent finding.

The "intrinsic" reinforcement used in this experiment consisted essentially of allowing the subject more control of how much time was spent in a particular activity (reading aloud to E). That they always chose to increase this time does not necessarily show that the activity itself was enjoyable; other reinforcers such as avoiding regular classroom activities, or some verbal formula such as "Practice makes perfect", may have influenced behaviour.

For a child of 10, after several years of pressure and inducements to read well, and of exposure of different training and entertaining reading material, it may be that the most effective, or even the only effective determinant of reading (of particular print), will prove to be the "automatic" or content-dependent reinforcement therein. The concept of "intrinsic" motivation bears a poorly defined relation to this type of reinforcement.

To dismiss the question of how intrinsic motivation is acquired, as Deci does, seems to eliminate most of what is interesting in the area. That there are species-specific limitations and propensities in what can be learned must be recognized (Seligman & Hager 1972). Perhaps the primate cortex predisposes us to find certain activities intrinsically interesting (i.e. the activity has a high natural probability of occurrence and will act as a positive reinforcer). On the other hand perhaps subtle
socio-cultural reinforcers, both verbal and non-verbal, shape
behaviours of high resistance to extinction, which subsequently
occur with tenacity in the absence of obvious or extrinsic
incentives. These two possible explanations are not of
course mutually exclusive. It appears to be unparsimonious
to deal with intrinsic motivation as a separate
phenomenon, without reference to the existing frameworks
of explanatory possibility.
Chapter XIII. Summary of main findings, and conclusions

The conscious application of operant principles to everyday behaviour has become widespread as a therapeutic measure during the last ten years. One undesirable behaviour which teachers, parents and children often wish to modify, is underachievement in reading. Although reading is a highly complex cognitive skill, it can be readily analysed in operant terms. It consists of specifiable learned components and is maintained by certain consequences. Deficits in reading skill can thus be presumed to be due to inadequate reinforcers in the reading or learning-to-read situation, with consequent failure to acquire the basic behaviour-units (the correspondences between given visual stimuli and the appropriate verbal responses).

Remedy of such deficits is therefore obliged to begin with a consideration of how to provide effective reinforcement for progress, the evidence being that all, except possibly a few cases of neurological handicap, can become literate if properly taught and motivated. The question which is central to the present work then, is that of how to reinforce cognitive behaviour, with reading as an example having considerable practical importance.

Many successes have been reported, and were discussed in Chapters III and IV, in the modification of classroom behaviour by operant techniques. The results of interventions
to change overt, readily observed behaviours have on the whole
been more impressive than those aiming to produce changes in
levels of cognitive skill (O'Leary and Drabman, 1971). Two
clear reasons for this provide a convenient focus for the
present discussion.

Firstly, the problem of observation and measurement with
cognitive skills vastly complicates the conventional procedures
of baseline-measurement, application of selected reinforcers
for target behaviours, contingency reversal to check behavioural
control, followed by reinstatement of the conditions
maintaining desired behaviour. Where the behaviour in
question is, like reading, difficult to define let alone to
measure, standard practices may still be applied but their
effectiveness becomes difficult to assess.

Reading is a behaviour where relative performance has
extremely high significance, both short- and long-term, in
the individual's life. Failure to attain acceptable
standards may limit educational and vocational opportunities
and materially affect both objective and affective indices of
fortune. However, at the same time, reading exposes the
greatest weakness of the operant approach: the failure to
deal convincingly with aspects of human language, especially
"meaning". A behaviour of great importance thus coincides
with an area of inadequate basic "theory". Both practically
and conceptually then, reading behaviour presents a challenge
to the shaper of behaviour.
The second reason for comparatively poor success in this field, stems partly from the first and involves the issue of human reinforcement. The teacher who undertakes to praise a pupil or deliver tokens for sitting quietly, can in fact replace the old, maladaptive ways of getting her attention with more appropriate ways. There appears to be no inherent reason why the new behaviours should not stabilize under these conditions, and become a permanent high-probability characteristic of the child's repertorie.

In the case of academic learning, however, the positive consequences may mainly consist of remote future promises (e.g. of a good job); the backward reader therefore has little immediate incentive, and may well receive little teacher attention in a large class. Poor reading does not distract the other pupils as does disruptive behaviour. Further, in inculcating a cognitive skill, all the teacher can reinforce is "going through the motions": paying attention, completing assignments and improving test scores. Control of these factors usually implies, but is not synonymous with, increasing discriminative control of the textual behaviour by the print. Reading can be regarded as a fundamental or tool behaviour in academic learning, in that after the first few years when little else is taught, reading fades from visibility as a "subject": the skill is presumed acquired. There are few opportunities in the normal classroom, therefore, for a poor reader to obtain information about, or rewards for, his efforts.
Although an ingenious variety of reinforcers has been used to control children's behaviour in classroom settings, tokens have several advantages when dealing with reading. They can be delivered immediately after the desired response and thus have a cue function as well as potentially great sensitivity to small signs of progress. No distracting consummatory response is required of the recipient at the moment of delivery, yet much flexibility is possible regarding the back-up reinforcers for which tokens are ultimately exchangeable.

When experiments using tokens to reinforce progress in reading are reviewed, however, some confusion about the exact mode of operation of this type of reinforcer becomes apparent. Basically, doubts centre on the relative weights of the incentive and informational roles of tokens, and the implications of the answer for remedial training practice.

A second major gap in understanding derived from the hypothesis that the backward or failing reader may have learned maladaptively to associate the visual stimulus of print, not with the correct phonemes and hence language, but with anxiety. Thus it becomes necessary to explore the effects of punishment on the learning-to-read process, bearing in mind that the backward reader may respond differently to the added information provided by mildly aversive consequences for errors, compared to
the reader who is confident in his ability to meet the expectations of the adult reinforcement agents of his environment.

The major themes of the present work having been recapitulated, the findings from the series of experiments performed can be summarized and discussed.

1. Tokens for reading

The first need was felt to be a thorough investigation of earlier claims by such experimenters as Staats, of the efficacy of token reinforcement for reading. After some preliminary studies had been done, Experiment I showed fairly conclusively that tokens in themselves do not necessarily aid learning, if adequate feedback on progress is concurrently being provided. Children receiving equivalent lessons, feedback on progress and social reinforcers, were indistinguishable in results (i.e. reading improvement after 20 remedial lessons) from those who additionally received tokens. Amongst the token groups, however, those penalized (by token-deduction) for errors did better only while the token system was in force; the only statistically significant effect of material reinforcement was that the recipients enjoyed their lessons less than did those subjects "rewarded" with paper stars.

Although published accounts of unsuccessful token programmes are predictably few, the failure in this case could reasonably be attributed to two main factors:
(a) the use of middle-class subjects who were therefore presumably anxious to read well, and requiring not motivation but opportunity within a supportive, individualized coaching situation,
(b) the use of a between-subjects design where subjects received only one treatment and were ignorant of those received by others, thus avoiding the psychological effects of contrast in reinforcement conditions.

Caution can therefore be advocated before token programs are introduced as remedial reading aids, and support given to the views expressed by Ferster (1971, 1972) that an overwhelming benefit of token systems may be to affect how the teacher organizes learning material, and attends to small idiosyncratic signs of progress in the pupils.

2. Effects of response-cost for errors

While the subjects penalized for errors in Experiment I could have been expected on the basis of other learning research to do better than those who were not, the reversal of this result when the experimental conditions changed and tokens were no longer operating, was an unexpected and very interesting result. Experiments II, III and IV aimed to explore this phenomenon further, in order to make clear if possible whether explanations in terms of an initial depression of performance under punishment conditions, or a later resurgence of formerly-punished responses, were more valid.
Experiments II and III used multiple-choice comprehension materials, where performance was expected to be at a stable level and little if any learning was involved in the task, while in Experiment IV subjects were required to learn lists of paired-associates. With within- and between-subject designs, normal and retarded readers, and with material backups for the "point" tokens or without, replication of the response-cost effect could not be achieved. Two differences between Experiment I and those that followed it were regarded as likely to account for this discrepancy:

(a) The longer time span of Experiment I (13½ hours spread over a period of 10 weeks) compared with 10-30 minute interactions of Experiments II, III and IV, obviously allowed for much deeper impact on the subjects of the experimental procedures, and for the possible contamination of experimental effects by many outside influences.

(b) The personal significance and relevance to real success and achievement of the training in reading contrasted strongly with the patent artificiality and triviality, from the subjects' point of view, of the experimental tasks.

It was finally concluded, therefore, that the response-cost effect was specific to the conditions in which it was originally found: a long series of trials using a task highly relevant to the subject's life.
3. **Positive reinforcement for cognitive learning**

By providing a wide variety of backup reinforcers, and including the option of asking for desired articles, it had been hoped to maximize the potency of the material reinforcers of Experiment I. Yet, there were no differences in progress in reading between those reinforced with valuable goods and those reinforced with paper stars. This result raises again the issue (first discussed in Chapter II) of what exactly is being reinforced in a token remedial reading program.

Experiment V provided contingent token reinforcement for one group of retarded readers having ten remedial lessons, and awarded members of a matched group the same monetary backups non-contingently. The lack of resultant differences in reading performance or in word-recognition progress, can be interpreted as support for the interpretation that the behaviour reinforced is the attention and effort required of the subject in the training situation, rather than the strengthening of particular grapheme-phoneme associations.

As there had been a suggestion of a delayed effect of the material reinforcement in Experiment I, such that recipients seemed to improve in word recognition at a faster rate than non-recipients for several months after the end of lessons, a long-term follow-up was incorporated into Experiment V. No replication was obtained, although an interesting reciprocity of different measures of reading skill was demonstrated (see Section 4 below).
The contingently-rewarded subjects of Experiment V showed a reinforcement- rather than a task-orientation; they were more noisy and competitive than those subjects receiving either non-contingent reinforcement (who on the whole were puzzled), or those receiving feedback on progress alone. This observation, plus the significant devaluation of lessons by Experiment I subjects who got material reinforcement (and rated lessons as more "bitter" and "sad"), was followed up in Experiment VI.

Here two subjects were once again given a long series of individual remedial reading lessons. In contrast to earlier manipulations, subjects experienced several different kinds of reinforcement during the experiment. The first, fourth and seventh weeks of daily lessons were baseline conditions and provided feedback and low levels of social reinforcement. Material backup reinforcers were provided in Weeks 2 and 3 for one subject and 5 and 6 for the other, while in the other fortnight for each subject an attempt was made to program for "intrinsic" reinforcers, by allowing tokens to be exchanged for time spent in the relevant activity (reading), either by increasing or decreasing it.

The two subjects covered work equivalent of 2-years Reading Age during the experiment while fluency and percentage of errors remained constant. Within the material-reinforcement condition, the alternating sessions of contingent and noncontingent reinforcement were again found not to affect
performance. Despite an expressed preference for material rather than intrinsic reinforcement, however, error rates were reliably increased under the former condition.

This result, and those of the earlier experiments, are congruent with an explanation in terms of over-arousal rather than distraction as the mediator of the generally unfavourable effects of supplying material reinforcement. The exact nature of "intrinsic" reinforcement, however, remains obscure. Clearly Skinner's (1957) "automatic" or content-dependent reinforcement, and the effects of long sustained intermittent social reinforcers for reading behaviour, are plausible components. The whole question of intrinsic motivation is central to the debate not only on how to reinforce cognitive skills, but also on how to achieve self-direction in those subjected to behaviour modification interventions (Bijou, 1970; Deci, 1972a; Lovibond, 1971; Lovitt, 1970). As such, it deserves a more careful and detailed investigation than has hitherto been attempted.

4. Methodological considerations

Several noteworthy procedural matters can be commented upon as a result of the variety of designs employed. Firstly, Staats' (1970) assertion that control subjects are not required in a "developing research series" (p. 9) can be challenged. The inclusion of control groups added much
valuable information, particularly in preventing the premature acceptance of real treatment effects in experimental groups in the present work.

The value of follow-up investigations was similarly demonstrated and, as has repeatedly been stressed by reviewers (e.g. Gelfand and Hartman, 1960; Kazdin and Bootzin, 1972; O'Leary and Drabman, 1971), should be a foremost concern of all applied researchers.

The failure of different measures of reading skill to co-vary (e.g. the negative relationship found between word recognition and fluency improvements), and the nullification of a remedial-help spurt by post-treatment decline in improvement rate, may not be welcome findings but will vitally affect the design and measurements chosen. The ethical need to aim at irreversible changes is, of course, a further complication which, with its attempted solutions, has been illustrated by these experiments. Examples have been provided of a multi-dimensional attack on the problems elaborated earlier, concerned with the observation and measurement of reading behaviour. One conclusion reached was the inescapable desirability of obtaining measurements of natural response frequency and the extreme difficulty in collecting these without the long-term involvement and cooperation of parents and teachers.
5. Summary on the reinforcement of reading

What reinforcement can then be expected to influence further skill development? The provision of tangible reinforcement, suitably tailored to individual need, may increase the visibility of progress, perhaps foster positive emotional responses to print, and control effort-output. The degree of generalization of the effect cannot be predicted, being dependent primarily perhaps on how well the reinforcement gained from reading can compete with that from other activities. Here the concept of intrinsic or automatic reinforcement, with its attendant obscurity, becomes important. The difficulties in specifying its influence separately from that of extrinsic reinforcers, possibly on very infrequent schedules, have already been discussed.

It is consequently suggested that discrimination should be employed before introducing material reinforcers for cognitive learning. They may prove not only ineffective but deleterious. This caveat may hold particularly strongly for children past the age of first steps in academic learning. The best reinforcer for cognitive activity and problem-solving may well be of a cognitive nature itself: information on progress and content-dependent stimulation. The motivating potential of detailed knowledge of results should never be underestimated.
6. Suggestions for future research

The present series of experiments must be seen as more provocative than definitive. It is a beginning, partly destructive of prior misconceptions and partly indicative of more fruitful directions. Many obvious limitations are a function of the requirement that outside help to the researcher should be avoided. The teams of paid post-graduate observers and specially trained and counselled teachers often noted in research of this type could not be utilized. Several recommendations can nevertheless be made for future work likely to bear a high yield of valuable information.

Several variables held constant in the present research need to be systematically investigated. These would include:

(a) Social reinforcement manipulation. A variable which was here as far as possible kept at a moderate intensity for all subjects, was the verbal and non-verbal approval and praise which can be delivered contingently upon desired responses. Clark and Walberg (1968) systematically varied it and obtained significant between-group differences on a reading test.

(b) Age and mental age of subjects. It has been pointed out that children of 10 years old, such as in the present research, may be most susceptible to control of reading behaviour by reinforcers which are inherent rather than
extrinsic to the activity. The situation might differ with younger children, where social reinforcement from the teacher in particular might be expected to have more power than it does later. Similarly, social reinforcers from peers might be suspected as the most potent contingencies of all for adolescents.

(c) Source of reading disability. The subjects used in the work reported have been undifferentiated for the cause of their reading backwardness. Low general intelligence, poor schooling and frequent illnesses may have intermingled, although attempts were made to eliminate them as factors, as well as emotional disturbance, and the lack of visual-motor co-ordination usually regarded as diagnostic of a "specific learning disability". The presentation of material and the incentive system appropriate for the latter case, would be very different, presumably, from the regime for a rebellious, over-anxious, or under-taught child.

(d) Reinforcement deprivation. A replication of the current work with working-class rather than middle-class children is an obvious need. The differences found (e.g. Chapter V) between lower- and middle-class children in responsiveness to different reinforcements could be a function of sub-cultural values as well as of varying levels of affluence. Certainly, straight comparisons between children of differing socio-economic backgrounds
working for extended periods at a "real" learning task such as reading, could provide much insight into what class-bound attitudinal differences, if any, do exist (e.g. towards money and educational achievement), and thus aid the implementation of more effective educational programmes for the disadvantaged. Other needed research centres on fuller exploration of reinforcement functions.

(e) More information about the hypothesized conditioned emotional reaction to print in failing readers. This could be obtained by means of physiological measures, reports by the subjects using ratings and possibly interviews, and invaluably, direct observations of reading behaviour in natural situations. This last would obviously require motivated cooperation with the researcher by parents and teachers.

(f) Where anxiety appears to play a major disruptive role in reading performance, the possibly of desensitization, probably by a combination of imagery and in vivo practice, certainly needs fuller investigation.

(g) It is evident from the results reported here that different measures of reading skill do not necessarily co-vary unidirectionally; this finding may present an opportunity to apply multiple-baseline techniques to the problem of demonstrating behavioural control. Accuracy and fluency, for example, could perhaps be subjected to separate contingencies. An equipment improvement which
would in this case become indispensable would be a
calculator or computer so that immediate feedback could
be provided to the reader (and to the extrinsic reinforcement-
dispenser where appropriate).

Many of the foregoing suggestions for future research
could be best explored in a rather different context from
that of "pure" research, implying as they do a very detailed
analysis of the individual retarded reader. Extensive
diagnosis of the difficulty, interviews with parents and
possibly observation of family interactions, and the
likelihood of manipulating teacher-behaviour would all be
necessitated. Such interventions are most appropriate to
a clinical undertaking, or at least one where the researcher
has an additional clinical status. The collection of
small-sample data over an extended period of time in such a
setting seems to hold the brightest potential for further
elucidation of the basic variables studied in the research
reported.

In technologically advanced societies many alternative
media compete with print to provide both information and
amusement. Print as yet has advantages in portability
and easy reference, but also has the disadvantages of
requiring long preliminary training, and of lacking the
sensory immediacy and impact of film. The schools of the
future may teach little reading, but operant research on
reading will still have been worthwhile if understanding is derived therefrom about how and why children learn.

The theoretical and practical obstacles which were uncovered in applying token reinforcement to reading are not specific to this particular endeavour. It is obvious that further long-term experimental investigation of individual cases is needed, to elucidate the prediction and control of not only reading but many other linguistic and symbolic human activities.
Appendix A: Technical diagram of apparatus.

- Power supply
- Transformer
- Lamp 24V
- Fuse 0.5A
- Mains

Key I: Add
Key II: Subtract

4 digit up/down counter
Buzzer on/off

Buzzer 24V

+24V

2500μF 40V

Dispa/down

260.
Appendix B

(a) Expt I Teachers Rating Scale

As a primary school teacher you are in a unique position as regards forming an overall impression of the children and their capabilities.

Could you please tick the following as appropriate for

. . . . . . . . . . . . . . (Name)

1. Classroom behaviour

Compared with other children in your class this year, this child is

- better behaved than average
- about average in behaviour
- more disruptive than average
- among the 2-3 most disruptive in the class.

2. Interest in reading

How often would this child pick up a book and read it voluntarily?

- never, avoids printed matter if he can
- occasionally shows an interest
- keen and appears to enjoy reading.

3. Ability in reading

a) Compared with others in your class this year, this child is

- well below average in reading
- a bit below average
- about average

b) Compared with his/her number work, this child's reading is

- weaker
- at about the same level
- stronger.

4. Do you think that this child has changed in any of the following areas in the past three months? Please put a tick if you havenoticed an improvement, a cross for deterioration, and leave blank for no change.

- reading
- spelling
- other subjects (which? . . . . . . . . . . )
- attitude to schoolwork
5. (Where child has been having extra lessons from H. Winsfield)

As time went on, did this child's willingness to attend extra reading lessons increase, decrease, stay the same, don't know? (Please tick one)

6. Is there any other comment you would like to make?

N.B. Teachers filled out Questions 1-3 before and after each subject had undergone the 20 remedial lessons, and questions 4-6 (presented on a separate page) only after the treatment, or after an equivalent period of time in the case of non-subject.

(b) Expt I Post-treatment subject questionnaire

1. Here are some lines with pairs of describing words on the ends.
   Let's think about the subject OUR LESSONS.
   On the first line, if you put a cross at this end, it would mean you think that OUR LESSONS were large. If you put the cross at this other end, you think OUR LESSONS were small. Right in the middle would mean you had no opinion about it.
   Put a cross somewhere on each of these lines.

   OUR LESSONS were:
   1. (large) small
   2. (bitter) sweet
   3. (strong) weak
   4. (happy) sad

2. How do you feel about reading now? Read these answers to yourself then put a tick by the one that is true:
   I hate it
   I don't like it
   It's all right
   I like it
   I love it

3. Do you know any of the other children who have been having lessons with me?
(c) Examples of teaching material, Experiments 1 and VI.

Schonell, F.J. and Schonell, F.E. "Essential Read and Spell" 1968

Reproduced here are the first lesson of Book II and the last lesson of Book III.
CAMPING

'Will you spend a week-end camping with me?' said Dick to George.

'Yes,' said George.

'Where are you going to camp?'

'Oh, West Island. I have spent a week-end there before.'

So Dick and George packed a tent, a lamp, cups, a frying pan and a string bag.

They took sleeping bags to keep out the damp and the insects that sting.

They made these into a big bundle so that Dick could swing it on his back.

'You look as if you had wings, Dick,' said George.

'No,' said Dick. 'I'm a walking parcel. Put a stamp on my back and post me to West Island.'

A. Write the names of the things Dick and George took to West Island.

B. Write the words in each line that have the last three letters like the first one, in the small box.

<table>
<thead>
<tr>
<th>sent</th>
<th>want</th>
<th>rent</th>
<th>camp</th>
<th>spent</th>
<th>bent</th>
</tr>
</thead>
<tbody>
<tr>
<td>send</td>
<td>spend</td>
<td>stamp</td>
<td>mend</td>
<td>lend</td>
<td>bend</td>
</tr>
<tr>
<td>lamp</td>
<td>camp</td>
<td>jump</td>
<td>damp</td>
<td>stamp</td>
<td>hamper</td>
</tr>
<tr>
<td>wing</td>
<td>sting</td>
<td>limp</td>
<td>thing</td>
<td>spring</td>
<td>swing</td>
</tr>
</tbody>
</table>

C. Write the ones you could do.

1. Grow some wings.
2. Put up a tent.
3. Cook fish in a frying pan.
4. Light a lamp.
5. Sleep in a sleeping bag.
6. Put a stamp on your back.
NELSON'S VICTORY

A famous British sailor, Lord Nelson, was killed on his own flagship, the Victory, during the battle of Trafalgar in 1805. He won a great victory for England and added glory to her naval history, but his death was a great loss to England. It was such a waste of a wonderful man, for he was only 47 years old. He received a mortal shot from a sniper hidden in the rigging of a French warship. It was a pity that he did not live to see his fine victory. Even when he was dying he was calm and brave. He did not lose control, and in a soft voice, but without haste, said good-bye to his captain, 'Kiss me Hardy . . .'

Today the actual Victory, not a copy of it, is preserved at Portsmouth, in the south of England, so that plenty of people have had a chance to see it. It is, of course, an empty ship now. But if you want a taste of the past you can visit Portsmouth. Fastened to the deck, near one of the masts of the Victory, is a plate to mark the spot where one of England's greatest heroes met his death.

A. See if you can find out:
1. In whose reigns Nelson lived.
2. Where he was born.
3. What naval victories he had in addition to Trafalgar.
4. What injuries he received in his naval battles.

B. Answer these questions:
1. How did Nelson meet his death?
2. What was the name of Nelson's flagship?
3. Where is this ship now?
4. Why was Nelson such a great sailor?

C. Write these words in a column, and then write their opposites:
present, unfasten, an original, pull, find, scarcity, whisper, even numbers.

D.| Noun | Describing Word | Noun | Describing Word |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>glory</td>
<td>glorious</td>
<td>pity</td>
<td>pitiful</td>
</tr>
<tr>
<td>plenty</td>
<td>plentiful</td>
<td>taste</td>
<td>tasteless</td>
</tr>
</tbody>
</table>

Write out the pairs. Use the describing words in sentences.
### Appendix C

Detailed statistical results of Experiment 1

<table>
<thead>
<tr>
<th>Measure</th>
<th>NDNB</th>
<th>END</th>
<th>DNB</th>
<th>ED</th>
<th>TC</th>
<th>HIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-post Schonell WRT change</td>
<td>7.50</td>
<td>6.25</td>
<td>1.00</td>
<td>3.25</td>
<td>5.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Pre-post Dominion WRT change</td>
<td>4.75</td>
<td>2.50</td>
<td>7.25</td>
<td>8.00</td>
<td>8.50</td>
<td>-3.25</td>
</tr>
<tr>
<td>Schonell WRT gains per month (pre-test to follow-up)</td>
<td>-35</td>
<td>1.15</td>
<td>-40</td>
<td>1.10</td>
<td>1.00</td>
<td>53</td>
</tr>
<tr>
<td>Percent change in fluency (words per corrected sec.)</td>
<td>42.10</td>
<td>44.85</td>
<td>46.03</td>
<td>51.47</td>
<td>49.52</td>
<td>12.53</td>
</tr>
<tr>
<td>Reading age work equivalent (years)</td>
<td>1.54</td>
<td>1.65</td>
<td>1.41</td>
<td>1.58</td>
<td>1.72</td>
<td>-</td>
</tr>
<tr>
<td>Percent Fluency change on harder material</td>
<td>24.90</td>
<td>24.25</td>
<td>31.32</td>
<td>42.52</td>
<td>6.35</td>
<td>-</td>
</tr>
<tr>
<td>Same diff, Pleasantness</td>
<td>95.50</td>
<td>58.25</td>
<td>91.50</td>
<td>80.25</td>
<td>84.25</td>
<td>-</td>
</tr>
<tr>
<td>Potency</td>
<td>77.75</td>
<td>55.25</td>
<td>86.75</td>
<td>69.50</td>
<td>81.00</td>
<td>-</td>
</tr>
<tr>
<td>New word test</td>
<td>86.00</td>
<td>78.25</td>
<td>77.25</td>
<td>85.00</td>
<td>70.00</td>
<td>-</td>
</tr>
<tr>
<td>Total errors</td>
<td>102.75</td>
<td>103.75</td>
<td>98.25</td>
<td>90.75</td>
<td>109.00</td>
<td>-</td>
</tr>
<tr>
<td>Average points per lesson</td>
<td>70.8</td>
<td>69.6</td>
<td>63.6</td>
<td>65.8</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
### 1) Schoenell WRT Change (Pre-Post treatment)

<table>
<thead>
<tr>
<th>Source</th>
<th>M.S.</th>
<th>d.f.</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>12.667</td>
<td>15</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Deductions</td>
<td>90.250</td>
<td>1</td>
<td>12.5202</td>
<td>.0043</td>
</tr>
<tr>
<td>Backups</td>
<td>1.000</td>
<td>1</td>
<td>.1387</td>
<td>.7160</td>
</tr>
<tr>
<td>D x B</td>
<td>12.250</td>
<td>1</td>
<td>1.6994</td>
<td>.1950</td>
</tr>
<tr>
<td>Error</td>
<td>7.208</td>
<td>12</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

### 2) Dominion WRT Change (pre-post treatment)

<table>
<thead>
<tr>
<th>Source</th>
<th>M.S.</th>
<th>d.f.</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>10.783</td>
<td>15</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Deductions</td>
<td>64.000</td>
<td>1</td>
<td>8.8786</td>
<td>.0112</td>
</tr>
<tr>
<td>Backups</td>
<td>2.250</td>
<td>1</td>
<td>.3121</td>
<td>.5321</td>
</tr>
<tr>
<td>D x B</td>
<td>9.000</td>
<td>1</td>
<td>1.2486</td>
<td>.2857</td>
</tr>
<tr>
<td>Error</td>
<td>7.208</td>
<td>12</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
3) **Ratings of pleasantness (semantic differential)**

<table>
<thead>
<tr>
<th>Source</th>
<th>M.S.</th>
<th>d.f.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>503,450</td>
<td>15</td>
<td>-</td>
</tr>
<tr>
<td>Deductions</td>
<td>324,000</td>
<td>1</td>
<td>.9258</td>
</tr>
<tr>
<td>Backups</td>
<td>2352.250</td>
<td>1</td>
<td>6.7215</td>
</tr>
<tr>
<td>D x B</td>
<td>676.000</td>
<td>1</td>
<td>1.9317</td>
</tr>
<tr>
<td>Error</td>
<td>349.958</td>
<td>12</td>
<td>-</td>
</tr>
</tbody>
</table>

4) **Average points per lesson**

<table>
<thead>
<tr>
<th>Source</th>
<th>M.S.</th>
<th>d.f.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>239.747</td>
<td>15</td>
<td>-</td>
</tr>
<tr>
<td>Deductions</td>
<td>1210.000</td>
<td>1</td>
<td>6.4041</td>
</tr>
<tr>
<td>Backups</td>
<td>10.000</td>
<td>1</td>
<td>.0529</td>
</tr>
<tr>
<td>D x B</td>
<td>108.900</td>
<td>1</td>
<td>.5764</td>
</tr>
<tr>
<td>Error</td>
<td>188.942</td>
<td>12</td>
<td>-</td>
</tr>
</tbody>
</table>
Appendix D

Materials used in Experiments II and III

1. Warm-up exercise (common to both experiments)

1. Jack and Jill went up the (road, hill, creek) to fetch a (pail, nail, bit) of water. Jack fell (over, down, hard) and broke (his, her, their) crown (Janet, Joan, Jill) came tumbling after.

2. Twinkle twinkle (little, tiny, big) star. How I (question, wonder, pretend) what you are. Up above the (roof, world, mountain) so high like a (pearl, necklace, diamond) in the sky. Twinkle (blink, twinkle, shine) little star. How (I, they, we) wonder what you are.

3. Jingle bells, jingle (alarms, bells, money). Jingle all the (way, day, night). Oh what (excitement, pleasure, fun) it is to ride in a one- (cow, horse, horse-power) open sleigh.

II. Multiple-choice comprehension material for Experiment II.

Block I

The Cutty Sark

1. One day in London I went for a trip by (camel, launch, balloon) down the river (Thames, Amazon, Murray). There I saw a fine old sailing ship named the Cutty Sark. The Cutty Sark, a sailing ship of 963 (pounds, feet, tone), was launched on Monday 23rd November, 1869. She was built by a (team, term, couple) of craftsmen in the days when (rowing, steam, plastic) ships were new. It was a treat to see her lovely (fim, figure, shape) and the (gleam, gloom, noise) of the ship’s bell in the sun. In those days, sailing ships or (clappers, coppers, clippers) had an (exciting, elegant, enormous) race around the Cape of Good Hope bringing (water, tea, rockets) to England from India. The Cutty Sark never (one, won, wind) the race but once she was in front for a (second, month, year). Then one day there was a scream from the (sailors, soldiers, drivers). The ship had lost her (rudder, rubber, robber) in a gale. Things looked (great, grand, grave), but the carpenter worked like a (shave, slave, shake) and made a new rudder. It was a wonder he was not (drowned, worn out, happy) while fitting it. It was a (heavy, fine, close) shave.

The Cutty Sark

2. The Cutty Sark was soon on her way again, (flying, skipping, sailing) round the (globe, stars, suburb). It was a fine day, with a stiff (leg, handle, breeze) beginning to play (among, between, above) the sails, making them (flatter, flutter, flitter). The Cutty Sark rode well, for she was loaded with (tea, T.V. sets, rubbish). The crew off duty were playing (Bingo, hide-and-seek) cards on the upper deck and cracking a joke now and then. All of a (sudden, suddenly, suffer) the Captain saw a raft (flickering, flapping, floating) on the water close to the ship.
He looked again, this time paying more (money, attention, time). On the raft he could see a tiny baby, lying asleep in a (haystack, basket, coat). "Sweer in danger," he cried, "over there, there, where" on that raft. At first the men could not hear what he was (saying, singing, shouting). So they saw the raft with the (kitten, tea, baby) staying so still in his basket. "Lower a (way, bottle, boat)," cried the captain, "and row to the raft." The men were (quiet, sad, quick) to offer help. They could not bear to see a baby (walrus, faintly, suffer). "We must pray that we can (save, shave, shake) him," they said.

The Cutty Sark

3. The men lowered the (jib, lifeboat, glasses) and pulled towards the raft. They could hear a faint (guitar, banjo, noise). The baby was (walking, wakening, taking) up as they lifted the basket into the boat. They looked into the (oven, envelope, sky) and then they got a (knot, shock, cake). Dark clouds had blown (up, around, below) in and seemed only a moment. There was a (peal, peel, pill) of thunder and the sound (hust, burst, burst) upon them. It was (plain, plane, plain) that they must get the baby to safety quickly, or the (raft, rescue, time) would be in vain. The men lowered rope and began to (pull, push, haul) up the basket. At first it struck the side and (stacked, stock, stuck) but soon swift and (shore, certain, sure), they hauled the baby to safety. "We must look after (it, him, them) the sailors said, but we must not (eat, spoil, spangle) him." On board, they wondered how they could (catch, buy, obtain) the right food for him. The storm ended and as they (talked, laughed, sank) there was a beautiful (rainbow, mirage, jet plane) in the sky. "A good luck (signal, sign, sigh) for baby," said a sailor.

Block II

At the Ballet

1. Jim and Jane were going to the State Ballet. A ballet of dances, it was not singing, but (dancing, opera, jazz). They looked (around, behind, forward) to going out for they liked the (smell, feeling, colours) of art in a crowd. As they began to leave (quickly, home, mother), a cold wind was blowing, bringing (sunshine, darkness, rain) with it. When they arrived at the (theatre, cinema, church) it was time for the ballet to begin. Then the music started, the curtain (fell, divided, rose) and the ballet 'Swan Lake' had (began, begun, begging). The story was about a princess called Odette. She and her playmates had been turned into swans by an evil (spirit, devil, misfortune). Only at night did they change again to young (women, swans, ballerinas). The spell could be broken only by the faithful love of a young (wizard, duck, man). One day a young prince, Siegfried, fell in (love, the lake, argument) with Odette, who warned him that he must always be (sweet, faithful, fantastic) to her. But the evil spirit (tricked, tricked, triggered) Siegfried into being unfaithful. He changed his daughter into the (sister, house, form) of Odette, and Siegfried mistook her for the real Odette. Hearing this Odette, in despair, threw (tears, flowers, herself) into the Swan Lake. And Siegfried in his (heroism, sadness, excitement) killed himself. But his (sacrifice, honour, trust) broke the spell and the two lovers went away together to another world.
Growing Grapes

2. You can (sow, sew, show) a grain of wheat in wet straw or cotton (cloth, thread, wool). If you keep it damp it will grow a tiny (shoot, chute, foot) like a claw. The same thing happens with a pea or a (turnip, turkey, bird) seed, but of course they grow much better in a (saucer, field, farm). But you cannot grow grape (plant, wine, vines) from seed. New grape vines are (grown, grown, grew) from cuttings of old grape vines. The vines are often trained (by, to jump, along) wire fences. When the vine is (red, green, old) enough it bears grapes - black, light green or (emerald, sapphire, ruby). Some grapes are sent to market if they (travel, feel, cook) well. Some are crushed and placed in (vats, bottles, pits) where they remain to be made into (vinegar, wine, raisins). Many people work in the vineyards, when the grapes are (ripe, ripen, ripened). When all the grapes are picked the (people, vines, carts) return home. But when they depart they (often, seldom, infrequently) take some bottles of (wine, Fanta, vodka) with them.

The Captain's Dogs

3. The Cutty Sark was in port, waiting for the (tide, captain, luck) to turn. Her Captain was on (bored, board, broad) and so were his dogs. You would not (declare, plead, think) that an animal would like life at sea, (would, will, have) you? These dogs were reared on board (for, with, by) the Captain. He wanted his dogs to be (heroic, polite, tasty), but he would not punish them if they were (disterpered, disappointed, disobedient). He would pardon and (pat, forgive, forget) them, because in a way this ship was like a (prison, cell, sill) to them. They could not run about as they would on (grass, board, shore). It is hard to (believe, trust, imagine) a monkey or a donkey or a lion taking to life at sea, but these (pets, dogs, cats) liked it. At last the Cutty Sark unfurled her (umbrella, newspaper, sails) and away she glided down the (track, river, bank). At one place they (past, passed, pasted) a swan. One of the dogs looked at it as (though, thought, through) he was making rude (poetry, proverbs, remarks) about it. He kept barking and barking at it, until the ship had (reached, finished, mapped) the sea coast.

Block III

Music for the Count

1. There was once a Count who lived in a large house, with marble (collections, stairs, servants). In his stable he had (beautiful, underfed, cowboy) horses. But he was (sometimes, medically, never) ill and could not sleep. This (nervous, anxious, worried) him. It was like a (horn, thorn, thistle) in his side. One day he said, "I wish someone (could, should, can) write some (words, hymns, music) for me. It must be in a simple form, but I would like something lively to (lift, cheer, hurry) me up as I toss and tumble in (anguish, the hay, bed)." A (musician, musician, conductor) heard him say this and he said to the (count, earl, duke), "I am a stranger to you, but if you will (understand, trust, flatter) me, I can compose some strange but (soothing, swooning, smooth) music by March or April or May. I like to (write, invent, play) music and I will not (send, charge, give) you a fee." When the Count heard the music, he was (very, excessively, rather) happy, and so he sent the composer a golden (goblet, gobbler, goblet) filled with one hundred gold (sunflowers, rhymes, coins).
Two hundred years (before, ago, once) Dirty Dick (had, not, was) in London. His real (home, address, name) was Nat Bentley. He was always very neat and clean until one (cloudy, calm, bright), stormy day, the lady he was going to marry died suddenly. "I cannot understand why this (should, might, would) happen," he said. He had planned a (party, sandwich, poodle) for her that very night. When he heard the sad (messages, news, paper), he closed forever the room where the feast had been set (hard, softly, ready), and said, "This must never be (opened, eaten, consumed) again." After that (we, she, you) never saw anybody in his shop. Nobody would go there (unless, because, however) they had to. He gave up sport and all that (variety, sort, sport) of thing, and sat in his dusty shop with only one (light bulb, candle, spotlight) to light his room. On Thursday, Friday and Sunday he (cleaned, burnt, closed) the shop and went for long walks in the (frosty, first, forestry) air. For fifty years nobody (polished, turned, saw) the handle of the closed door, and when he (died, departed, terminated) they found dirty cups, and plates in the (sink, room, shop), and cobwebs and dirt and dust. No (surprise, example, wonder) he got the name of Dirty Dick!

The Lazy Maids

There was once an old (baby, lady, laddie) who used to wake her maids every (morning, afternoon, evening) when the cock crowed. They were always very (quiet, crushed, sleepy), and some of them did not even (curl, comb, colour) their hair before they started work. They (are, were will be) busy all day. She said to one of them, "You clean that (dusty, old, boring) stove," to a second, "You cook the (sheep, lamb, calf) for dinner." To a third she said, "You go to market." It was a long, (steady, ready, stealthy) climb up a hill to market, without any (comfort, shadow, shelter) from the hot sun, and the maid would become very (thin, pale, tired). A (fourth, fifth, sixth) maid would have to sew and darn, perhaps mend a hole in the thumb of a (finger, hand, glove), or make a cushion cover for her (master, mistress, maid). None of them dared to (eat, sweep, soak) even a crumb of bread without being told. So one day they killed the (fowls, sparrows, cock), thinking that they could then stay in their warm (house, beds, coats) a little longer. But after that the (nice, old, tiny) lady woke them at midnight! "We did not know how (lucky, happy, joyful) we were before", the (wrote, sang, cried).

Materials used for Experiment III

Growing Grapes

You can (sow, sew, show) a grain of wheat in wet straw or cotton (cloth, thread, wool). If you keep it (clean, damp, dark) it will grow a tiny (shoot, chute, foot) like a claw. The same thing happens with a pea or a (turnip, turkey, bird) seed, but of course they grow much better in a (saucer, field, farm). But you cannot grow grape (plants, bushes, vines) from seed. New grape vines are (groan, grown, grew) from cuttings of old grape vines. The vines are often trained (by, to jump, along) wire fences. When the vine is (fed, green, old) enough it bears grapes - black, light green or (emerald, sapphire, ruby). Some grapes are sent to market if they (travel, feel, cook) well. Some are crushed and placed in (bottles, vats, pits) where they remain to be made into
(vinegar, wine, raisins). Many people work in the vineyards, when the grapes are (ripe, riper, ripened). When all the grapes are picked the (people, vines, carts) return home. But when they depart they (seldom, immediately, often) take some bottles of (vodka, wine, lemonade) with them.

Three experimental passages:

The Cutty Sark I

One day in London I went for a trip by (camel, launch, balloon) down the river (Thames, Amazon, Murray). There I saw a fine old sailing ship names the Cutty Sark. The Cutty Sark, a sailing ship of 903 (pounds, feet, tons), was launched on Monday 23rd November, 1869. She was built by a (team, few, couple) of craftsmen in the days when (rowing, steam, plastic) ships were new. It was a treat to see her lovely (film, figure, shape) and the (gleam, gloom, noise) of the ship's bell in the sun. In those days, sailing ships or (clappers, croppers, clippers) had an (exciting, elegant, enormous) race around the Cape of Good Hope bringing (water, tea, missiles) to England from India. The Cutty Sark never (one, won, beat) the race but once she was in front for a (second, month, year). Then one day there was a scream from the (sailors, soldiers, drivers). The ship had lost her (rudder, rubber, robber) in a gale. Things looked (great, grand, grave), but the carpenter worked like a slave and made a (good, strong, new) rudder. It was a wonder he was not (drowned, worn out, happy) while fitting it. It was a (heavy, fine, close) shave.

The Cutty Sark II

The Cutty Sark was soon on her way again, (flying, skipping, sailing) round the (globe, stars, suburb). It was a fine day, with a stiff (leg, handle, breeze) beginning to play (among, between, above) the sails making them (flutter, flutter, flitter). The Cutty Sark rode well, for she was loaded with (T.V., sets, tea, rubble). The crew off duty were playing (Bingo, hide-and-seek, cards) on the upper deck and cracking a joke now and then. All of a (quickly, suddenly, sudden) the Captain saw a raft (flapping, swimming, floating) on the water close to the ship. He looked again, this time paying more (money, attention, time). On the raft he could see a tiny baby, lying asleep in a (hammock, basket, coffin). "Baby in danger," he cried, "over (here, there, where) on that raft." At first the men could not hear what he was (saying, singing, sighing). Then they saw the raft with baby staying so (still, small, long) in his basket. "Lower a (way, bottle, boat)," cried the Captain, "and row to the raft." The men were (quiet, amazed, quick) to offer help. They could not bear to see a baby (walrus, suffer, faintly), "We must pray that we can (catch, save, find) him," they said.

The Cutty Sark III

The men lowered the (lifeboat, glasses, jib), and pulled towards the raft. They could hear a faint (lady, banjo, noise). The baby was (walking, waking, taking) up as they lifted the basket into the boat. They looked into the (oven, envelope, sky) and then they got a (present, shock, cake). Dark clouds had blown (up, around, below) in what seemed only a moment. There was a (peal, peal, pill) of thunder and the storm (burst, burst, burnt) upon them. It was (plain, plane, plan) that they
must get the baby to shelter quickly, or the (raft, rescue, time) could be in vain. The men lowered rope and began to (pull, put, move) up the basket. At first it struck the side and (stacked, stuck, stuck) but soon, swift and (shore, certain, sure), they hauled the baby to safety. "We must look after (it, him, them) the sailors said, but we must not (eat, spoil, spangle) him." On board, they wondered how they could (catch, buy, obtain) the right food for him. The stage ended and as they (talked, laughed, sunk) there was a beautiful (shower, jet plane, rainbow), in the sky. "A good luck (sign, signal, sight) for baby," said a sailor.

Dirty Dick

Two hundred years (before, ago, once) Dirty Dick had a shop in London. His real (home, address, name) was Nat Bentley. He was always very neat and clean until one, (cloudy, calm, bright), stormy day, the lady he was going to marry died suddenly. "I cannot understand why this (should, might, would) happen," he said. He had planned a (dinner, party, luncheon) for her that very night. When he heard the sad (messages, newspaper), he closed forever the room where the feast had been set (hardly, softly, ready), and said, "This must never be (opened, eaten, consumed) again." After that (we, she, you) never saw anybody in his shop. He gave up sport and all that (variety, sort, effort) of thing, and sat in his dusty shop with only one (light bulb, candle, spotlight) to light, his room. On Thursday, Friday and Sunday he (claimed, liked, claimed) the shop and went for long walks in the (frosty, first, forest) air. For fifty years nobody (polished, turned, saw) the handle of the closed door, and when he (died, departed, emigrated) they found dirty cups and plates in the (sink, room, shop), and cobwebs and dirt and dust. No (surprise, example, wonder) he got the name of Dirty Dick!
Appendix E

Experiment II - Detailed instructions

1) Introduce self.

You'll be seeing me just once, and we're going to be doing some reading exercises.

It's not a test; I want to see how many points you can get. You get points, (Demonstrate) which show up here on the counter, when you give the right answer.

Let's have a look at (warm-up sheet). Can you read out what it says? Which of the words in the bracket is the right one? Put a line under it. You must only put a line under one of them. You would get a point for that, like this. Now finish the rest of that verse.

In the next rhyme, I'll show what will sometimes happen if you underline the wrong word. Let's say you forgot the rhyme and underlined 'tiny' rather than little. Then this happens - (demonstrate) - a buzz and you lose one of your points. I'll let you know when to expect buzzers and points off. Finish the rhyme.

Now do the last rhyme all by yourself for practice.

Here are some passages of reading where you have to choose the best, most sensible word out of the ones in the bracket, each time. We'll do five passages and then have a rest. Don't spend too long over any one word, but do underline one of the words in every bracket.

Experiment III

2) Name?

"We have some reading passages here. The idea is to choose the right word out of the ones in the brackets.

This (showing warm-up exercises) will show you what it's like.

Jack and Jill went up the . . . road, hill or creek? (E underlines subject's response - continue to end of rhyme).

"Now you do number two for yourself. Read it to yourself and underli the best word to come next." Then, "Good, you've got the idea fine."

"Now here's the first passage, 'Growing Grapes'. Go ahead and see how well you can do it." (timed with stopwatch).

Then

1) for Controls: good. Now the next 3 passages are all about a ship called the Cutty Sark. Go ahead and do them the same way.

At end: Good, now there's one last passage to do, 'Dirty Dick'.

Off you go.
2) for A's

Good. Now in the next 3 passages, which are all about a ship called the Cutty Sark, you can get points for your answers. Whenever you choose the right word, you get a point (demonstrate) - the number of points you have appears in this window. You have to see how many points you can collect. Remember, there's one point for each right answer. Off you go.

3) for B's

... and when you choose the wrong answer you lose a point, like this (demonstrate). So it's (demonstrate click) and point on for right answers, and (demonstrate buzz) and point off for wrong. See how many points you can collect.

A's and B's at end of Cutty Sark passages:

How many points? That's very good.

Now for the last passage, 'Dirty Dick', we go back to not having points. (Switch off generator and remove counter from desk). In that way it's like the first passage. Go ahead.

At end, for all:

That's all, thank you very much for helping me with this. Please don't talk about it with your friends, will you?

3) Experiment IV

For all subjects: Ask name, age and birthday. Then (with suitable pauses and emphasis) "I want you to learn some pairs of words. First we'll go through the list and I'll read out both words in each pair, and then we'll go through several times showing you the first word and you guessing the second word each time. You have a short period of time, like 1-2, to guess the second word, and the rule is that you must guess something each time. You can't possibly be right if you don't guess but you might be if you do.

For A's and B's: Every time you guess right, you get a point. The number of points you have appears in this window (demonstrate).

Add for B's: You have a bank balance of 20 points to start with. When you guess wrongly or, if you don't guess at all, you lose a point (demonstrate).

For A's and B's: The aim is to see how many points you can end up with at the end of a list.

For all subjects: O.K.? So first we'll go through showing you both words - then you'll see the first word and have to guess the second. Any further queries of child answered with "I think you'll get the idea once we start."
For first list:
In this list the first word in each pair is an animal, and the second is a number from 1-20.

Second list: In this list the first word in each pair is something you use for eating, and the second word is a country.

Third list: In this list the first word of each pair is a colour, and the second is a preposition, or joining word.

For A's and B's:

After the first list:
"That's good. Let's see how many points you can have by the end of the next list."

After second list:
"That's 'y' points that time. Good. Now for the last list we turn this off (do so) and see how you go without points."

For all subjects:

Reply to any self-deprecatory comments by subject re own performance with - "Most people find it harder than they think it will be; you're doing fine."

At end: "That's all. Thanks very much for helping me with this. Please don't talk about it with the other children, will you?"

4) Experiment V

Introduce self, and children to each other.

"We are going to be meeting - every morning at (time) for two weeks - i.e., for 10 sessions - of reading practice.

I'll read to you and you'll read to me. You get marks for how well you read - one mark for every line you read correctly. Your marks show in this window and when you've read I'll mark your graph or progress chart here (demonstrate) to show how you're going - the aim of course, is to make your graph go upwards, i.e., to get better as we go along.

Add for Ct group

You can earn cents according to how many marks you get. 1c for every 10 marks, to the nearest mark. You'll get these at the end of the session each time.

Add for Nct group

Each time you come you'll get some cents. How many you get will vary from day to day - I'll tell you how many at the start of the session and you'll get them at the end."


Buchwald, A. M. Variations in the apparent effects of "Right" and "Wrong" on subsequent behaviour. *Journal of Verbal Learning and Verbal Behaviour*, 1962, 1, 71-78.


Clark, M., Lachowicz, J. and Wolf, M. A pilot basic education programme for school dropouts incorporating a token reinforcement system. Behaviour Research and Therapy, 1968, 6, 183-188.


Hewett, F. M. Teaching reading to an autistic boy through operant conditioning. The Reading Teacher, 1964, 17, 613-618.


Mira, M. Results of a behaviour modification training programme for parents and teachers. Behavior Research and Therapy, 1970, 8, 309-311.


Schonell, F. J. Backwardness in the basic subjects. Edinburgh, Oliver and Boyd, 1942.


Shores, R. E. Performance of learning disabled and normal children from differing social classes under tangible or intangible reinforcement conditions. Exceptional Children, 1969, 25, 643-644.


Stauffer, R. G. Thorndike's "Reading as reasoning": A perspective. Reading Research Quarterly, 1971, 6, 443-448.


Vernon, M. D. *Specific dyslexia*. *Slow Learning Child*, 1965, 12, 71-75.


Walker, H. M. and Buckley, N. K. "Effects of reinforcement, punishment, and feedback upon academic response rate." 
_Psychology in the Schools_, 1972 (8), 2, 195-199.

Walker, H. M. and Buckley, N. K. Programming generalization and maintenance treatment effects across time and across settings. 

Warburton, F. W. and Southgate, V. 1.5.3. As independent evaluation. 

Ward, M. H. and Baker, B. L. Reinforcement therapy in the classroom. 

Wardenberg, H. L. The teaching of reading. 
_Albany, the University of the State of New York, 1974._

Walsk, B. H. The application of Proseck's generalization on reinforcement to the management of classroom behavior. 

Walsk, B. H. and Simons, J. T. Management of small-group behavior within a first grade classroom. 

Weiner, H. Human S-R responding as a function of the type of reward (money vs. points) and the presence or absence of a noncontingent monetary wage. 
_The Psychological Record_, 1972, 21, 497-500.

White, R. W. Motivation reconsidered: The concept of competence. 
_Psychological Review_, 1959, 66, 297-333.

Whitlock, C. Note on reading acquisition; An extension of laboratory principles. 

Whitlock, C. and Bushell, D. Some effects of "back-up" reinforcement on reading behavior. 

Winer, M and Cromer, W. Reading and reading difficulty: A conceptual analysis. 

Williams, J. P. From basic research in reading to educational practice. 

Willis, J. W., Morris, B. and Crowder, J. A remedial reading technique for disabled readers that employs students as behavioral engineers. 
_Psychology in the Schools_, 1972, 9, 67-70.

Winer, B. J. Statistical principles in experimental design. 


