**DOCTOR OF PHILOSOPHY**

# Stimulus equivalence and developmental dyslexia

Morgan, Jane L.

*Award date:*
1996

*Awarding institution:*
University College of North Wales, Bangor

**University of Wales**

# Stimulus Equivalence and Developmental Dyslexia

**Jane L. Morgan**

**Ph.D.**
**1996**

## Acknowledgements

My first thank-you must go to the members of my thesis committee without whom this thesis would, quite simply, not exist. To Dr. John Everatt and Dr. Nick Pidgeon for their invaluable help with the 'dreaded' statistics; to Prof. Miles for his expert insights into dyslexia and his willingness to provide a much-needed objective opinion; but most of all to Dr. Neil Dugdale for his ceaseless guidance, encouragement and patience.

I am also indebted to Lisa and Patrick for helping me collect some of the data and to all members of the School, too numerous to mention, who knowingly or unknowingly contributed to the progress of this Ph.D.

Thank-you to all the students and in particular the members of the general public who gave up their free-time to act as my subjects. I am especially grateful to all the dyslexic individuals who willingly volunteered to come along and endure hours of tests which were cruelly designed to play on their weaknesses.

Lastly, a huge thank-you belongs to my family, boyfriend and friends whose unconditional material and financial support carried me through my time as a postgraduate and specifically this last year.

I hope the end-product sufficiently justifies all your efforts.

## Summary

The Naming Hypothesis (as detailed by Dugdale and Lowe, 1990; Horne and Lowe, 1996) states that in order to successfully demonstrate stimulus equivalence subjects must name the stimuli. Horne and Lowe quantify this by stating that if the subjects assign common names or name intraverbally equivalence will emerge.

It was decided to evaluate the role of naming in stimulus equivalence by employing a population of subjects who are developmentally stable, of average or above average intelligence, but characterised by a deficiency in naming. For this purpose developmentally dyslexic adults were employed.

It was hypothesised that if naming is indeed necessary for the emergence of stimulus equivalence then dyslexic subjects should differ from control subjects on the equivalence test session. Specifically, the dyslexic performance should be more prone to errors (Snowling, Wagendonk and Stafford, 1988) and longer response latencies should be observed on the test session trials, in line with the dyslexic tendency to be significantly slower in naming (Fawcett and Nicolson, 1994; Watson and Brown, 1992).

Five matching-to-sample experiments were undertaken to compare the two groups' performances using a variety of stimuli and protocols which did or did not encourage the subjects to name. No statistically significant differences were found between the groups on the tests of equivalence although more dyslexic subjects were found to persistently fail.

Looking across the studies it was found that subjects' performances on Experiments 1 and 2 (in comparison with Experiments 3, 4 and 5 which were explicitly verbal) were consistent with them using a nonverbal strategy which presents the possibility that equivalence can be demonstrated without naming. The data from Experiments 3-5 also indicated that intraverbal naming is not always sufficient to bring about equivalence.

It was concluded that these data question the predictions made by Horne and Lowe (1996) and point to the need for more research regarding the role of naming in stimulus equivalence.

# Contents List

# Chapter One

## Stimulus Equivalence

# Stimulus Equivalence

## A Brief History

Stimulus equivalence has been discussed extensively during the last few decades amongst behaviour analysts and in many ways is a comparatively modern term in the field. It was, however, founded on principles that originated centuries ago. It was during the nineteenth century that the study of the human mind became established as an empirical science. The psychology which was born at this point in history was closely linked to, and as a result strongly influenced by, the major philosophical camps of the period. According to Miller (1962) these were:

(1) Positivism - the philosophy of August Comte which encompassed facts which could be determined definitively. Therefore, it was entirely knowledge based.

(2) Materialism - which deemed that the universe could be understood in the real terms of physical characteristics (for example, weight, length, time and so on). It is these laws of pure science (traditionally physics and chemistry) which underpin behaviour, and it is structure and function which should be analysed.

(3) Evolutionism - which viewed society as evolving gradually over time from earlier forms. It was most famously advocated by Charles Darwin.

(4) Empiricism - by which it was thought that the only source of knowledge was gathered via sensory experience.

It was the latter which supplied the most fertile ground for the emerging science of psychology as it provided: firstly, a method, stating that knowledge is best acquired by observation, experience and experimentation; and secondly a theory, emphasising experience as the foundation of all knowledge. This was in direct contrast with the nativist stance which trumpeted innate skills as being the most important to human learning (as characterised by the nature/nurture dichotomy).

An extreme form of empiricism was adopted by a group of philosophers known as the British Associationists, which was the direct predecessor of experimental psychology (Rachlin, 1991). As a reflection of this their fundamental beliefs were that all knowledge is accrued via the senses, yet they went further by stating that isolated sensations are insufficient to convey the essence or meaning of an object/event. For this reason, there is the need for something extra to gel direct sensations together; this something extra was seen to be *association*.

Various conditions were identified under which ideas came to be remembered. Contiguity was deemed to be the most essential of these and emerged as the basic factor for each associationist standpoint, meaning that if sensations occur together frequently, one alone can elicit the memory of the others (Rachlin, 1991, p.12). It is worth noting that many of these early definitions have been interpreted from a cognitive standpoint, therefore, for Behaviour Analysts the language may sit uneasily. However, it cannot be disputed that the fundamental principles could be of use to all branches of psychology. With the onset of the Age of Reason in the eighteenth century, associationism was put to extensive use not only in the field of philosophy and later psychology, but also in literature with the Romantic movement and even theories of economics (Leahey, 1992).

David Hume (1711-1776) categorised the human mind into perceptions which could be divided into impressions (sensations) and ideas (less vivid impressions). Using this dichotomy each environmental event (or sensation) activates ideas in the mind which are images of the stimulus itself or of other stimuli which have been previously associated with it. Simple and complex perceptions were distinguished between. Hume's doctrine of the association of ideas accounted for how complex perceptions are built up (see Hume, 1911). Expanding on notions put forward by John Locke (1632-1704 - see Locke, 1877) he set out three principles of associationism: similarity, contiguity, and cause. Causes are inductive (for instance, sun and sunglasses are linked because bright sunlight causes us to put on sunglasses) and arise through experience which allows us to adapt to our environment (Leahey, 1992).

Associationism as a psychological paradigm was introduced by a British physician, David Hartley (1705-1757) whose work paralleled Hume's but surprisingly was not influenced by it. He saw association as a basic cognitive operation which presented itself in two forms: (i) successive associations - meaning that sequences of ideas become attached to each other; and (ii) simultaneous associations - or contiguity, ideas which occur together, at the same time, become related (Hartley, 1749).

At the same time, James Mill (1773-1836), another associationist, created a totally mechanistic picture of the mind, describing ideas as following automatically on from each other. As Leahey (1992, p.45) relates, Mill, on the subject of abstract mental operations would theorise that, "Reasoning is no more than the associative compounding of the ideas contained in syllogisms." Such acts are seen as no more than automatised logic (see Mill, 1969). This is of importance because it reflects procedures and paradigms adopted by researchers into mediated transfer and stimulus equivalence hundreds of years later. Such associationists as Mill would suppose that if an idea A was associated with another idea B and similarly, if idea B was associated with idea C, then, this necessitated some kind of association between ideas A and C (Jenkins 1963). The nature of this association between A and C is what was to interest future researchers in Learning Theory. Mill continued to show his forward thinking when he accounted for the meaning of a word as being the sum of associated ideas recalled when the word is encountered. Rachlin (1991, p.14) interprets his theory as follows:

> In our lifetime, we experience chairs by seeing them, touching them, sitting on them, and so forth. All these activities in relation to chairs produce their own sensations and contain many sensations in common. The "sittableness" of a chair is an association of the visual experience of chairs with the kinesthetic sensation of sitting. Also, the simultaneous seeing and hearing of the word *chair* produces sensations that become associated with each other and with the sensations resulting from the sight and feel of chairs. These all mix together in a huge bundle so that when we hear the word *chair* the memories or ideas of all the other sensations come to our minds.
>
> (p.14)

If the last four words are substituted for, "are elicited" then these, "huge bundles" could be described as equivalence classes in behaviourist terminology.

It is hopefully becoming evident how the origins of Learning Theory can be traced back to the associationist philosophy. Schwartz (1989) says that the two influential aspects of associationism were: firstly that the mind is shaped by experience; and secondly that basic units combine to produce more complex ones. According to Neel (1971) associationism influenced researchers such as Thorndike and Pavlov in that they adopted the associationist idea that repeated presentations of stimuli in the environment leads them to become associated and to elicit each other. More explicitly, Pavlov at the beginning of the twentieth century renamed the association of ideas/sensations 'conditioning.' In the 1930s Thorndike deemed learning to be a series of connections but instead of associations between ideas he talked of relations between stimuli and responses (Miller, 1962). A portion of the discussions in Catania and Harnard (1988) is concerned with whether Skinner's analysis of behaviour has its roots in associationism. The consensus was that the influence of associationism is apparent, but Skinner's model takes its crude ideas much further.

A direct extension of the basic associationist philosophy came in the field of research dedicated to mediated associations (see Jenkins, 1963; Jenkins and Palermo, 1964). The scenario that if ideas A and B become associated, as do ideas B and C, then this suggests an association between ideas A and C, was labelled as a mediated association because it is mediated by a common idea, that being B. This differs from an immediate association which would only arise via a direct contiguity of ideas (Jenkins, 1963). Jenkins describes how mediated associations can be acquired more rapidly using paired-associate learning methodology. AC is learned faster if it is preceded by AB and BC training, rather than, for example, AB and DC.

If viewed from an historical perspective it can be seen how some of the fundamental principles applied to modern day Learning Theory can be traced back to the philosophical doctrines of associationism. However, although still fundamentally concerned with associations, Learning Theory (as would be expected over a span of hundreds of years) has evolved and modified its descriptions of behaviour. Without going into too much detail, Mill's premise works fine when applied to reflexes and other such behaviours encompassed by classical conditioning, yet they can not so easily be applied to operant conditioning. Behaviour analysts rarely use the term association, preferring instead the term 'relation'. Operant discriminative stimuli do not come to be associated simply through repeated presentations, nor do they automatically elicit behaviour. Operant behaviour arises from environmental contingencies of reinforcement or punishment.

The one specific realm of operant behaviour which will be discussed in detail is the paradigm of stimulus equivalence which shares many similarities with the principles of mediated association. Here associations between two sets of stimuli are generally not being described, as in stimulus equivalence, in the paired-associate research (as described by Jenkins, 1963; and Jenkins and Palermo 1964) subjects although relating one list of stimuli to another, do so via response terms. One member of the list is presented visually and this is paired to a verbal response from a second list of stimuli. Such responding could actually be seen as chaining when applied to several sets of stimuli (see Sidman, 1994, Chapt.4 for further details). Stimulus equivalence differs in that it is viewed in terms of stimulus-stimulus responding and when verbal responses are discussed (as in a verbal mediation account of equivalence) they serve as response terms which link the two stimuli in a relation, they are not explicitly trained as part of the relation.

To summarise, to take one area of psychology such as Behaviour Analysis and reflect on its historical roots (and it is important to note that only a small part of the history has been related here) is a fruitful occupation. It can lead to a clearer understanding of the field and to a greater appreciation of developments which have evolved to form modern day Behaviourism, and more specifically for this thesis, stimulus equivalence. Yet with the constant accumulation of knowledge, psychology is an ever changing science, generating new research hypotheses, and new behavioural principles. One such development is the emergence of the stimulus equivalence paradigm, the behaviour it generates, and the necessary emergence of new behavioural theories to accommodate it.

**Matching-to-sample procedures**

The term stimulus equivalence has been in existence for much of this century, but is applied nowadays purely in the sense of the behaviourist definition. The basic units of the paradigm are conditional relations, where one stimulus (a sample) is conditional upon choosing another (a comparison). Such relations are typically arbitrary (in that the stimuli bear no formal or physical resemblance to one another) and are taught using matching-to-sample (MTS) procedures.

More specifically, a sample is presented singularly to the subject (most commonly the procedure is automated and stimuli are presented using a computer) and the subject is required to acknowledge its appearance with an appropriate response (for instance, by touching the sample or pressing a key/button). This response results in the appearance of the comparisons, usually two or more, from which the subject must select. If the selection is correct, then appropriate reinforcement is delivered (an incorrect response typically leads to no reinforcement). Once the subject is able to match the correct stimuli (sample and comparison) together according to set criteria, it can be said that the conditional discrimination has been learnt.

**An auditory-visual task**

Figure 1 show two sets of stimuli. Set 1 are all features of the stimulus, flower, and Set 2 of the stimulus, heart.
The subject is first presented with the sample (A1), the auditory name, “flower” spoken by the Experimenter (or alternatively a prerecorded vocalisation), once the subject has

Figure 1: An illustration of two stimulus sets.
A stimuli = auditory names
B stimuli = printed words
C stimuli = pictures

| | Set 1 | | Set 2 | |
|---|---|---|---|---|
| A1 | "Flower" | A2 | "Heart" | Auditory Name |
| B1 | FLOWER | B2 | HEART | Printed Word |
| C1 |  | C2 |  | Picture |

responded in the appropriate manner two comparisons are presented (B1, the printed word, FLOWER; and B2, the printed word, HEART). The correct discrimination would be to select comparison B1. In the second instance, the sample is the auditory name, "heart" (A2) and in this case the correct comparison is B2 (the printed word, HEART). If the subject selects the wrong comparison no reinforcement is delivered (see top half of Figure 2). A second set of conditional discriminations could also be taught in the same way. This time the samples (A1 and A2) are matched to comparisons C1 and C2, the pictures of the stimuli (see bottom half of Figure 2). In this way two conditional discriminations have been taught for each stimulus set (see Figure 3).

Moving from the hypothetical to the empirical, such a procedure was first employed by Murray Sidman in 1971 using a severely retarded subject who was unable to read aloud, as a means of evaluating reading comprehension (Sidman, 1971). The problem he faced was how could understanding be assessed in such an individual?

He theorised that in order for a subject to match a written word to a picture (reading comprehension) this requires two skills: firstly, the conversion of the written word to an auditory name; and secondly the ability to match the auditory name to its visual referent. Therefore, in order to achieve reading comprehension, all that need be required is to match the picture and written word to its auditory name and reading comprehension should emerge without direct training and most importantly without requiring the subject to read aloud. So,

**Figure 2:** A diagram to depict two sets of trained conditional relations. A1-B1 and A2-B2 are seen on the top half: A1-C1 and A2-C2 are on the bottom half. Black arrows denote correct responses. Grey arrows denote incorrect responses.

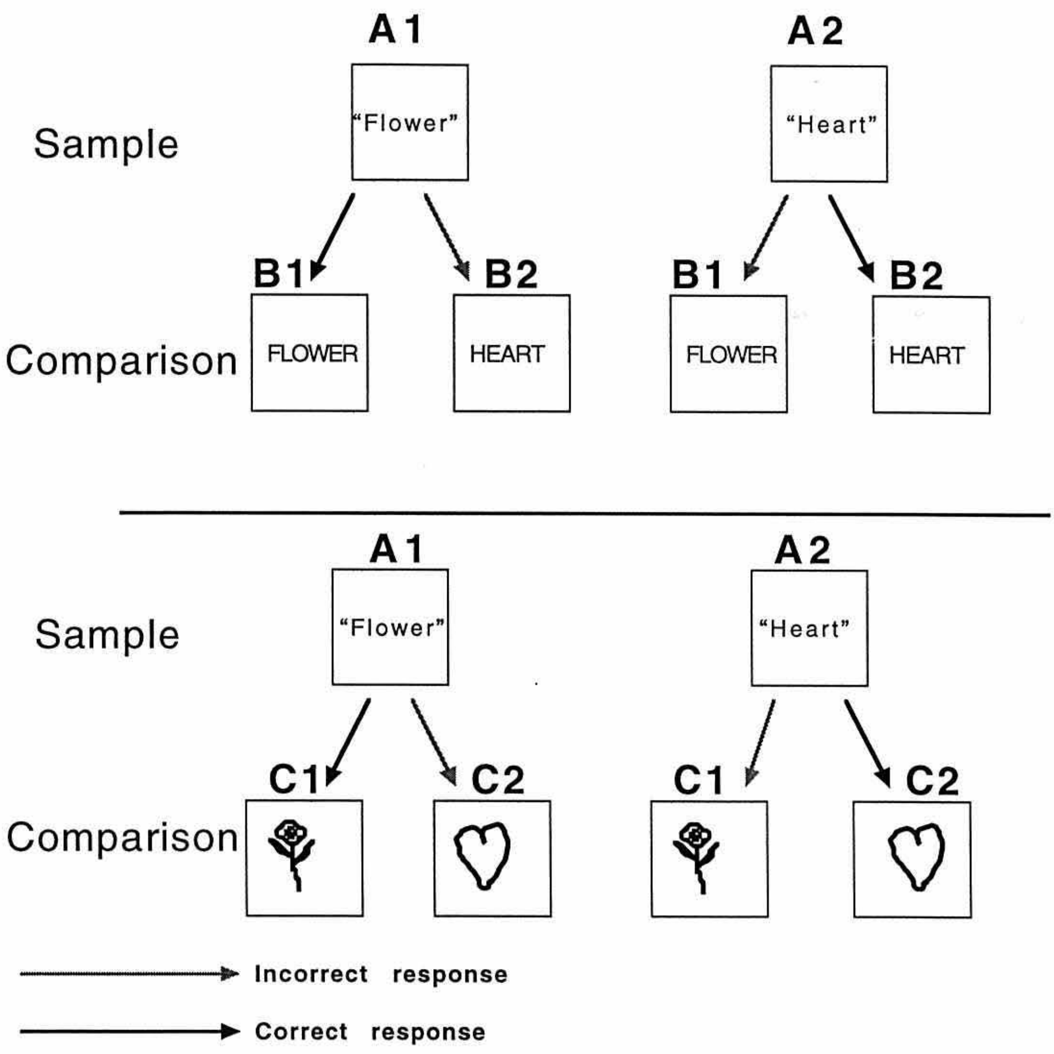

A1
A2
Sample
"Flower"
"Heart"
B1
B2
B1
B2
Comparison
FLOWER
HEART
FLOWER
HEART
A1
A2
Sample
"Flower"
"Heart"
C1
C2
C1
C2
Comparison
Incorrect response
Correct response

**Figure 3:** **The two conditional discriminations trained for each stimulus set.**

Train A ⟶ B

Auditory Name ⟶ Printed Word

Train A ⟶ C

Auditory Name ⟶ Picture

using matching-to-sample procedures Sidman established in his subjects the AB and AC relations illustrated in Figure 3 for twenty stimulus sets. The subject came into the experiment with the ability to match auditory names to pictures (AC) and to orally name pictures (CA). He was then trained to match auditory names to printed words (AB). To investigate whether reading comprehension was present the subject was tested on his ability to match printed words to pictures (BC), and to orally name the printed words (BA), without further training. The subject was found to be able to do so. In other words, visual reading comprehension and oral naming had emerged without direct teaching. These findings were verified further by a second, more experimentally rigorous study (Sidman and Cresson, 1973).

The paradigm was extended further. Sidman, Cresson and Willson-Morris (1974) demonstrated the emergence of equivalence in two institutionalised Down's Syndrome youths. They were taught forty conditional discriminations (AB and AC) and eighty derived (BC) relations emerged 'for free'. If this learning reliably emerges, then this suggests far-reaching implications for teaching methodology (see Sidman 1994, chapt.3, for fuller discussion).

What can be gleaned from this paradigm? In the Sidman studies subjects with severe learning difficulties, after being appropriately trained to match sets of stimuli in a certain way (AB and AC), when tested are able to exhibit many more derived relations without being explicitly taught to do so. Further, they have shown the ability to behave symbolically (that is demonstrate reading comprehension and orally name printed words) for stimuli which had previously been outside their comprehension.

If this paradigm does act as a measure of symbolic activity then it is potentially a valuable tool for behaviour analysts. It has been demonstrated that verbal behaviour is an important source of control in human performances on operant schedules of reinforcement and can differentiate human and nonhuman patterns of responding (Harzem, Lowe and Bagshaw, 1978; Lowe, Beasty and Bentall, 1983; Bentall, Lowe and Beasty, 1985; and Bentall and Lowe, 1987). Skinner (1957) accounted for verbal behaviour within the framework of the three-term contingency model. Cognitive theorists, such as Chomsky, have criticised this approach for failing to account for two important aspects of symbolic behaviour: (i) meaning; (ii) novel utterances which appear with no apparent source of reinforcement (Wulfert and Hayes, 1988). There seems to be some overlap between behaviour exhibited by subjects on the matching-to-sample task outlined above and these two aspects of symbolic behaviour suggesting that the stimulus equivalence paradigm may be valuable in bridging the gap between traditional behaviourist theories and accounting for symbolic activity. To examine this a closer look at what is actually occurring when the subjects learn the critical relations is needed.

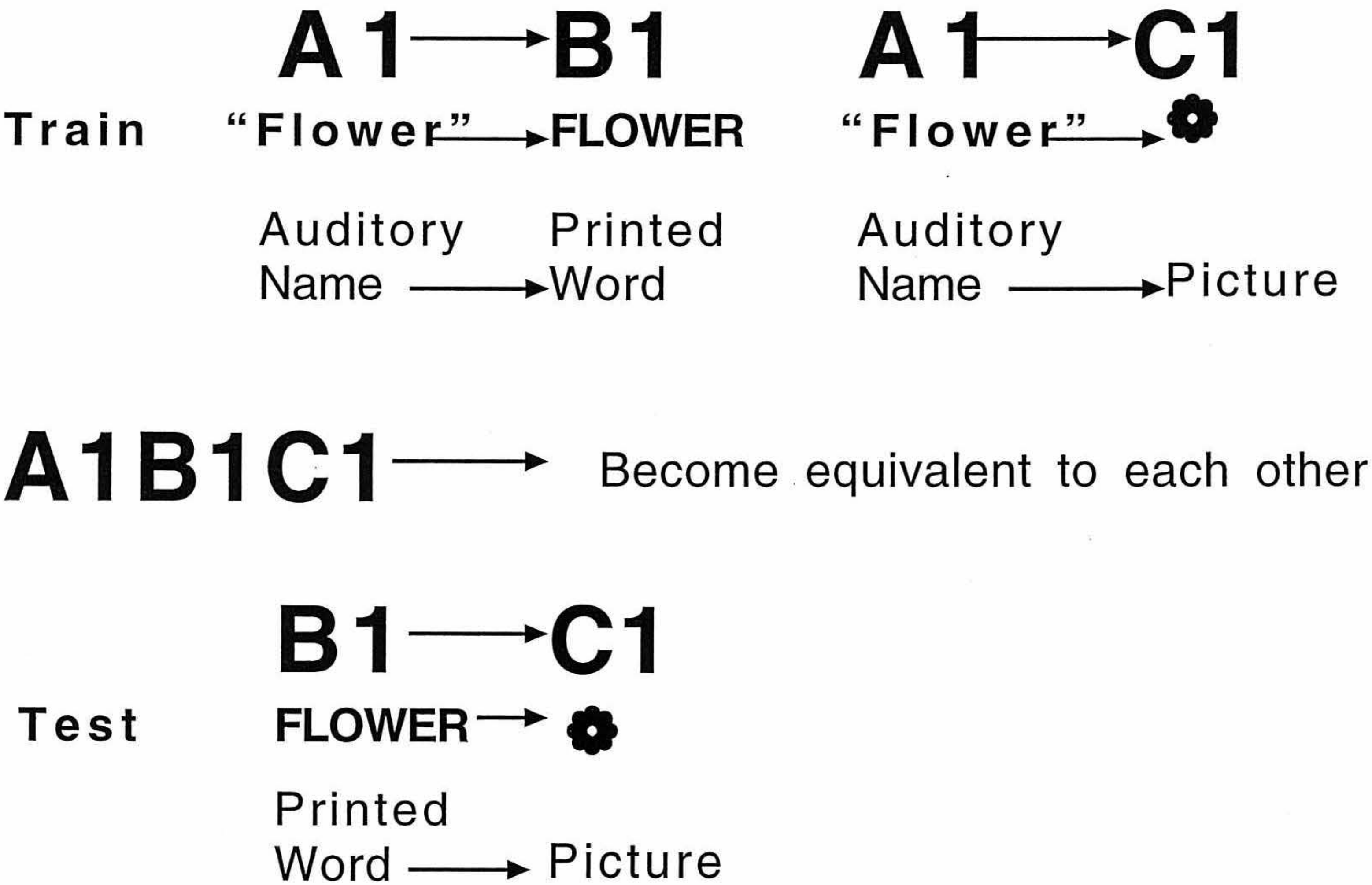
Returning to the hypothetical example, a nonhuman subject (such as a pigeon, rat or monkey) could be taught, using appropriate learning techniques to respond correctly to the initial conditional discriminations (AB and AC). Yet if the same arbitrary behaviour (for instance, selecting the printed word FLOWER, B1, on hearing the dictated name, "flower", A1, and selecting the picture of a flower, C1, when presented with the printed word, B1) is observed in a human subject it would be inferred that the individual is acting symbolically. In other words, for them, "flower" (auditory name) means, or acts as a symbol, or represents, FLOWER (printed word) or picture. This logic indicates that the same should be said of the nonhuman's behaviour, if relating the stimuli in this way was indeed symbolic. Yet looking at these trained relations alone does not mean that the inference is correct even in human subjects. The subjects are merely relating the stimuli conditionally according to the reinforcement contingencies, in the same way as a dog playing dead in response to a command in return for a reward or learning not to jump up because this will lead in a smack.

So, what is it that makes this behaviour symbolic? Symbolic behaviour is defined by substitutionality or put another way, certain features of a stimulus represent it in various contexts. For example, the auditory name, "flower" is a substitute for the printed word, FLOWER when reading, or the picture flower when naming/speaking; they all represent the same specific stimulus, flower. What is crucial is that all the features of the flower are equivalent to each other. They all (the auditory name, printed word and picture) form a class of equivalent stimuli which represent, flower. Only if this equivalence class is formed will the training of certain relations lead to the emergence of untrained relations. For example, if train A1B1 and A1C1, B1C1 will be derived because A1B1C1 all form a class of equivalent stimuli (see Figure 4).

Sidman and colleagues (Sidman, Rauzin, Alzar, Cunningham, Tailby and Carrigan, 1982; Sidman and Tailby, 1982) laid out a specific criterion for stimulus equivalence based on the mathematical definition of equivalence from set theory. The conditional relations AB and AC should also be equivalence relations if they demonstrate the properties of:
(i) Reflexivity, the relation a stimulus holds with itself, for example, if A=B then A=A; (ii) Symmetry, if A=B then B=A; and Transitivity, if A=B and A=C then B=C (or an alternative protocol, if A=B and B=C then A=C). Finally there is an overall test for equivalence which combines the properties of symmetry and transitivity, CA (if B=A and C=B then C=A, Sidman and Tailby 1982). As well as a conditional relation possessing the above properties, the other key factor of symbolic behaviour is that if there is equivalence between stimuli these properties should emerge without direct training.

Taking symmetry as an example: if A1 then select B1, if "flower" (auditory name) select FLOWER (printed word); the symmetrical version is, if B1 select A1, if FLOWER (printed word) select "flower" (auditory name). If these two relations have to be explicitly taught then they could function as two separate unidirectional relations, one does not reflect a property of the other. If the subject is acting symbolically then AB is directly trained and BA should emerge without direct reinforcement (because A is equivalent to B). So, what

Figure 4: An illustration of how stimulus equivalence can be derived from one stimulus set.


A1 → B1
A1 → C1
Train
"Flower" → FLOWER
"Flower" →
Auditory Name → Printed Word
Auditory Name → Picture
A1B1C1 → Become equivalent to each other
B1 → C1
Test
FLOWER →
Printed Word → Picture

is seen here is only a single relation (AB) which exhibits the property of symmetry (as measured by BA). Similarly, if A is equivalent to B, and A is also equivalent to C, then BC (transitivity) and CB (equivalence) should emerge without reinforcement because B is also equivalent to C.

Returning to Sidman (1971) the subject displayed reading comprehension (BC) and oral naming of the printed words (BA) without further training because the original taught relations were equivalent. Or put another way the stimuli in each set displayed stimulus equivalence. So it can be seen that Sidman's definition provides a means of empirically assessing whether conditional relations are also equivalence relations.

It must be noted that the Sidman interpretation of mathematical equivalence is somewhat limited, as Saunders and Green (1992) point out. The mathematical definition also views behaviour controlled by, for example, oddity/S-control (where it is not the reinforced comparison which is controlling behaviour but the unreinforced comparison, if A1 then do not select B2, leads to the correct response B1, but it is B2 which is directing the response) as equivalence; and also considers the possibility of one large equivalence class developing which would be characterised by erroneous responding according to the experimental contingencies but does not strictly speaking exclude the possibility of equivalence between the stimuli. Saunders and Green also cast doubt over the value of the reflexivity test as it is at odds with the nature of the equivalence relation. In other words, in arbitrary equivalence relations, matching does not occur on the basis of physical similarity or because each member of each equivalent class is identical to each other. The example the authors cite is in the instance of transitivity, "...the basis for similarity between A1 and B1...may be unrelated to the basis for similarity perceived between B1 and C1.." (p. 236). The authors conclude that caution should be exercise because conditional relations can be more complex than the Sidman definition implies.

The question of what is stimulus equivalence and how it can be measured has been addressed. The remaining issue is how does it come about? Given that it has already been stated that this new symbolic behaviour extends beyond the remit of traditional accounts. In other words, how does the training AB and AC lead to the formation of an equivalence class? Or how do properties of equivalence emerge? Several researchers have attempted to address this question.

One way of interpreting this is in terms of the nature of the relationships between the stimuli. This is the approach Sidman has taken. Briefly, he sees equivalence as a "fundamental stimulus function" (Sidman 1994, p.111). In other words, a 'given' or a basic behavioural unit which cannot be broken down into anything smaller. In his opinion a child is born into a world were everything is equivalent to everything else. Reinforcement contingencies or context break down the environment into equivalence classes or sizable, meaningful chunks. Sidman (1994) expands upon this by suggesting that responses and reinforcers too can become members of equivalence classes.

Hayes talks about equivalence in terms of stimulus-stimulus relations when outlining his Relational Frame Theory (Hayes 1991; Steele and Hayes 1991; Hayes 1992; Hayes and Hayes 1992). What he describes as Arbitrarily Applicable Relational Responding (that is the type of responding seen as on a MTS task; matching stimuli) is characterised by the properties of stimulus relations (those being mutual entailment; combinatorial entailment; relational reflexivity/irreflexivity; and transfer of function) which is similar to the Sidman definition. Hayes, however, says that this type of responding then leads to the formation of relational frames (in the case of equivalence relations this would be a frame of sameness/coordination) but other frames exist in the subject's behavioural repertoire such as opposition, distinction and comparison (see Hayes, 1991 for examples of such frames) which serve to operate on behaviour. In this way, stimulus equivalence is a result of higher order responding (the activation of the frame of coordination) which arises from the subject's behavioural history of directly trained exemplars and not a fundamental feature of the stimuli as Sidman stipulates.

Boelens (1994) has presented an explanation of stimulus equivalence in terms of the three-term contingency model. In this way, equivalence can be explained as examples of generalised performances given the appropriate training histories. Trained examples (in the everyday human environment) lead to generalisation to novel samples and comparisons. The behaviour observed during an equivalence test is thus under the discriminative control of earlier responding.

Another way of looking at the paradigm of stimulus equivalence is in terms of <u>responses</u>. What is meant by a response is not the selection response (that is to choose the appropriate stimulus) which is always the same (for instance, touching the screen or pressing a key) and is merely a component of the stimulus-stimulus relation; but a topographically distinct response which can mediate between the stimuli and in this way facilitate the emergence of equivalence.

Returning to the original auditory-visual task (see Figure 1), the auditory name (produced by the subject in imitation of the sample word) could be argued to be mediating between the stimuli. For instance A1B1C1 all form an equivalence class because they all share the same name, "flower" and similarly A2B2C2 are matched together because they share the same name, "heart". It is this mediating response which leads to the emergence of equivalence between the stimuli. According to this view it is the mediating response/verbal behaviour which facilitates the successful demonstration of equivalence and not the other way around (as stipulated by Sidman and Hayes).

It must be noted that the first possibility Sidman and colleagues considered was that behaviour on a matching-to-sample task is mediated by a common/shared response. Sidman, Cresson and Willson-Morris (1974) investigated this using two institutionalised Down's syndrome youths as subjects. However, the authors concluded from the resulting data that the observed emergent behaviour could not have been mediated by the subject's common naming responses because the subjects were incapable of producing reliable names for the stimuli on a post-experimental naming test (see Dugdale 1988 for how this conclusion can be demonstrated to be not entirely justifiable). This theory, was advocated by Sidman and colleagues for the next decade (Sidman and Tailby 1982; Sidman, Kirk and Willson-Morris 1985; Sidman, Willson-Morris and Kirk 1986).

In his later discussions Sidman does not dismiss the possibility that naming (in one form or another) may facilitate performance on the equivalence test. For example, Bush, Sidman and deRose (1989) talk of subjects implementing verbal rules in order to demonstrate equivalence. However, in common with Hayes he maintains that naming is merely facilitating the subjects' performances and is not necessary for it to emerge. Hayes (1991) says that in the context of a stimulus equivalence task, naming (which is an example of arbitrarily applicable responding) merely signals the appropriate frame of behaviour, the frame of coordination, and is therefore, in itself not necessary for equivalence.

The role of naming in the demonstration of stimulus equivalence has been recently debated with the publication by Horne and Lowe (1996) of their Naming Hypothesis which examines the origins of verbal behaviour and the resulting implications for stimulus equivalence. According to this hypothesis naming is necessary for the demonstration of equivalence and it is the properties of the name relations (responses) which promotes stimulus equivalence.

The origins of this theory can be found in Dugdale and Lowe (1990) where the inference is first drawn that it is a subject's ability to name the experimental stimuli, and to use language to relate the sample and comparison stimuli which leads to the successful demonstration of equivalence. When such verbal behaviour is absent they claim that the subject must fail the test of equivalence. They address the question of why should such verbal behaviour be of importance in the demonstration of equivalence by making a distinction between subjects merely labelling the stimuli and subjects actually naming the stimuli.

It is of worth here to define what is meant by the term naming in this particular context in contrast to labelling or tacting. Labelling or tacting involves unidirectional responding usually in the form of: "see object, say name of object". Dugdale and Lowe define naming as follows, "..naming is itself a symbolic skill that involves bidirectionality." (p.132). By this they mean between an arbitrary stimulus and verbal response. The keyword to their account is <u>bidirectionality</u>. It is detailed how a naming relation comprises of two components which are linked symmetrically. These are:
(i) the arbitrary stimulus elicits the speaker's verbal response (see /a/ leads to say "a").
(ii) the verbal response evokes behaviour in the speaker/listener (say "a" leads to select /a/).
The critical feature of this naming relation is that one component emerges unreinforced once the other has been

trained. Then it can be said that the subject is naming which can be described as stimulus-response symmetry which is qualitatively different to other verbal behaviours such as tacting or labelling. These may on the surface appear to look the same as naming but they lack the critical derived symmetrical element which defines symbolic behaviour.

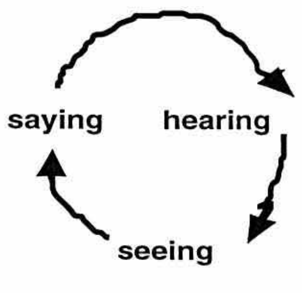
The flaw in Dugdale and Lowe's argument is that it is essentially circular. Naming, (as described above) which is a symmetrical relation, facilitates the emergence of equivalence relations which are also symmetrical by definition (Sidman et al, 1982; Sidman and Tailby, 1982). To overcome this the authors differentiate between the two kinds of symmetry which are occurring :

(i) naming, which is stimulus-response symmetry, and involves topographically distinct responses (seeing a specific stimulus - saying its specific name).
(ii) the symmetry between two visual stimuli in an equivalence context, which is stimulus-stimulus symmetry and incurs stimulus selection based responses (pointing or pressing a button) in other words, the same response to every stimulus. In these terms the former type of symmetry gives rise to the latter. The question still remains of how the first type of symmetry (naming) comes about?

Dugdale and Lowe hypothesise that it emerges through the route of normal language development with trained exemplars of stimulus-response symmetry. Given enough experience, one component of the naming relation will then emerge in the absence of direct reinforcement. Dugdale and Lowe state, "Naming would thus involve both language production and comprehension, and it would require the subject to function both as a speaker and listener" (p.133).

Horne and Lowe (1996) take these basic principles and produce a much more elaborate Naming Hypothesis which spans the fields of both Developmental Psychology as well as Behaviour Analysis. Their initial task (as eluded to in the Dugdale and Lowe paper) is to establish the origins of naming, the second (and most important for the purposes of this thesis) is how the Naming Hypothesis relates to stimulus equivalence.

In order to appreciate the latter it is necessary to briefly summarise the former account. Horne and Lowe (1996) describe naming as a circular relation. Like Dugdale and Lowe (1990) they attribute importance to both speaker and listener behaviours, but for them this entails stretching what was once a two-term model (stimulus-response symmetry) to accommodate a third term. The end result is a relation as follows:




The action of naming is depicted as an individual seeing an object; saying a verbal label; and hearing their own utterance which prompts the appropriate listener behaviour and then reengaging with the object.
The components of this act are hypothesised to evolve as follows. The first essential unit is the evolution of listener

behaviour which is shaped and reinforced by the verbal community surrounding the child. Ultimately this involves responding differently to different items, for instance, by orienting, pointing, fetching, dropping and so on.

The second crucial process in the name relation is the link between a child's listener and speaker behaviour. Critical to this is the echoic response. The echoic has been of fundamental importance to the Skinnerian account of language development but in Horne and Lowe's account it takes on added significance in that a caregiver's utterance may not only occasion an echoic response in the child but also the previously acquired listener behaviour for that particular object. They talk of the echoic as being a functional class consisting of: the caregiver's utterance; the child's echo; the child's listener behaviour; all of which is reinforced and cemented in the child's linguistic repertoire. However, this echoic behaviour is not of course naming it is not a response to the stimulus itself but to the caregiver's utterance. It does, however, act as a prerequisite for more complex behaviour.

The third element of the name relation is tacting which is verbal behaviour evoked by a specific stimulus in the environment. For example, the caregiver points to a doll and says, "doll"; the child hears "doll", looks at the doll (listener behaviour) and says to herself "doll"; the sight of the doll becomes the discriminative stimulus for the response, "doll". Eventually, the child will be able to produce the utterance herself with or without the presence of the discriminative stimulus.

This name relation, according to Horne and Lowe (1996) is, "...a qualitatively new bidirectional relation in the child's behavioural repertoire" (p.200) and is also a , "higher-order behavioural relation" (p.203) which can be elicited once a few examples have been directly established via the primary route (as described above). The authors go on outline how once a child has acquired the ability to form name relations this powerfully changes their behaviour. For example, arbitrary stimuli come to share properties because they have a common name. Similarly, a couple of reinforced exemplars leads to generalisation to novel members of a stimulus class because they share a common name. In addition, stimuli occasion a name which in turn, because of the bidirectionality of the relation, evokes previously established listener behaviour.

The second task for Horne and Lowe's naming hypothesis is to account for the emergence of stimulus equivalence. In auditory-visual matching (the task described earlier) the stimuli classes are given a common name (the A stimuli; the auditory name) and are reinforced for matching this name to other stimuli (AB and AC training). If a common name is applied to all the stimuli in each class this should promote equivalence/substitutionality between members of the class. In that if a subject is assigning a common name to the stimuli this evokes the appropriate listener behaviour which in the matching-to-sample context is responding appropriately (that is, see sample A1 and say "flower", see comparison B1 say "flower" and select key B1). In other words, the listener behaviour is selecting the other members of the class whenever any one member appears as the sample. If this occurs then the criterion of stimulus equivalence will be met.

## A visual-visual task

A second commonly used matching-to-sample task involves stimulus sets comprising of abstract visual stimuli. Sidman, Willson-Morris and Kirk (1986) moved on to this methodology in an attempt to investigate the role of naming in the emergence of equivalence. Subjects' performances were compared on an auditory-visual task and on a visual-visual task. In the visual-visual task all but one subject failed to consistently name the stimuli leading to the conclusion that common naming was not necessary for the emergence of equivalence (see Dugdale, 1988 for a critique of this conclusion).

The visual-visual task does mean that the stimuli are less likely to be named because there is no obvious common name. However, the Naming Hypothesis still predicts that if names are assigned to the stimuli, stimulus equivalence must emerge. So, how could naming promote the emergence of stimulus equivalence between visual stimuli? Horne and Lowe (1996) argue that the most probable way by which common naming would occur in this task is through the subject assigning a name to a stimulus prompted by its physical resemblance to some real-life object and this same feature is then sought in the other members of the class (which has been established through the contingencies of reinforcement). The relations may take longer to establish because the subject could be amending the strategy throughout the training session as the stimuli are matched appropriately, but it is likely to occur because as

Horne and Lowe (p.216) say, "The subject's naming of the stimuli thus transforms the task from 'arbitrary' to 'non-arbitrary' match-to-sample."

Another form of naming which may occur specifically in visual-visual matching is intraverbal naming. Horne and Lowe describe such naming as follows, "...a relation where, like the echoic, a verbal stimulus is the occasion on which a particular verbal response receives generalised reinforcement" (p.209). This type of naming can also encompass both speaker and listener behaviour and is not necessarily unidirectional. For example, a child on hearing/saying "moo" may intraverbally produce "cow". With self-repetition "moo-cow" becomes "moo-cow-moo-cow" and so on. Here is the bidirectionality, in that if either component is produced the other is emitted along with the appropriate listener behaviour.

Intraverbal naming could occur, according to the Naming Hypothesis, when a subject assigns individual names to each stimulus. Through the contingencies of reinforcement the subject will learn to match up pairs of stimuli (for example, A1B1) and the subject will be regularly saying the names together (for instance, A1 "square" and B1 "cross") covertly as a self-echoic. So an intraverbal forms, "square-cross". Repetitions over numerous trials of the sequence, "square-cross-square-cross" will lead to the relation becoming bidirectional (square leads to cross but also cross leads to square). At the same time a bidirectional relation will be established for the appropriate listener behaviour (that is responding in the required manner). This means that when confronted with B1A1, "Cross-square" this also produces the appropriate listener behaviour, that is responding to the "square" key.

During the course of training other stimulus pairs will be related in this manner (for example, B1C1, "cross-hand"). This results in a chain of intraverbal relation, A1B1C1, "square leads to cross leads to hand". With self-repetition of this sequence any one of these stimuli will produce the other two resulting in an intraverbal equivalence class which will lead to successful responding on the formal equivalence test.
A slight variation of this which might occur is that the subject may describe the relation between the stimuli. For example,"square goes with cross" (A1B1) or "cross goes with hand" (B1C1). The names are not bound intraverbally but by meaningful propositions which will function in the same manner as intraverbals to occasion equivalence.

Horne and Lowe also go on to state that merely naming the stimuli is not always sufficient in bringing about equivalence, "...to ensure success, it is not sufficient for subjects to have naming skills...but more specifically, their stimulus class naming must be congruent with the experimenter-defined classes" (p.226). So, subjects' naming must be consistent and grouped according to the experimental criterion for equivalence to emerge

It should be noted that there is a certain amount of ambiguity attached to defining naming in the sense that naming can mean many things to many people. As Levelt (1989) points out the same object may be referred to differently depending on the given context and depending on how much information the listener needs to know. What is crucial on a matching-to-sample trial is that the name is consistent and topographically distinct (see Michael, 1985).

To summarise it has been shown how the occurrence of stimulus equivalence can be viewed in terms of stimulus-response relations. This is by no means the only way of characterising the emergent properties of equivalence (it has already been noted how Sidman and Hayes view stimulus equivalence in terms of stimulus-stimulus relations) but the major strength of this hypothesis over the other accounts is that it can essentially be falsified and this is why it it the hypothesis this thesis is to test.

For example, what evidence could be gathered to support Sidman's view that equivalence is a 'given' from birth and that a combination of reinforcement contingencies, context and history break down the environment? The point at which a child begins to demonstrate equivalence could be measured using a longitudinal experimental design, but what can be said? How can it be shown that the child is breaking down their environment and not building it up (as predicted by Hayes via relational framing or the Naming Hypothesis via verbal behaviour)? Another problem is that Sidman's theory is non-specific. He states that the environment is broken down by context, but does not elaborate as to the actual context which would be sufficient to do this. Hayes' theory states that relational frames emerge from a history of directly trained exemplars but the only example he relates is for the frame of coordination. There is no specification as to how the frames of opposition, distinction, comparison and so on would arise. On the other hand Horne and Lowe set out three unequivocal means by which their theory could be falsified.

**Empirically testing the Naming Hypothesis**

The Naming Hypothesis predicts that if a subject assigns names to the stimuli (as described above) they will successfully demonstrate equivalence. If a subject does not assign names or does not assign consistent names, the equivalence will fail to emerge. Horne and Lowe set out three definitive tests of their predictions:

(i) If naming gives rise to stimulus equivalence then nonhuman subjects should fail standard tests of equivalence.

(ii) Nonverbal humans (that is preverbal infants who have not yet learned to name; or language disabled individuals who possess no functional speech should fail to successfully demonstrate equivalence too.

(iii) If a population of subjects who initially fail an equivalence test are then taught to assign names to the experimental stimuli this should result in equivalence relations emerging if naming is the key determinant to success.

Evidence is cited in support of theory on each of these three points.

(i) Nonhuman Evidence. This is the most testable of the three predictions. It appears difficult to say unequivocally that it is a human subjects' naming which promotes equivalence in that: firstly it often occurs covertly and as a result is difficult to measure; and secondly many other cognitive processes are occurring any one of which could be influencing the emergence of equivalence. In contrast it is easy to disprove the theory. If nonhumans show evidence of stimulus equivalence then something other than name relations must be controlling the subject's behaviour.

Horne and Lowe state that so far no convincing evidence of the emergence of stimulus equivalence has been demonstrated using animal subjects. The one exception they single out is the Schusterman and Kastak (1993) study with the sealion, Rio. For 12 three-member stimulus classes the sealion was trained on AB and BC baseline relations and the corresponding symmetry, transitivity and equivalence relations. On 18 three-member class the sealion just received baseline training and equivalence was tested for (with reinforcement). On 16/18 first test trials Rio selected the correct response. Schusterman and Kustak attribute this success to Rio having been taught a sufficient number of examples which demonstrated to her that the sample and comparison stimuli were interchangeable resulting in the formation of the critical bidirectional relations needed to pass the equivalence test. Horne and Lowe discuss the possibility that because of some procedural abnormalities the sealion was treating each stimulus pair as a compound (that is not discriminating between the individual elements) and through reinforcement on the non test pairs, has learnt that regardless of the location of the stimuli to respond to the 'outer' element (the comparison) as no response to the sample stimulus was required. This means the subject would respond correctly on the symmetry trials without further training in that she sees it in effect as an AB trial but knows that she has to respond to the outer element in order to receive reinforcement and therefore, responds correctly. If the AC transitivity trials are also seen as compound stimuli this should occur on the CA equivalence trials too (given that "associative transitivity" can be demonstrated as arising from the training alone and not due to equivalence). In this way they suggest the sealion's performance may be due to some simpler form of responding rather than equivalence.

The Schusterman and Kastak study provides the best evidence so far regarding the emergence of stimulus equivalence in nonhuman subjects. Yet it is by no means conclusive for the reasons expoused by Horne and Lowe and quite simply due to the fact that the procedure undertaken was entirely nonautomated, so the possibility of experimenter cueing (no matter how subtle) can never be ruled out entirely.

Deviating for a moment from the Horne and Lowe rationale, other studies since the sealion experiment have presented evidence of the possible demonstration of equivalence in nonhuman subjects. For example, Zentall and Urcuioli (1993) describe how pigeons (albeit different samples of pigeons) have demonstrated reflexivity, symmetry and transitivity using procedures where food is incorporated as a member of the stimulus class. Similarly, Manabe, Kawashima and Staddon (1995) claim to have demonstrated emergent relations (in line with the Sidman definition of equivalence) after training certain conditioned discriminations in budgerigars. In addition, they claim that their subjects produced, ' name' responses and this may have facilitated the emergence of the new relations. Finally, Yamamoto and Asano (1995) report the emergence of symmetry and transitivity in a chimpanzee who had previously received extensive language training using matching-to-sample protocols.

However, this evidence (like the Schusterman and Kastak sealion study) is far from being unequivocal. Generally speaking, all the procedures described deviate at some level from the standard MTS tasks employed with human subjects. For example, in the Zentall and Urcuioli (1993) study reinforcement continues to be given to the subjects throughout the test session. Such a tactic would be admissable if the subject's percentage of correct responding was significantly above chance from the outset of the test session. This, however, is not the case in the data presented. In fact, much of the performances cited are at no more than chance level. If any improvement in performance is seen it seems to be the product of responding on a large number of test trials, with reinforcement and therefore, is not strictly speaking, 'emergent'.

A further insight into the pigeon's performances on the MTS tasks was provided by the authors using a different tactic by which they compared different training schedules and using consistent and inconsistent stimulus pairs culminating in a test for the transfer of learning. Urcuioli, Zentall and DeMarse (1995) compared a many-to-one (MTO) procedure (where a variety of sample stimuli are related to one comparison) to a one-to-many (OTM) where one sample is related to many comparison stimuli. The general conclusion was that transfer of learning was seen using the MTO procedure but no evidence was found using the OTM schedules. From this evidence it can be seen that whether the pigeons passed the test is dependent on how the baseline relations are trained. What is crucial is that such order effects are not seen in human subjects suggesting that the behaviour observed in pigeons may have been brought about by something other than stimulus equivalence.

In the Manabe et al study the emergence of equivalence was again tested for in the presence of reinforcement. Correct performance on the test trials also seems to emerge gradually over a number of trials which suggests that the behaviour the budgies were exhibiting is directly learned behaviour. Again such responding is not directly comparable to the stimulus equivalence demonstrated in human subjects which emerges in the absence of reinforcement and relatively early on in the test session.

Finally, the evidence presented by Yamamoto and Asano is also flawed. Firstly, the procedure was nonautomated and thus does not have adequate controls against experimenter cueing. Secondly, the effects reported were small and by no means consistently observed over the test sessions. The best evidence presented is seen with regard to transitivity. The properties of symmetry and equivalence were less reliably observed, in fact they usually only reached chance level.

Crucially, in these studies, there is no instance where all the properties of stimulus equivalence (as outlined by Sidman and Tailby, 1982) are evident. In Schusterman and Kastak's study, Rio initially failed the symmetry test. Zentall and Urcuioli (1993) do not actually report any evidence of symmetrical responding in their pigeons. Similarly, little evidence of symmetry was seen in the Yamamoto and Asano study. As seen already in the

Manabe et al study it can be argued that none of the properties of equivalence exhibited are actually emergent. What is common here is a lack of evidence for the emergence of symmetrical responding in nonhumans.

Bentall and Dickens (1994) argue that being able to respond symmetrically is the key difference between human and nonhuman behaviour. In that symmetry is necessary for effective language because verbal relations are symmetrical. So, it could be concluded that the key test of equivalence is symmetry and to date there has been no convincing nonhuman evidence of its occurence,

Such criticisms are bound to be put forward at this stage as there are few reliable replications to add weight to these findings. However, even if these data are more reliably replicated or procedures more stringently applied and stimulus equivalence is demonstrated in nonhuman performances, this according to Horne and Lowe would not falsify their Naming Hypothesis. In their paper and in their reply to the subsequent commentaries (Lowe and Horne 1996, same volume) they talk about the "formalistic fallacy" (Skinner 1969) which warns that findings from animal data should not be applied wholesale to human subjects as the controlling behaviour may be different in both cases. They suggest two possible routes by which stimulus equivalence could manifest itself. The first is the verbal route whereby naming is necessary for equivalence to emerge and this is the route taken by the majority of verbal humans. The second could entail nonverbal behaviour which is contingency-governed and could explain the animal data. If this holds then there is no challenge to the Naming Hypothesis by any nonhuman findings.

Horne and Lowe emphasise that there is far more convincing evidence for the verbal route of equivalence. Which can be seen by considering the latter two tests of the theory.

(ii) If naming is necessary for equivalence then nonverbal humans should fail to demonstrate equivalence.

One study which is most-widely cited to support this assumption is Devany, Hayes and Nelson (1986) who showed that a group of retarded children who possessed no spontaneous language failed to demonstrate equivalence, whereas the other two groups (a group of normally developing preschool children and a group of retarded children who had the ability to produce spontaneous speech/signs) did. Although the data is correlational it could lend support to the view that the ability to use language and success on tests of equivalence are related.

Barnes, McCullagh and Keenan (1990) studied children with and without a hearing impairment on a standard matching-to-sample task. The three subject groups were made up of: (i) normally developing preschool children aged 3-4 years; (ii) normally developing partially hearing children with a verbal age of two years and above; and (iii) normally developing partially hearing children with verbal ages of two years and below. All subjects in groups one and two successfully demonstrated equivalence, but only one subject from group three passed the equivalence test. This again is correlational yet it infers a relationship between verbal ability and the successful demonstration of equivalence.

It must be noted that this evidence is not unequivocal. Dugdale (1988) pinpoints some procedural inadequacies which suggest that there could be alternative explanations behind the subject's performances other than stimulus equivalence in the Devany et al (1986) study. The most salient of these are as follows. The sessions with the normally developing children took place with both the experimenter and subject sat together on a rug with the stimuli spread in front of the subject or held by the experimenter. This opens up the possibility of experimenter cuing, of a kind which could not have been present with the retarded subject groups who sat rather more formally at a table to take part in the experimental sessions. The retarded subject group with no language were at a major disadvantage in comparison to the other groups because all the experimental instructions were verbal and thus were potentially easier to comprehend by the language-able groups. As a result it can not be said for sure that the language-disabled group's poorer performance on the equivalence task was due solely to their inability to employ verbal strategies or because they were unable to comprehend the task requirements as well as the other two groups due to their lack of verbal experience.

Similarly, the stimuli employed were line drawings of animal-like figures in different coloured ink. This could possibly encourage differential naming and as a result enhance learning in the language-able groups based on the animal or colour names, but not in the language-disabled group.

Finally, Dugdale (1988) scrutinises the actual data and observes that the retarded/no language group took the longest to learn the baseline tasks and therefore, as a result it should not be surprising that they needed longer to produce correct responses to the test-trials. The sample may have consisted of general 'slow learners' which had little to do with their verbal abilities.

In the Barnes et al (1990) study the groups with normally developing verbal ability successfully demonstrated stimulus equivalence, whereas only one subject in the non-verbal partially hearing group did the same. Subjects' performances cannot be attributed to lack of intellectual capacity as all the subject were non-retarded (unlike in the Devany et al 1986 study). However, it must be noted that very small numbers of subjects were employed; only two subjects per group.

Therefore, in the low ability group one subject passed and one subject failed (which is what would be expected to occur by chance). The subject who failed (Claudia) had the lowest verbal age of all the subjects and on her first test session scored 0% which suggests that she was responding consistently despite the fact that it was the incorrect comparison which was controlling her behaviour (S-minus control). So it is possible that she was demonstrating equivalence between the incorrect responses. However, in the second test session she performed at chance level, which led the author to conclude that it was unlikely that she had formed erroneous equivalence classes in her first test session. This may seem a reasonable conclusion, yet on the other hand, Claudia may have just abandoned her first strategy when it came to the subsequent session especially as she was receiving no reinforcement for her responding. Even if it can be argued that Claudia definitely did not demonstrate any evidence of stimulus equivalence, she is the only subject not to do so. The assertion would be stronger if there were more evidence to base it on.

So, to summarise, there is evidence which demonstrates that subjects with little or no verbal ability fail to demonstrate equivalence where comparable subjects with normally developing language do. However, firstly, this evidence is only correlational, the two phenomena may merely co-occur. From this evidence alone it can not be said definitively that one causes the other. Secondly, there are procedural issues which may suggest that: something else other than equivalence is controlling the subjects' behaviour; the fact that the subjects with verbal abilities have language puts them at an advantage to understand the task requirements or to discriminate between the stimuli; in the Barnes et al (1990) study only small subject numbers were employed meaning that 50% of the group with low verbal ability passed and 50% failed, which is chance level. To say unequivocally that subjects with no language fail to demonstrate equivalence more research is needed.

The third test of the Naming Hypothesis is as follows:
(iii) If naming is the sole determinate of stimulus equivalence, then subjects who initially fail and are then taught to assign names to the experimental stimuli should go on to demonstrate equivalence.
This prediction yields the most evidence. Horne and Lowe outline the following studies. For example, Lowe and Beasty (1987) tested three groups of children aged between two and five years. The younger subjects initially failed the equivalence test and were taught to appropriately use some of the verbal strategies which had been successfully employed by the older subject who had passed immediately. After this intervention the younger subject went on to pass too. This evidence indicates that what the subjects say during the training sessions or how they describe the relations between the stimuli dramatically effects their performances.

Similar evidence has been presented by Eikeseth and Smith (1992) with autistic subjects. Four autistic children (aged three to five years) were taught the conditional discriminations AB and AC and were tested for evidence of equivalence. Initially all subjects failed on the test trials.

They were subsequently taught to assign common names to the A, B and C stimuli after which two of the subjects went on to demonstrate equivalence. Of the two remaining subjects one performed above chance level. The subject who failed in every phase of the experiments also scored the lowest on measures of both expressive and receptive language.

The reported improvement in performance may not be entirely clear cut. For example, looking at the graphs of the data in the Lowe and Beasty study, the effect of the training intervention appears dramatic. It seems that if the subjects who fail are then presented with the instructional training intervention on the baseline trials only, they go on to pass. A multiple baseline design is employed so that two subjects receive the intervention and two do not until after a second equivalence test (so there is a comparison between the performance of subjects who do and do not receive the intervention). The subjects who do not receive the initial verbal training continue to to demonstrate equivalence. However, during the intervention, test trials are also presented giving the subjects extra exposure to these trials which may have facilitated their performance. Mandell and Sheen (1994, p.31) make a similar point, "The subsequent emergence of equivalence relations might thus be attributable to the additional training as well as to the naming instruction." The point is that it may be the instructional training which gives rise to equivalence but it cannot be said conclusively.

The same point can be applied to the Eikeseth and Smith (1992) study. These data are also open to interpretation. Although one subject (Trey) showed a marked relationship between naming and the successful demonstration of equivalence, it was not as defined for the remaining three subjects. Further, again (as in the Devany et al 1986 study) this is data, which although not without importance, has been gleaned from a retarded population with impairments which possibly go beyond just verbal limitations.

Taking a different approach, Saunders, Saunders, Williams and Spradlin (1993) looked at the performance of adult subjects with mild learning difficulties on various training procedures. They found that their subjects were more successful using a "many-to-one" or "comparison -as-node" (CAN) procedure whereby multiple samples were paired with one comparison stimulus. An example stimulus class would be as follows: B1A1, C1A1, D1A1. This when compared to the more widely used "one-to-many" or "sample-as-node" (SAN) protocol. For instance, A1B1, A1C1, A1D1. It was hypothesised that the instructions given to the subjects on the CAN procedure facilitated their performance. That is on the first four training trials subjects were instructed as follows and names assigned to the stimuli as follows, "when the X (sample name) comes up you push the button under the Y (comparison name)." Six subjects received such instructions and five subjects passed the equivalence task. Five subjects were not given any instructions; from this group only one subject demonstrated equivalence.

The authors concluded that the instructions facilitated the CAN procedure because they prompted covert anticipatory naming of the comparison stimuli (whereas this did not occur with the SAN procedure as not all the sample-comparison examples featured in the instructions). The subjects who failed were then given the above instructions and went on to successfully demonstrate equivalence. In this way it is possible that it is the instructions which made the task easier for the subjects. However, it must be noted that the latter effect of implementing the instruction intervention only applied to certain stimulus sets so it can not be said definitively that the improved performance was entirely down to the instructions, but from this evidence it can be said that they play a role.

Critics of the three studies cited so far may say that definitive conclusions can not be drawn from these data because they employ children and/or retarded populations meaning that other cognitive skills (apart from the targeted verbal abilities) may be impaired or not yet fully developed. Mandell and Sheen (1994) address this by investigating a normal population. The stimuli utilised are manipulated in order to effect the likelihood of the subjects producing names for them. In their first study three different sets of stimuli were employed: (i) phonologically correct nonwords (that is, ones that are pronounceable); (ii) non-phonological nonwords; and (iii) strings of punctuation marks. Their hypothesis was that if naming was an important mediator then the subjects who were taught the conditional discriminations using the first set of stimuli should require less training and make fewer errors. This was found to be the case. In the second study, subjects who were pretrained to read non-phonological nonwords as opposed to writing them down made significantly fewer errors on the subsequent tests of equivalence.

Yet can it be said definitively that the effects of the first study were the result of the stimuli being more 'nameable.' As the author themselves acknowledge, it was possible to assign names to the non-phonological stimuli (for example, by arranging the letters into a recognisable string) and similarly to the series of punctuation marks (for example, by attending only to the first element only of the sequence). Looking at the second experiment the read-aloud group made significantly fewer errors on the subsequent test of equivalence. This was attributed to these subjects being more predisposed to name the stimuli. This inference was drawn though from post-experimental interviews which can be notoriously unreliable. There is no means of assessing whether the subjects named the stimuli during the matching-to-sample procedure so this is evidence which can be only viewed at the most as being correlational.

## Conclusions

When approaching the wealth of stimulus equivalence research one is confronted with, generally speaking, a single paradigm (adhering to the Sidman et al 1982 definition) and at the same time a variety of terms, interpretations, and theories which attempt to explain its occurrence. In order to extricate a valuable research hypothesis some rationalisation must take place and it must be decided which of the accounts of stimulus equivalence is the most pertinent and can be empirically tested.

For the purposes of this thesis the Naming Hypothesis as detailed by Horne and Lowe (1996) has been adopted. As noted in the previous section, the evidence which is most usually cited in its favour is by no means unequivocal. However, the fundamental premise (that naming results in the successful demonstration of stimulus equivalence) is potentially falsifiable which makes it a valuable theory. In addition, it is the route most likely to be taken by verbally competent human subjects and because of this alone justifies further investigation.

Horne and Lowe set out three predictions which can be tested. The first is that nonhumans should fail to demonstrate equivalence. This would present definitive evidence if nonhumans did pass a test of equivalence. However, this prediction is weakened because Horne and Lowe also state that there is a possible second route whereby equivalence can be demonstrated without the use of language.

Secondly, there is the assumption that human subjects with no language or who are language disabled should similarly fail to exhibit equivalence. The problem which arises here is that subjects may not only be deficient in verbal ability but other cognitive skills may be lacking or not yet fully developed (for example, memory or attention span). This could effect performance on a matching-to-sample task and could cast doubt on any potential conclusions.

The third test of the Naming Hypothesis is that if subjects who initially fail to demonstrate are then taught to assign names to the stimuli then equivalence should emerge. The key problem here is that any intervention may mean extra exposure to the stimuli which could also benefit subjects' performances. Once again evidence cited to support the above premise employs children or subjects with learning difficulties which present difficulties as outlined above.

The solution is to employ developmentally stable adults of average intelligence. This is what Mandell and Sheen (1994) sought to do by utilising adult subjects but by manipulating the stimuli making then easy or difficult to name. Their findings suggested that there exists a relationship between how nameable the stimuli are, how much training is required and how many errors are made during the test session. However, this methodology gives little insight into whether naming does or does not play a role in the formation of equivalence classes. In order to examine this it seems fruitful to employ a population who are fully-developed (that is adults); who have no limited intellectual abilities; but who have deficiencies which concern manipulating symbolic material or language. Such subjects should provide evidence regarding the extent of the role of naming in the successful demonstration of stimulus equivalence.

# Chapter Two

## Developmental Dyslexia

# Developmental Dyslexia

## Background

When introducing the paradigm of stimulus equivalence a historical perspective was adopted in order to demonstrate that although the current definition of the term as it stands has only been in usage for the past decade or so, it has its basis in the deep-rooted philosophies of the eighteenth and nineteenth centuries. A brief exploration into the modern day term uncovered many differences and deviations from its initial employment.

In many ways the same can be said of the term developmental dyslexia. During the nineteenth century clinicians had noted that injury or disease of the brain could lead to language disorders (see Head, 1926 for a discussion of the early history of aphasia) leaving patients unable to produce appropriate speech. Kussmaul (1878) first talked of 'word-blindness', a language disorder whereby sufferers were unable to read text. A German optometrist, named Berlin used the term 'dyslexia' in 1887 as a substitute for 'word blindness' in an attempt to escape the suggestion of literal blindness in such patients, to describe individuals who were deficient in reading due to damage or disease to the brain. What they were both discussing, however, was an inability to read text, despite the non impairment of other cognitive faculties. Such observations include fundamental characteristics of dyslexia as it is identified nowadays, yet such early studies were based on instances of acquired problems.

Soon after congenital or developmental cases became subject to discussion whereby individuals were born with their deficiencies and had not acquired them as a result of subsequent brain injury or disease (Hinshelwood, 1917; Morgan, 1896; Orton, 1937). Hence emerged the term developmental dyslexia as it is applied today[1] (see Miles and Miles, 1990; Richardson, 1992; and Thomson, 1990, for a detailed discussion of the historical lineage of the term). However does it still represent the same concept as it was originally applied to?

The first documented cases describe individuals who had severe reading and spelling difficulties. Taken literally, a dictionary definition of the word *dyslexia* tells of its origins from the Greek *dys,* meaning badly or with difficulty, and a derivation of the Greek or Latin word *lexis* (or *lexicon*), meaning pertaining to words or speech. Again there is much emphasis on language processes such as reading and spelling. Nowadays unusually poor reading remains still at the centre of the majority, if not all, definitions. For example, Pavlidis (1990) writes:

> Dyslexia could be described as a syndrome that is best exemplified by an unexpected severe reading retardation which is not caused by any known intelligence, psycho-educational or environmental factors.
>
> (p.3)

In this way, dyslexia is most consistently characterised by reading failure, which cannot be explained by external, educational, social, or emotional factors. The qualification is

[1] Throughout the text the term dyslexia will be used to refer to developmental dyslexia unless otherwise stated. In addition other terminology such as, 'reading disabled' or 'poor readers' may be used in accordance with the terminology chosen by specific authors. In these cases the subjects they refer to fit the above definition of dyslexia.

also made that dyslexic individuals must be of average or above average intelligence as it is complicated to view it otherwise. However, it must be noted that there is no disputing that low ability people can also be dyslexic.

This reliance on the reading element is typified in research where subjects who would satisfy the above dyslexic criteria are alternatively referred to as poor readers, backward readers or as being reading disabled. Yet in many ways this is an over-simplification of the matter. The above definition could be seen as unhelpful because it is exclusionary in that dyslexia is described in terms of an absence of a skill (for instance, reading) and more importantly for the dyslexic individual by the absence of a skill which is not usually measured until a child is of school age.

In recent literature there has been a call to move away from such conceptualisations because clinical observations of dyslexia (such as those reported by Miles, 1993) suggest that it can no longer be seen as a deficiency solely of the skill of reading. In this way the label, dyslexia, as it was literally applied by early pioneers in the area, has over the years, as will be shown, expanded to become a much broader and useful concept.

After setting out the traditional definition of dyslexia Pavlidis (1990) goes on to embellish this viewpoint and predict that resulting from current neurological findings, "...dyslexia should also manifest itself in tasks other than reading. Such tasks, however, should simulate important

components of the reading process, i.e. sequencing, timing, attention and occulomotor control.." (p.4).
He goes on to outline how dyslexic individuals have been observed to be deficient in skills other than reading such as: poor spelling; directional confusion; short term memory problems; tasks involving symbolic material; timing and coordination problems; and eye movement irregularities (Pavlidis, 1990 p.6-7).

A similar characterisation of dyslexia has been expounded over the years by Prof. Miles (specifically Miles, 1993). He urges us to remember that dyslexia is a syndrome as applied in the medical sense, comprising various symptoms which may or may not be evident in each dyslexic individual. It is thus preferable to identify a pattern of deficits as opposed to focusing in on just one shortcoming (reading failure). Miles (1995, 1994a, 1994b) suggested that dyslexia should be defined as a 'mismatch' between skills. After reviewing anatomical evidence which suggested categorical differences between the structure of dyslexic and control brains (for example, Livingstone, Rosen, Drislane and Galaburda, 1991) he concluded, "...that for constitutional reasons dyslexics have a distinctive learning style and an unusual balance of skills: they may be slow at tasks which involve the processing of symbols at speed and yet often show high reasoning and creative powers" (Miles, 1995, p.25).

Therefore, the dyslexic would show the mismatch of being good at certain tasks which involve creativity or reasoning (Thomson, 1990 p.243 describes some subject areas in which some dyslexic individuals have been seen to excel) yet poor at language tasks especially if they must be performed under the pressure of time. Moving away from an exclusionary definition by emphasising not only the absence of some skills but the presence of others, Miles (1994b) supports the view that dyslexia should not be associated solely with a reading deficit as this shifts the emphasis away from other relevant difficulties or relevant strengths; added to this is the fact that many older dyslexic children are able to read at a reasonable level but are still indisputably dyslexic.

The picture that emerges of dyslexia is one which goes beyond a mere reading disability, that of a combination of impairments and the presence of, in some cases, above average abilities. This leads to the question of what underlies the syndrome of dyslexia. On one hand it must be a deficit which can account for all the wide range of observed difficulties outlined above, yet on the other hand, it must be a specific impairment in that it only effects certain skills whilst leaving others intact. The exact nature of the deficit which could meet the above criteria is the topic of much discussion.

## Underlying Deficits

The question of what is the deficit which underlies the dyslexic symptoms is a complicated one and one which has not yet been resolved satisfactorily. Researchers have attempted to address the question at various levels.

At its most fundamental level dyslexia could be seen as having a neurological basis. Clinical observations such that there is a higher predominance of dyslexia in males than females and that there is a high familial incidence of dyslexia have prompted research into the genetic origins of the syndrome. For example, LaBurda and Defries (1990) reappraised the numerous studies involving twins which have been carried out and discovered that there was a higher concordance rate for reading disability in monozygotic twins than in dizygotic twins. Linkage analysis (which maps the location of genes on chromosomes) is currently being undertaken to connect dyslexia to a specific chromosome(s), (for further discussion see Pennington, 1991).

Such research is still highly speculative but it is possible that certain genetic factors could influence neurology and result in some of the anatomical irregularities observed in dyslexic brains by A. Galaburda and colleagues. In a series of post-mortem examinations of dyslexic brains ( which are reported in Galaburda, 1994) certain abnormalities were identified. In the temporal lobe (which comprises of the Wernicke speech area) the two plana were found to be symmetrical (in control brains asymmetry has been reported in 70-80% of cases examined). Cortex malformations, for example, ectopias (cells intruding into other layers) and dysplasias (the disorganisation of cells within a particular layer) were also discovered in areas pertinent to language.

Livingstone et al (1991) observed that neurons in the magnocellular layer (which processes fast, low contrast information) were smaller than usual. This could manifest itself in a processing impairment of the visual and auditory systems especially when rapid processing of information is required.

As mentioned previously this line of research is still very much in its infancy and much of the data needs to be replicated. Its importance, however, must not be underestimated as this is the starting point from which the cognitive impairments which manifest themselves in the observed dyslexic difficulties could originate.

The next level which has been of interest to researchers suggests that there is a specific cognitive impairment (which could be a consequence of an impairment at the previous level, in other words, have an anatomical basis) which results in the observed dyslexic deficits. For example, a dyslexic individual's poor reading is the consequence of a breakdown in one (or more) of the component skills which constitute effective reading. It has been argued in the previous section that dyslexia should be regarded as a cluster of symptoms and thus more than a reading disability, therefore, if a cognitive impairment should be identified it must be able to account for all the dyslexic difficulties.

The most consistently well-documented contender for this impairment is a phonological deficit. Phonological processing is an integral part of language production Miles and Miles (1990) simply describe it as follows, "...'phonological processing' refers to operations by which stimuli are interpreted in terms of the speech sounds involved"(p.65).

In other words, it is the transforming of incoming stimuli irrespective of their modality into sound-based representations for use in speaking, reading and spelling. Galaburda (1994) tentatively links such a hypothesis to his neurological findings by suggesting that developmental dyslexia, "...originates as a disorder of perception affecting the brain at the vulnerable time (before the age of 1 year) when phonological structures relating to the native language are being organised in the developing brain"(p.134).

Prior to such anatomical research the phonological deficit hypothesis grew from observations that dyslexic subjects differed from control subjects when verbal material was presented. For example, Ellis and Miles (1978) presented 10-15 year old dyslexic and non-dyslexic boys with a variety of visual and verbal tasks. Group differences were only observed only on tasks which involved the use of language. On what they deemed purely visual tasks no differences were found. Vellutino (1987) cites similar findings. In one example, groups of dyslexic and control subjects were required to copy unfamiliar Hebrew letters and words following a brief exposure to them. This task was thought to be nonverbal or visual in that the letters were totally meaningless to both groups. As a result no differences in performances was noted between the two groups. In a further study cited by Vellutino (1987), two groups of poor and average readers were in the first phase of the experiment required to copy words, scrambled letters and numbers (for example, DNV, 832) from memory after a brief exposure. In the second phase the subjects were asked to recall the name of each character of the stimuli presented, in the correct order. Poor and normal readers did equally well on the first task but not on the naming task where the poor readers' performances fell below that of the normal readers. The author concluded that the deterioration in the performance of the poor readers was due to the verbal as opposed to the visual nature of the task. Given that dyslexic individuals exhibit a deficiency in the handling of verbal material the onus is on researchers to specify which particular stage of the verbal process is problematic and many authors have sought to do so.

In one of the earlier examples, Ellis and Miles (1981) presented their interpretation of the data, taking a cognitive perspective and using information processing models. This has been updated over the years but it is useful as a basic explanation of the structures involved in reading which broadly speaking remain unchanged. The premise behind this approach is that if a model can be formulated which represents the various components involved in the skill (in this instance, reading) then the point of breakdown can be pinpointed. In the first instance the authors outlined model A which is paraphrased as follows. In order to read a single word, a visual stimulus ('WORD') is presented: this enters the Visual Information Store (or VIS) wherein a visual/physical representation of the stimulus is held. This is not actively maintained so it fades rapidly. If further processing is appropriate the representation is transformed into a more lasting visual code; followed by a lexical code; an articulatory code; resulting in speech, "WORD" (NB the model is presented as a sequence of events but the authors suggest the coding processes could be performed also in parallel). At the lexical coding stage, a cognitive/semantic 'lexicon' is accessed. It is hypothesised that within the cognitive system there is an internal lexicon which contains visual/auditory representations against which incoming stimuli are compared. The authors proceed to test this model and evidence is presented to support the notion that there is no real disturbance at the VIS level for dyslexic subjects. Similarly there are no reliable differences at the level of visual coding, "It must be concluded that there are no essential differences between dyslexic and control children in respect of speed of visual coding, visual code capacity or rate of decay of visual code"(p.185).

A reliably reported dyslexic deficit is that the dyslexic individuals demonstrate a slowness in producing the names of stimuli. The authors say because this deficit is observed across a variety of stimuli (for example, words, objects, colours and so on) it cannot be a pattern recognition problem, therefore, it must be one of lexical or articulatory coding. They go on to present evidence which suggests that dyslexic subjects utilise the articulatory loop (the facility in short term memory for the subvocal rehearsal of the material to remembered) as well as control subjects and are affected by tasks involving articulatory suppression in the same manner also. If the groups differed with regard to articulation they should as a result differ on these tasks. Therefore, articulatory dysfunction may not be implicated in dyslexic subjects impaired performances.

This notion that the dyslexic problem is not one of articulation is reiterated by Anderson, Podwall and Jaffe (1984). They compared dyslexic and control subjects' performances on a rapid naming test in an attempt to discover which aspect of the process resulted in the dyslexic subjects slower naming. In order to do this the subjects' speech patterns were analysed and it was found that the dyslexic subjects paused more often and articulated more slowly. The authors suggested that this was due to the fact that the dyslexic children needed more cognitive preparation time (i.e. preparing for the naming of the adjacent stimulus) and not because they took physically longer to produce the names. They concluded, "It is this preparation task which occurs partly during pausing and partly during vocalising, rather than the task of articulating the name, that is the primary source of naming difficulty in these children"(p.83).

This leaves the stage of lexical coding as the point of impairment. To support this theory Ellis and Miles present a study reported in N.C. Ellis' (1980) thesis involving a Posner experiment. The dyslexic subjects were found to be reliably slower and more prone to errors at judging same or different letter pairs presented in different cases (for example, Bb). There were no group differences when the letters were presented in the same case (BB). The study was repeated with the addition of an articulatory suppression task, the results however, remained the same suggesting that the dyslexic subjects' impairment was a result of a lexical coding deficiency.

A detailed model B was presented and the deficit was narrowed down to one involving the, "lexical encoding of visual events" (p.209). In other words, in this context the dyslexic deficiency lies in accessing or retrieving the appropriate sound-based codes from the lexicon.
At this point it may be advantageous to return to the original notion of a phonological deficit which is what has been broadly described by Ellis and Miles. A lexical encoding hypothesis as they describe it, encompasses a deficit in accessing the sound-based representations (from the lexicon) of visually presented material.

Vellutino (1987) although employing different terminology, describes a similar model. He uses the analogy of a library. In short term memory/ working memory, stimuli are transformed into symbolic/abstract representations for storage in long term memory where it is either filed or discarded. For a dyslexic individual, words are stored with incomplete, "...phonological codes - file cards, in the library model. Asked to call up the proper word, the child finds that he or she has not retained enough clues to the name of the word"(p.23).

Similarly, in a more recent account, Rack (1994) specifies further the processes involved in reading. In his version the individual sees the novel, written word and a certain extent of letter-sound (or grapheme-phoneme) information is derived a this point. In parallel to this, visual processing is occurring whereby specific segments of the new word activate identical segments of other known words. The phonological information associated with these visual representations are accessed and it is this phonological information along with any derived from the initial letter-sound analysis which influence the speaker's utterance when reading the word. This model is heavily reliant on phonological skills and it is these that are pinpointed by the author to be deficient. In this way, such models predict that the source of a dyslexic individuals' difficulties is in coding incoming stimuli into sound-based/phonological representations or retrieving them from memory. This includes incoming auditory stimuli. Brady, Shankweiler and Mann (1983) and Tallal, Miller and Fitch (1993) both present evidence to support the notion that the phonological deficit occurs with auditorily as well as visually presented information.

According to Wagner and Torgesen (1987) such a phonological processing deficit could manifest itself in three areas: (i) the development of phonological awareness as measured by tasks such as those involving the segmentation, blending or rhyming of words, that is being aware that words are made up of distinctive phonemes (sounds); (ii) phonological recoding in lexical access, in other words, transforming visual stimuli (for example, words) to sound codes for access to the lexicon for speech (for instance, producing names). Reading research suggest that there are two routes to lexical access: (a) nonlexical/ phonological route whereby words are 'spelt' out via grapheme-to-phoneme conversion; or (b) lexical/visual/nonphonological route whereby whole words are compared to whole word representations in the lexicon. (iii) Phonetic recoding to maintain information in working memory by which visual stimuli are coded as sounds to maintain them in short term memory, via rehearsal.

In further discussions of a phonological deficit in dyslexia, other researchers identify the same areas of impairment which if such a deficit existed would prove problematic (see Kamhi, 1992; Miles, 1995; Rack, 1994; Richardson, 1992; and Vellutino, 1987). A multitude of evidence to support the existence of such impairments has been presented elsewhere (see Rack, 1994; and Wagner and Torgesen, 1987 for detailed reviews), to enable the conclusion that there exists a deficit in phonological processing in dyslexic individuals. However, does such a deficit account for all the observed dyslexic difficulties?

There is evidence to suggest that poor phonological skills leads to poor reading ability in later development. Bradley and Bryant (1978, 1985) undertook a longitudinal study working on the premise that the experience a child has with rhyme at the preschool level might affect their subsequent reading and spelling ability. It was found that when a rhyming test (an odd-one-out task whereby subjects had to identify which one of a series of words was different to the others because it did not rhyme) was presented to pre-readers their ability to perform the task was correlated with later reading ability.

Maryanne Wolf and colleagues also embarked on a longitudinal study in the 1980s (for example, Wolf, 1984; Wolf, 1986; Wolf, Bally and Morris, 1986; Wolf and Goodglass, 1986; and Wolf and Obregon, 1992). They isolated another phonological-based skill on which dyslexic individuals have been shown to be deficient, that is naming, and investigated its relationship to subsequent reading ability. Their hypothesis was that the disruption of specific stages of the naming process could impede the development of reading because they share similar properties. It was found that performance on certain naming tasks (for example, rapid naming tests) was the best predictor of later reading ability. Children who took longer and produced the most errors on the rapid naming test went on to demonstrate detrimental reading profiles.

In this way, researchers have set out how a phonological deficit hypothesis can account for the dyslexic reading and spelling impairments. What about the other observed difficulties? Dyslexic symptoms such as mirror writing or letter reversals were once thought of as visual errors. Yet it makes more sense to view these occurrences in terms of a linguistic deficit. Miles and Miles (1990) argue how some letter reversals are not solely due to visual weaknesses:

> ...confusion between 'b' and 'p' might be due more to auditory similarity than to visual similarity, since they are voiced and unvoiced forms of the same plosive; or again, it is possible that the differences between 'd' and 'b' may be observed by unclear speech, since although these two letters are not very close auditorily they are close in their places of articulation.
>
> (p.66-67)

Added to this are the errors often cited as commonly appearing in dyslexic children's writing (for example, SAW for WAS) are not mirror images of each other. Vellutino (1987) prefers to regard such errors as ones of inaccurate word retrieval due to weak phonological representations on memory. Miles and Miles (1990) similarly propose that dyslexic errors in mathematical time tables or left/right directional confusability can too be explained by poor phonological processing or in the case of the latter weakness the lack of concrete phonological representations.

At this point, it should be noted that over the last few years alternative deficit accounts have been proposed. However, these are not necessarily mutually exclusive to a phonological deficit account. For instance, researchers such as Lovegrove in Australia and Stein in Oxford (see the relevant chapters written by the above authors in Fawcett and Nicolson 1994a for a full discussion of their work) have been investigating the extent of visual deficits in dyslexic individuals and have found differences in the dyslexic processing of rapidly presented visual information (Lovegrove); and in ocular dominance suggesting that dyslexic subjects show more unstable ocular dominance which could lead to reading impairments (Stein). Miles (1994a), however, states that it is possible for phonological deficits to occur *alongside* visual difficulties and that they may share the same neurological basis. He says that it should not be seen as a case of 'either..or' but instead 'both..and' (p.7).

A further claimant for the specific cognitive deficit which underlies the dyslexic difficulties has been recently put forward predominantly by R.Nicolson and A.Fawcett. These researchers claim that a phonological deficit hypothesis cannot account for all the observed dyslexic difficulties. A point Miles (1995) reiterates, "Although the hypothesis of a phonological deficit in dyslexia has, rightly, gained a large amount of support, there is reason to think that as it stands it is incomplete"(p.29). He goes on to suggest that although a linguistic deficit makes sense of the majority of the dyslexic impairment, some remain outside its scope. An example is dyslexic musicians' weaknesses in adhering to the correct rhythm or estimating time intervals.

Nicolson and Fawcett in a series of research papers pertain to a more fundamental deficit which underlies a broader spectrum of dyslexic difficulties. Nicolson and Fawcett (1990) reported findings which indicated that dyslexic subjects demonstrated weaknesses in gross motor skills such as balance when performed under dual task conditions (in other words in conjunction with a task tapping an unrelated domain such as backward counting or an auditory-choice task). The authors suggest that the noted decrement in the dyslexic children's performances could be attributable to what they term DAD or, "dyslexic automatisation deficit" (p.161). This relies on the cognitive theory that in order to acquire skills efficiently automatisation is needed. If a skill is automated it requires less cognitive load which results in: a higher speed of processing; the skill being well-founded in LTM and less susceptible to interference; and a diversion of cognitive capacity away from the fundamental skill (such as reading) to higher processes (such as comprehension). The authors hypothesise that dyslexic individuals do not automatise skills as efficiently and by this they mean cognitive and motor skills, hence their poor motor balance performance when a secondary task is introduced. Cognitive load is diverted to the new task (counting) to the decrement of the initial task (balancing). If the balancing task was fully automatised then the added cognitive load would not pose a problem, as demonstrated by the undisturbed control group's performance under the same conditions. The dyslexic subjects can perform the initial task as well as the control subjects due to conscious compensation, in that they 'work harder' to attain proficiency. The secondary task uses up this conscious processing capacity leading to impairment of the primary task.

This automatisation deficit, the authors suggest, accounts for the dyslexic deficiencies ranging from motor to cognitive skills. Further support for this hypothesis is provided by Nicolson and Fawcett (1994) where the administration of twenty-two measures to dyslexic and control subjects aged from 8-16 years is reported. Tests included those of phonological skill, working memory, information processing speed, and motor skill. Only on a measure of simple reaction time did the dyslexic group perform on par with their reading-age and chronological-age control groups on the other tasks, the dyslexic group performed worse than their reading-age controls.

Therefore, it was concluded that indeed there exist phonological deficits but also deficits in the processing of speed, memory and motor tasks - again implicating a non-phonological deficit. Nicolson, Fawcett and Dean (1995) demonstrate dyslexic deficiencies on a time estimation task and put forward the suggestion that this is indicative of a cerebella dysfunction where such skills are thought to neurologically originate. Such a neurological impairment if identified would account too for the motor skill and automatisation deficits already observed in dyslexic performances.

Although this view of dyslexia on the surface differs from that of a purely phonological deficit, the two are not necessarily exclusive. It is a matter of conceptualisation. If dyslexia is viewed in its widest sense as Nicolson and Fawcett purport then a far reaching account is indeed required. If dyslexia is seen in its narrowest sense as principally a reading deficit then a phonological deficiency would suffice and account for additional dyslexic difficulties too. It could be that the non-phonological weaknesses described by Nicolson and Fawcett lie outside the dyslexic cluster and merely co-occur with the syndrome. Or there could exist a sequence of deficit, for instance, a neurological impairment gives rise to automatisation disruption which leads to phonological deficits, which in turn lead to reading and spelling problems.

It is not the place of this thesis to speculate which, if any, is the most plausible account, if nothing more, at this stage there is too little evidence to form any definitive conclusion. The observation of most relevance to this thesis is that whatever the underlying cause, a phonological deficit is a major characteristic of developmental dyslexia. Nicolson and Fawcett (1994) when reviewing their studies state, "...phonological skill and motor skill appear to be least susceptible to improvement with age (with significant deficits even compared with reading age controls) on several phonological and motor tasks..."(p.229).

Thus when sampling a group of dyslexic individuals a phonological deficit will be a reliable and persistent finding. Rack (1994) sums up the position, “We would not want to restrict a definition of dyslexia, at this stage, to one of a phonological disorder. However, it is quite clear that phonological skills play a central role in the majority of dyslexics' difficulties”(p.30).

**A Naming Deficit**

A strong case has been made that developmental dyslexia is characterised by a phonological deficit irrespective of whether this is the chief underlying cognitive cause or merely one consequence of a more primitive impairment. Wagner and Torgesen (1987) when identifying the three main areas in which a phonological deficit would manifest itself included, ‘phonological recoding in lexical access’, in other words, accessing and retrieving phonological representations from the lexicon which constitutes naming. Seeing a stimulus (be it a word, nonword, object or colour) and accessing the corresponding phonological representation to produce a speech pattern - a name.

The actual processes involved in producing this name is not of relevance here. What is of importance is that a phonological deficit implies a naming deficit. Miles (1995) confirms this, “What has been suggested is that it is the SPOKEN REPRESENTATIONS of stimuli in the environment which are the essential stumbling block for dyslexics. In more simple terms, dyslexia is a weakness at verbal labelling”(p.28), in other words, naming.

There is much evidence to demonstrate that dyslexic individuals suffer from an impairment of naming be it overt naming or covert naming. Returning to the stimulus equivalence paradigm briefly, it was argued that according to the Naming Hypothesis in order to successfully pass a test of equivalence subjects must assign consistent names to the stimuli. If this occurs the subjects will always (because of the properties of the name relation) demonstrate equivalence. It was hypothesised that if naming results in the demonstration of stimulus equivalence then, a population of subjects who are characterised by a naming deficit should perform differently on such a task. Developmental dyslexia, as it will be shown, is characterised by such a naming deficit wherein dyslexic individuals are more prone to make errors and thus name inconsistently when naming visual stimuli and take longer to produce the appropriate names. If naming is intrinsic to demonstrating equivalence then the dyslexic subjects' performances should be more prone to errors and produce longer latencies on the equivalence test trials.

Consider overt naming first. In, 1972, Denckla observed that boys with developmental dyslexia acquired colour names later than non-dyslexic children. Furthermore, once acquired the production of these colour names remained slow over time (as measured by a repetitive timed colour naming task). As a result it was hypothesised that this slow colour naming could correlate with the slow and inefficient reading strategies which characterise dyslexia. Subsequent researchers set out to replicate this finding and to ascertain whether other naming tasks (for example, those involving objects, letters and numbers) were similarly affected in dyslexic subjects.

In a further paper Denckla and Rudel (1976a) sought to do this. They hypothesised that other visual stimuli would also show slow name retrieval times. They employed Rapid Automatised Naming tests (RAN tests) or continuous tests. These are defined by Anderson, Podwall and Jaffe (1984) as, "...rapid, repetitive responses to over learned stimuli.."(p.71). Subjects are presented with sheets of familiar stimuli (the names of which are often learnt at an early age and which are thought to become progressively 'automatised' throughout development) and are asked to name them as fast as possible moving from left to right across the page. Usually fifty stimuli are presented at once which include five different stimulus types repeated at random, ten times each (see Appendix I). The time taken to produce names for all the stimuli on the sheet is measured with the emphasis strictly on speed.

Denckla and Rudel (1976a) measured 128 subjects aged 7-12 years on four subtests: colours (red, green, black, blue, yellow); numbers (2, 6, 9, 4, 7); 'use' objects (comb, key, watch, scissors, umbrella); letters (high frequency, lower case letters p, o, d, a, s). The resulting data were analysed and the dyslexic group was found to be slower on all four subtests. The object subtest took the longest for all subjects. Colours were then named more slowly followed by letters and numbers (which were named equally fast). This same pattern of responding was found in a large scale study of normal readers undertaken by Denckla and Rudel (1974). Similar findings were revealed by Spring and Capps (1974) who measured the performances of 48 dyslexic and non-dyslexic boys aged 7-13 years using a RAN format for digits, colours and pictures. The dyslexic boys took longer than the control subjects on all tasks.

Wolf and colleagues carried out a comprehensive study of the naming deficit and its relation to reading using a longitudinal sample. Wolf (1984) reports how a battery of naming and reading tests was administered to a sample of 115 children before, during and after reading acquisition. Generally, it was found that the poor readers were significantly slower than the average readers on the rapid naming tests. Wolf (1986) reported findings gathered from a RAS naming test (rapid alternating stimuli). This was developed to investigate the ability of readers to direct their attention to contextual patterns whilst performing the continuous naming task under the pressure of time. This measure presented a combination of stimuli in a consistent sequence (for example, A-B-A-B or A-B-C-A-B-C). This was claimed to mirror reading more realistically because they share similar skills, such as combining an automatic task (for instance, naming a familiar stimulus ) with the skill of directing attention to higher level contextual patterns (to derive meaning). This task proved to hold powerful discriminating properties between average and dyslexic children in that dyslexic subjects took reliably, significantly, longer.

Katz, Curtiss and Tallal (1992) looked at the performance of 67 language impaired children and 54 age-matched controls on the RAN and also on a manual version of the RAN which involved nonverbal pantomime response. The LI children performed significantly poorer on both versions of the RAN suggesting that the dyslexic deficit is not limited to verbal output but generalises to other motor domains (supports the Nicolson and Fawcett hypothesis).

In this way, dyslexic children can be seen to be slower than control subjects at completing such rapid naming tests. However, is this type of naming analogous to the type of naming which should occur on the matching-to-sample sessions if stimulus equivalence is to occur? It is, in the sense that Horne and Lowe (1996) predict that if subjects assign individual names to the stimuli then they will name intraverbally and with self-repetition bidirectionality will emerge along with the appropriate listener behaviour. In this way subjects should be continuously naming.

On the other hand, the RAN tests present stimuli randomly, where intraverbal naming occurs in a set pattern, and are administered under the pressure of time where the MTS trials are not. The MTS task involves the presentation of stimuli in a discrete-trial format. So, although it can be said that dyslexic subjects demonstrate a naming impairment, as predicted by the phonological processing hypothesis, it must be established that this deficit is not restricted to the RAN type of task. Therefore, it is necessary to also look at the performance of dyslexic subjects on a naming task more analogous to the matching-to-sample format (that is, discrete-trial presentation). A variety of different measures fit this description but they all share the same characteristic, that of the stimuli to be named being presented individually.

Denckla and Rudel (1976b) compared dyslexic and non-dyslexic children on a simple naming test which consisted of 36 black and white line drawings of objects bound individually in a booklet. The experimenter asked the subjects to name the object pictured as quickly as possible and vocalisation onsets were recorded. The results demonstrated that the dyslexic children produced longer response latencies than the groups of non-dyslexic subjects suffering from minimal brain-dysfunction and the non-dyslexic controls. The dyslexic group differed from the control group by approximately 500ms on the less familiar names.

Perfetti, Finger and Hogaboam (1978) compared the vocalisation latencies of what they termed, 'skilled' and 'non-skilled' readers (aged 5-9 years), on a variety of stimuli which were manipulated according to set size, stimulus material and number of syllables. No group differences were found on subtests involving colours or digits but differences were observed on the printed words and pictures subtests. A mean of 400ms difference was observed between the two groups for producing the names of one syllable printed words, and this difference increased according to word length. There were no group differences for producing one syllable names of pictures but a difference of approximately 100ms was seen for producing two syllable picture names.

Stanovich, Freeman and Cunningham (1983) used a discrete trial methodology to measure the speed of letter naming in 5 year old children. On a task using lower case letters less skilled readers took an average of 140ms longer to produce names and 60ms longer to produce the names of upper case letters. Both were statistically significant differences. Bouma and Legein (1980) measured letter and word naming latencies using Dutch dyslexic and control subjects aged 11-15 years. The dyslexic group was on average 120ms slower at naming letters and 220ms slower at naming words than the control group.

In addition to the dyslexic children being found to be slower in naming than the control subjects on the discrete-trial methodology they have also been found to produce more errors. Snowling, van Wagendank and Stafford (1988) replicated a study carried out by Katz (1986) in which he looked at object naming in children with reading disabilities. He found that poor readers' performances were affected by: their ability in that they named fewer words; and by the word length and difficulty. Snowling et al measured the subjects' (aged 8-10 years) reaction times in addition to the number of errors made. No significant difference in latencies was found between the groups on the picture naming task or on the subjects naming responses to oral definitions. The dyslexic readers, however, did make more errors. The authors claimed that this was due to inadequate phonological representations of words.
In this way the dyslexic naming deficit may not be solely characterised by slow naming but also inaccurate naming. This may be of importance to the subjects' performances on the matching-to-sample task if this means that the dyslexic subjects are not naming the stimuli consistently. Inconsistent

naming would mean that the baseline relations may take longer to learn and ultimately would prevent the emergence of stimulus equivalence according to the naming hypothesis if consistent equivalence classes are not formed.

So far it has been demonstrated that a naming deficit occurs in dyslexic children of all ages using both the rapid automatised naming format and the discrete-trial methodology.
It has been questioned whether these two different types of naming tests measure the same latent variable. For example, Stanovich et al (1983) have said that the observed differences between the groups is less marked on the discrete-trial format because it is a purer measure of naming and that the RAN test measures much more than just naming. For instance:

> ...the discrete trial procedure provides a much clearer measure of name access time than does the continuous list procedure. The continuous list procedure involves complex scanning, sequential response and motor production strategies that could differentiate good from poor readers. (p.200)

This is reiterated by Wolf and Goodglass (1986) who state that the larger differences observed between impaired and average readers on continuous-list procedures can be attributed to the added cognitive demands of the serial presentation.
Specifically, Perfetti (1985) talks of a subject, who is confronted with a RAN type task, not only having to retrieve the name of a particular stimulus but also to produce it whilst at the same time preparing to process the next stimulus. In this way, it can be argued that the RAN protocol involves many more cognitive skills than purely the retrieval of a name code and as a result is not the most 'pure' measure of naming.

However, Bowers and Swanson (1991) have shown that dyslexic subjects (aged 6-7 years) demonstrate differences in naming speed on both the RAN and discrete-trial methodologies using single digits and letters. So, it can be said that if the two types of naming tests measure different aspects of naming they possibly co-occur in dyslexic individuals.

It should be noted that the naming deficit is not confined to artificial naming tests. Murphy, Pollatskak and Well (1988) took a group of dyslexic subjects (aged 10-11 years) who were characterised as slow namers as measured by the RAN. On a less formal naming task which required the subjects to retell a story the dyslexic subjects were found to be significantly slow and made significantly more errors. Therefore, if the naming which occurs during the typical matching-to-sample session is less formal or requires more than simply producing the name of a particular stimulus, there is a suggestion that the dyslexic subjects will still demonstrate a deficit.

The evidence cited so far concerns children up to the age of 15 years of age who are still developing cognitively. It can be argued with some certainty that such dyslexic individuals demonstrate a naming impairment using both the discrete-trial and RAN protocols. However, what is of particular pertinence to this thesis is whether this deficit persists into adulthood?

Studies involving dyslexic adults are scarce as they can no longer be as easily tested 'en masse', as in school, for example. If it is accepted that developmental dyslexia is deficit and not the result of a developmental lag which can be eventually overcome then the naming weaknesses should still be observed in adult populations, this a view voiced by Kamhi (1992), "Although individuals with dyslexia may learn to read fairly well, the phonological processing deficit that underlies the disorder never goes away"(p.50). There is empirical evidence to support this notion.

Wolff, Michel and Ovrut (1990) examined the performances of adolescents and adults on the colours and objects subtests of the RAN employing: dyslexic adolescents (13-18 years old); dyslexic adults (18-32 years old); non-dyslexic learning disabled controls (in order to determine whether any differences were specific to dyslexia); and non-dyslexic controls. In general, the adults performed with greater accuracy and naming speed than the adolescents but, more importantly, the dyslexic subjects in both age groups performed significantly more slowly than the control subjects.

Kinsbourne (1990) reports a study where severe and mild dyslexic adults were compared to a normal reading group on various tests of language. The severe dyslexic adults were found to be worse on the majority of the tests and especially so on the RAN objects subtest.

Similarly, Felton, Naylor and Wood (1990) compared the performances of 115 adults (mean age 33.1 years) diagnosed as dyslexic in childhood to control subjects on a battery of neuropsychological tests. The reading disabled subjects performed below the control subjects on all tasks. It was found that the tasks which most clearly differentiated the groups were those which required rapid naming and nonword reading.

So, dyslexic adults take significantly longer to name stimuli using the RAN style format. Fawcett and Nicolson (1994b) looked at naming speed in dyslexic children using a computer version of a discrete-trial format of the RAN tests. The experimental task required subjects to name objects, colours, letters and digits. The subjects who took part were three groups of dyslexic and control individuals aged 8, 13 and 17 years. This grouping allowed chronological-age and reading-age match comparisons. The dyslexic group when taken as a whole were significantly slower than their chronological-age matched controls on all four subtests. A average reported differences approximated 100ms between the group's latencies. This difference lessened with age but the authors took this as sufficient evidence that the dyslexic discrete-trial naming deficit persisted into adulthood.

Watson and Brown (1992) investigated single word reading in dyslexic individuals at college. Primarily, they were interested in the priming effects of, 'friends' and 'enemies' on word production ('friends' being orthographic words which share the same spelling-to-sound characteristics, for instance, *pill* has many 'friends' such as *mill, hill, till*; 'enemies' are orthographic words which do not share these properties and as a result do not facilitate the retrieval of the correct phonology of the target word, for example, *pint* has many 'enemies' such as *mint, hint, tint* - Watson and Brown's example). The dyslexic group's overall word naming times were found to be significantly slower than those of the other groups tested, on average ranging from 300-400ms slower. So, there is evidence of the naming deficit in adult, high-achieving dyslexic subjects.

In this way the dyslexic naming deficit is in evidence on a variety of overt cognitive tasks which require the subject to produce aloud the names of various stimuli (be it objects, words, colours, letters or digits) under various conditions and there is some amount of evidence to assume that this impairment persists into adulthood.

Another question to consider is whether overt naming accurately reflects the naming which occurs during a typical Matching-to-sample task? Subjects should also successfully demonstrate stimulus equivalence if they covertly name the stimuli. In fact, unless explicitly asked to do otherwise, most subjects will not name aloud stimuli during task performance. Therefore, it is necessary to demonstrate that the dyslexic naming deficit is not restricted to overt tasks but applies equally to covert naming tasks too.

This is a much more difficult case to argue for as measures of covert naming are not as widely reported. There is some evidence, however, to indicate that dyslexic individuals, because of their weakness at phonological processing, do demonstrate impaired performances on tasks which involve to all intents and purposes covert naming.

Ackerman and Dykman (1993) presented a battery of naming tests to 7-12 year old readers. One task involved a computerised rhyme decision task. Pairs of stimuli were presented and subjects had to decide whether the words rhymed or not. They responded by pressing the letter 'R' for right and 'W' for wrong. Poor readers made significantly more errors on this task than the control group. This suggests a weak internal phonological processing mechanism.

Murphy et al (1988) reported the performance of dyslexic subjects on what they term a 'receptive task'. Subjects were asked to categorise words spoken by the experimenter, using a manual response: pushing the switch up to indicate 'yes'; and down to indicate 'no' (this should involve the subjects covertly naming to some extent be it in processing the auditory stimulus or accessing the required category). The dyslexic group was significantly slower at performing this task. Specifically, they were 26% slower at categorising a correct response which resulted in them being on average 129ms slower than the control group.

Ellis (1981) employed a Posner type task. Stimuli were either visually identical (OO); visually dissimilar (OB); visually similar (OQ); phonologically identical (Bb); phonological dissimilar (Ba); or phonologically similar (Bd). Subjects had to determine whether the letter strings presented were the same or not by responding 'yes' or 'no'. No group differences were observed on the visual judgment tasks, yet the dyslexic group was significantly slower at responding when the name code was involved. On average they were 100ms slower than the control group. To confirm that it was the subjects covert naming which prolonged their responding Ellis and Miles (1978) report a replication of the above study using highly confusable, novel, visual stimuli, which were the type of stimuli which are typically presented in tests of equivalence. No group differences were found on this task. Thus it could be concluded that the problem was not one of processing visual information.

Bruck (1992) described a series of studies employing adult dyslexic subjects which assess the persistence of phonological deficits employing subjects aged 19-27 years. Some of the tasks reported reflect measures of inner speech (as defined by Levine, Calvanio and Popovics 1982; and Nebes 1975). One such task involved syllable counting. Subjects heard a nonword auditorily and used blocks to indicate the number of syllables the word possessed. A similar task was presented involving phoneme counting. Adult dyslexic when compared to non-dyslexic college students, made significantly more errors on the above task.

Therefore, although the evidence is more sparse, it is reasonable to hypothesise that adult dyslexic subjects persist in their phonological processing deficits and covert naming impairments. Added weight is provided given that as reported earlier, Anderson et al (1984) found that articulation time was irrelevant in discriminating between dyslexic and control groups; and Ellis and Miles (1982) claim that dyslexia is characterised by a lexical encoding deficit and not an articulatory loop problem.

To summarise, it has been argued that dyslexic individuals reliably demonstrate a deficit in both overt and covert naming. Such a deficit manifests itself in a more error-prone performance and a slowness in producing the appropriate name. Dyslexic subjects have been showed to be significantly slower than control subjects depending on the task requirements producing latencies ranging from 100-500ms higher and this weakness appears to persist into adulthood.

Therefore, there is firm reason to believe that if subjects name during the matching-to-sample equivalence test session then the dyslexic subjects' performances should differ from those of the control subjects.

# Chapter Three

## Allying the Theoretical and the Practical

**Allying the theoretical and the practical**

The purpose of this chapter is to build a link between the theoretical (as expounded in Chapters 1 and 2) and the empirical (the proposed experimental measures).

Chapter 1 outlined the paradigm of stimulus equivalence and focused on the naming hypothesis (Horne and Lowe 1996). Here it was stated that naming is necessary for the successful demonstration of equivalence. Put another way, if the subjects are assigning names to the stimuli then stimulus equivalence must automatically follow and the hypothesis sets out how the properties of the name relation bring about the equivalence relation.

The following studies which are to be reported employ a visual-visual matching-to-sample task. Horne and Lowe (1996) present two specific types of naming which are sufficient for stimulus equivalence to emerge on such a task. Those being: (i) by assigning a common name to the members of each equivalence class; and (ii) intraverbal naming.

Common naming is most likely to come about if the subject views the stimuli in each class as sharing some kind of physical property (for example, they may all contain straight lines, or curved lines). In this way, the same listener behaviour (which is occasioned by the shared name) is applied to each member of the class and in this way they are matched together appropriately in a MTS setting (for instance, A1 "straight" - select - B1 "straight" - select - C1 "straight").

Intraverbal naming would occur if the subject assigns individual names to the stimuli. This way the subject is during the training sessions learning to match together stimuli with different names in accordance with the reinforcement contingencies (for example one stimulus set would be trained as follows, A1 "square" - B1 "cross"). In order to maintain this relation the names of the stimuli would be rehearsed together until intraverbal naming is established (for example, "square-cross-square-cross") to the point where any one member of the pair occasions the other. When the second relation (BC) is trained (B1 "cross" - C1 "hand") an intraverbal equivalence class should emerge mediated by the name common to both relations ("square-cross-hand-square-cross-hand") whereby any one member of the class should occasion any other and along with it the appropriate listener behaviour (that is when one member appears as a sample, select the other member of the class).

According to the naming hypothesis if either of these types of naming occur stimulus equivalence should emerge. If this does not happen and subjects fail the equivalence test, then there are two possible explanations: (i) the subjects are not naming; (ii) the subjects are not naming consistently or congruently with the experimenter defined classes.

Evidence was presented to support this hypothesis which although valuable is by no means definitive. One criticism lay in the fact the majority of cited studies employed populations of subjects who were either children or individuals with some degree of learning disability. As a result it cannot be claimed unequivocally that these findings were the result of the presence or absence of naming skills as other cognitive features may have been impaired or not yet fully developed.

Chapter 2 introduced the main characteristics of the syndrome, Developmental Dyslexia. The use of such a population of subjects rules out the above problem because, by definition, dyslexia is observed only in individuals of average or above average intelligence (eliminating the issue of the possible presence of a secondary impairment); and using adult subjects eradicates any developmental problem.
It was noted that one of the most reliable characteristics of dyslexia is a phonological processing deficit which manifests itself in a naming impairment (both covertly and overtly) resulting in dyslexic individuals taking longer to produce the names of stimuli and being more prone to making errors. Evidence was presented to support the claim that this deficiency persists into adulthood.

Therefore, there exists an hypothesis which predicts that naming is required in the successful demonstration of stimulus equivalence. Juxtaposed to this is the notion that a chief symptom of developmental dyslexia is a deficit in naming. Taking theses two pieces of evidence together, it can be seen how studying the performances of dyslexic subjects could provide some insight into whether naming does or does not play a role in the emergence of stimulus equivalence. If naming is required for the demonstration of equivalence then dyslexic subjects (who possess a deficit in naming) should perform differently on a test of equivalence.

The first question to be addressed is how to quantify this difference? The nature of a dyslexic subject's phonological deficit suggests that she/he will be more prone to making errors when naming is required due to inadequate phonological representations. For instance, Wolff et al (1990) looking at the performance of adult and adolescent dyslexic subjects on a RAN test found that they made more errors than the control subjects. Katz (1986) and Snowling et al (1988) found that dyslexic readers made more naming errors using a discrete-trial methodology. This would affect a dyslexic subject's performance on a matching-to-sample task if their error-prone naming meant that the names assigned to the stimuli were inconsistent (that is, sometimes one name is produced but on another occasion an error is made).

Manifestations of this impairment may include a dyslexic subject initially taking longer to learn the baseline relations during the training sessions. It is not being argued that naming is necessary for these relations to be learnt (for example, nonhumans can be taught to respond correctly on such conditional relations), just that naming should accelerate learning here as it serves as a means of discriminating between the stimuli. If consistent names are being assigned to the stimuli then this may facilitate the task of forming the correct relations between the stimuli. On the other hand, if inconsistent names are applied, this could mean that more experience of the correct relations (in other words more training trials) are needed before they are appropriately learnt.

This has been empirically demonstrated using nonhumans (Cohen, Looney, Brady, and Aucella 1976; Sidman et al 1982; and Urcuioli, 1985) where subjects who have been taught sample-specific responding (that is, responding in a topographically distinct way to each sample) beforehand have been found to learn arbitrary conditional discriminations quicker (than subjects who have not received such training) if they continue to respond consistently to the sample stimuli.

The closest evidence to support this prediction comes from studies which measure dyslexic/non-dyslexic performance on paired-associate learning tasks. These are not procedurally comparable to a matching-to-sample task where naming is involved as the verbal response does not mediate performance. However, a PAL task does involve learning associations between often abstract stimuli and maintaining this performance over time in the sense that performance is measured as the number of trials required until no errors are consistently observed. Vellutino (1987) reported the performance of poor and average readers on two PAL tasks. One comprising of visual-verbal pairs (novel abstract, pictorial stimuli and nonsense stimuli) and the second consisting of purely visual pairings. The poor readers did less well on the visual-verbal task. Therefore, the dyslexic difficulty lies not in forming and remembering the associations (the poor and normal readers performed equally well on the visual-visual tasks) but in assigning and consistently applying the names to the stimuli.

Similarly, Done and Miles (1978) compared dyslexic and control adolescents (mean age 14 years) on a PAL task involving the pairing of nonsense shapes and words. Once again the dyslexic group needed significantly more trials before the associations were reliably formed. The authors claim that this represents a weakness at building up phonological representations and this is what leads to more errors being made. Further, the task is not an impossible one for the dyslexic subjects, just more practice is needed.

So, translating this back to the training of the baseline relations, if the dyslexic subject assigns names to the stimuli it can be expected that more exposure to the stimuli is needed before an error-free (or consistent) performance is observed. Therefore, it is important to look at the number of trials needed to reach criterion and the number of errors made during those training sessions.

If the dyslexic subjects' naming is weak this could mean that their performance may be more vulnerable to disruption when any reinforcement is dropped (which is standard practice during the training sessions to ensure that the learned relations can be maintained without feedback in preparation for the test session). As a result dyslexic subjects may make more errors on these unreinforced trials due to their weaker relations between the stimuli.

On reaching the test session the procedure requires that subjects must be consistently responding on the baseline relations. Regardless of whether the subjects named during the training sessions, according to the naming hypothesis naming

must occur during the test session if the properties of equivalence are to be demonstrated. In this manner, inconsistent naming will lead to failure on the test session and more errors being produced across the trials.
Should the dyslexic subjects elect not to assign names to the stimuli (in that they may be conscious of their impairments and avoid using such a strategy) this according to the naming hypothesis should also lead to failure to demonstrate equivalence. However, if the dyslexic subjects choose an alternative strategy (for instance, a nonverbal one) which is just as effective and go on to pass the equivalence test this would not support the fundamental premise of the naming hypothesis.

The only direct means of ascertaining whether the subjects were or were not naming on these test session trials is to examine the latencies. For the subjects who successfully demonstrate equivalence (who by definition make relatively few or indeed no errors) the difference in the dyslexic/non-dyslexic naming speed should be reflected in the response latencies on the test trials, if the subjects are naming. Put another way, if dyslexic subjects are slower at producing the names of stimuli, and naming is an integral part of forming equivalence relations, then the dyslexic subject should be appropriately slower on the test session trials (that is the time taken between the comparison stimuli appearing on the screen and the subject making his/her choice).

This assumption, that a subject's response latency reflects an underlying naming strategy has been investigated by Bentall, Dickins and Fox (1993). The purpose of their study was to distinguish between two theoretical stances. The first was the Associative Network Theory (Fields, Verhave and Fath 1984; Fields and Verhave 1987; and Fields, Adams, Verhave and Newman 1990) wherein the control a derived relation holds is inversely related to the number of nodes which link the stimuli. For example, in the stimulus class ABC, AB/BA and BC/CB are single/no node relations; whereas AC/CA trials which involve transitivity are distanced by one node and thus making their control less strong. This diminished control can be characterised by more errors on these trials (Fields et al 1990) or increased response latencies (Wulfert and Hayes 1988).

The second strategy Bentall et al outlined a verbal mediation/naming hypothesis whereby a subject's matching behaviour is controlled by a common response/name. Bentall et al hypothesised that such a strategy would manifest itself in equal response latencies regardless of the trial type (compare with Dugdale and Lowe 1990, Horne and Lowe 1996). Fields et al (1990) predictions encompass this pattern in that a common name will counteract the nodal distance effect by transforming AC/CA relations into a single/no node relation (in that the responding is controlled by mediating response and not a mediating stimulus A-B-C).

Bentall et al assigned three groups of subjects to the following conditions: (i) preassociated stimuli - where the stimuli fell into clear semantic categories (for instance, plants or planets); (ii) non-associated but nameable stimuli;

and (iii) abstract stimuli which were difficult to name. Subjects in groups (ii) and (iii) made significantly more errors on the trials which involved transitivity and produced significantly longer response latencies on these trials. Few errors and little difference between the response latencies on the various trial types were seen for group (i) who were presented with the preassociated stimuli.

These findings were replicated in a second more experimentally rigorous study. The authors concluded that the latencies did not differ when preassociated stimuli were employed because the subjects were assigning common names to the stimuli and that the patterns demonstrated by the non-associated and abstract stimuli groups reflect the responding predicted by the Associated Network model.

The latter pattern of responding is not necessarily at odds with a verbal mediation account as there is the possibility that subjects are assigning individual names to each stimulus (see Dugdale and Lowe 1990; Horne and Lowe 1996). If this is occurring the nodal distance effect would still be observed. For example, the subject may intraverbally link:
AB/BA - AB - “house-pink”
BC/CB - BC - “pink-box”
The AC/CA relation would entail: A-B-C - “house-pink-box” or C-B-A - “box-pink-house”; and involve more complex naming which should initially take longer to produce, resulting in the trials involving transitivity taking longer and being more prone to errors.

Bentall et al, however, dismiss the possibility that this form of naming would be more prone to errors (and as result dismiss as the strategy underlying the AC/CA inflated latencies pattern). They propose that even if the subjects assign individual names to the stimuli this should lead to an error-free performance on the test session (which is not observed in the AC/CA inflated latencies pattern where significantly more errors are observed - see Experiment 3, Bentall et al 1993). This observation relies on the assumption that the equivalence classes are formed prior to the test session and that they do not emerge during the actual test session, if names are assigned to the stimuli. However, merely applying the names to the stimuli (during training, for example) is not sufficient in bringing about equivalence. The crucial feature is that the names are put together intraverbally (Horne and Lowe 1996) and it is this which could take time to emerge during the test session, resulting in initial errors being made.

Therefore, it could be argued that both patterns of responding predicted by Bentall et al could reflect underlying naming strategies. This is of direct relevance to this thesis which is exploring the effects naming differences have on subjects performances. It is important to note whether these patterns are observed in the subjects who pass the equivalence test who according to the naming hypothesis should be assigning names to the stimuli.

To summarise, if subjects who successfully demonstrate equivalence do assign names to the stimuli it is predicted that the aforementioned differences in performances should be observed between groups of dyslexic and non-dyslexic subjects.

If such differences are observed this will add support to the hypothesis that naming plays a role in the successful demonstration of equivalence. If no differences are observed between the groups this could signal two possibilities:
(i) the dyslexic subjects show no deficit in naming which would challenge the notion that this particular dyslexic characteristic persists into adulthood;
(ii) the dyslexic subjects are not naming but still passing the equivalence test suggesting that naming is not always necessary for the emergence of stimulus equivalence which would challenge the assumption that this is the route by which verbally competent human subjects demonstrate equivalence. The dyslexic subjects could be using an alternative strategy (that is remembering the relations visually) which is just as effective as the control subjects' strategy. Alternatively, the performance could be simply controlled by the features of the stimulus-stimulus relations (as predicted by) overriding any need for the subjects to name.

# Chapter Four

## Experiment 1

Background

Method

Results

Discussion

## Experiment 2

Background

Method

Results

Discussion

## Experiment 3

Background

Method

Results

Discussion

## General Discussion

## Experiment 1

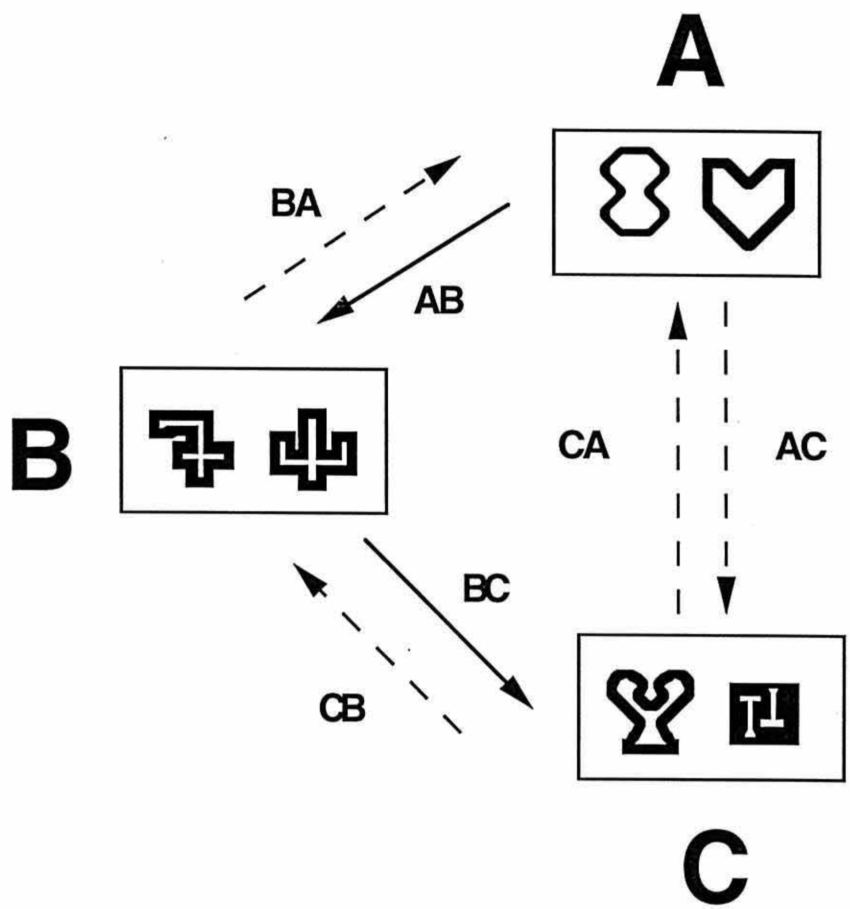
### Background

In order to demonstrate any differences in performance between dyslexic and non-dyslexic adults, the simplest matching-to-sample protocol was adopted, in the first instance. This involved the presentation of one sample stimulus and two comparison stimuli. There was no previous literature on which to consult, so employing just two equivalence classes was deemed to be an appropriate starting point as there was no means of predicting how difficult the dyslexic subjects would find the task.

Novel, abstract, pictorial stimuli were employed in this study. In experiments involving adult subjects familiar stimuli present extraneous variables such as them having strong preestablished meanings and associations, which could lend themselves readily to being grouped subjectively rather than according to equivalence classes.

Subjects were to be presented with a standard matching-to-sample visual-visual task (as outlined in Chapter 1) which consisted of the training of two AB and BC conditional relations, and then the testing of: symmetry (BA and CB), transitivity (AC), and equivalence (CA); all of which should emerge without further training (see Figure 5).
It was hypothesised in Chapter 3, that if the subjects are assigning names to the experimental stimuli then the naming differences between the dyslexic and non-dyslexic groups would be reflected in differences in performance in the

Figure 5: The stimulus equivalence paradigm utilised in Experiment 1.
Solid lines denote trained relations.
Dotted lines denote emergent relations.


A
BA
AB
B
CA
AC
BC
CB
C

training phases (although naming is not necessary for a successful performance here it was argued that naming may facilitate subjects' learning), and on the test session trials, where according to the naming hypothesis, the subjects who successfully demonstrate equivalence should be naming.

The groups were to be compared on the following measures:
(i) the number of errors needed to reach criterion during training,
(ii) the maintenance of this learning,
(iii) amount of exposure to the training relations required by the subjects,
(iv) the number of errors made on the test session,
(v) the test session latencies to directly assess whether the dyslexic naming deficit was reflected here. If the dyslexic subjects take longer to produce the names of the stimuli, this should, according to the naming hypothesis, lead to longer test trial latencies during the tests of symmetry, transitivity and equivalence.
The latter two measures are to be employed using subjects who passed the tests of equivalence as it is these subjects who, according to the naming hypothesis, should be naming.

Within-subjects effects were also to be considered in order to investigate whether examining subjects' response latencies on the equivalence test can determine anything useful concerning subjects' underlying strategies. In line with Bentall et al (1993), if subjects assign a common name to the stimuli, then no significant differences should be observed between the various test session tasks. If the tasks involving transitivity (those being AC transitivity, and CA equivalence) take longer, then this indicates that something other than common naming is controlling the subject's behaviour. Bentall et al suggest that this pattern can be explained by Field's, 'Associative Distance Effect'. However, this pattern of responding is not incompatible with the hypothesis that the subjects are assigning individual names to the stimuli (see Chapter 3).

If either of these patterns are observed in the latencies of the subjects who pass the test of equivalence then there would be evidence to suggest that these patterns are reliably observed when the above strategies are implemented. If not, it would question whether there is any value in interpreting response latencies in this way.

A further important function of this first experiment was to serve as a pilot study which would detect trends in the data and thus guide future experiments. So, it was proposed that the data were to be analysed on an individual as well as group level. It was reasoned that taking group measures of central tendency may mask subtle individual differences in the data which, although not statistically analysable, may point to valuable trends which are worth pursuing in more detail.

## Method

**Subjects** (see Appendix A for full subject details)

Eighteen subjects took part in this experiment (nine in the dyslexic group and nine in the non-dyslexic group). Each group consisted of seven females and two males. The age range for the dyslexic group was 21-32 years (mean age = 25.89 years); and for the non-dyslexic group 18-24 years (mean age = 21.89 years). This mean age difference of four years, was considered acceptable because all the subjects were adults and were deemed to be developmentally stable.

All subjects were undergraduate students from various departments at the University of Wales, Bangor. Participation was voluntary, although five subjects (two dyslexic and three non-dyslexic) were psychology undergraduates who received course credits for taking part in departmental experiments. This, however, at the time was not a compulsory course requirement. No subject possessed prior knowledge of stimulus equivalence theories or methodology.

Subject Selection: The dyslexic subjects were recruited via the Dyslexia Unit where they were members of a university support group. All subjects had been previously diagnosed as being dyslexic either in Bangor or elsewhere during their educational careers.

For the purposes of these studies a standard definition of developmental dyslexia was also adhered to. As discussed in chapter Two this is widely accepted to be the observation of a 'mismatch' of symptoms (Miles 1994b); more specifically high or average intelligence coupled with below average reading and spelling (Miles, 1993 p16). As Critchley and Critchley (1978 - cited in Critchley 1981) write:

> Developmental Dyslexia is a learning disability which initially shows itself by difficulty in learning to read, and later by erratic spelling and by lack of facility in manipulating written as opposed to spoken words. It is not due to intellectual inadequacy or to lack of socio-cultural opportunity, or to emotional factors, or to any known structural brain-defect.
>
> (p.1)

In accordance with the above definition all the subjects in this study were matched for intellectual attainment and socio-cultural opportunity by virtue of the fact that they were all studying on degree courses at the University of Wales (it was therefore safe to assume that they were all of above average intelligence). None reported to have suffered any known emotional or neurological deficiencies.

For five out of the nine subjects their diagnosis of dyslexia had been confirmed by staff associated with the Dyslexia Unit or elsewhere. Such assessments included: an intelligence test; a reading test; a spelling test; and in some cases the Bangor Dyslexia Test (Miles, 1982). The remaining four subjects who had not recently been formally assessed were tested pre-experimentally using the Bangor Dyslexia Test (Miles, 1982) and the last thirty items of the Schonell S1 spelling test (Schonell and Schonell, 1952).

The Bangor Dyslexia Test (Miles, 1982) is an orally administered test comprising of ten items/dyslexia indicators on which a subject is scored positive, neutral or negative according to standardised criteria in the test handbook. Scores are noted down by the experimenter on a formal response sheet which also includes experimental instructions. The more positives that are observed the more likely it is that the individual possesses dyslexic tendencies (see Miles,1993, for fuller interpretation of the test scores).

The Schonell spelling test comprises of a list of printed words which the experimenter reads out aloud to the subject. The last thirty words only were selected because these were deemed to be the most pertinent in discriminating between adult subjects of average or above average intelligence. The whole Schonell test can be used with children from five years upwards, so some of the items would produce ceiling scores for adult subjects. Each subject was required to record her/his responses on a blank sheet of paper. There was no time restraint, the subject merely indicated to the experimenter when she/he wished to move on. Some words were semantically confusable (such as *colonel* and *coarse*). In these instances the experimenter also provided a short definition (for example, *"colonel - as in the army"* or *"coarse - as in rough"*).

As time was limited in this first study these two tasks were chosen because they were both efficient and pertinent indicators of dyslexia. Miles (1993, 1994b) presents a detailed validation of the use of these particular diagnostics measures. With regard to the Bangor Dyslexia Test, Miles (1993) presents data from a large scale study which compared the performances of dyslexic and non-dyslexic groups on the task. A sample of 31 subjects aged 13-18 years were tested; the dyslexic group scored a mean of 4.87 positives; and the control group a mean of 2.05 positives. This difference was highly significant, suggesting that this test can discriminate between samples of dyslexic and non-dyslexic individual. Typically studies of childhood dyslexia look for discrepancies between intelligence and reading/spelling scores. The decision to rely on just a spelling test rather than a reading test here with dyslexic adults was again validated by data presented in Miles (1993). He reports the scores obtained by 48 dyslexic adults on various diagnostic measures. It was found that many of the subjects obtained high scores on the Schonell R1 Word Recognition test, in fact, 20/41 reported scores were over 90 which is the expected norm for 18 year olds. In contrast, only 4/40 dyslexic subjects reached the expected score on the Schonell spelling test.

Even though a major defining characteristic of childhood dyslexia is poor reading, the same yardstick cannot be employed as readily in studies involving dyslexic adults who have had the benefit of years of practice to overcome their reading problem. The same could be argued for spelling but the above data suggests that it is less resilient to compensation and therefore will serve as a more reliable diagnostic criterion. Miles (1994b) reports, "Practitioners know that even when dyslexics have learned to read adequately they almost always continue to have a spelling problem" (p.74).

The resultant dyslexic criteria employed in this study was as follows: a dyslexic subject must score 5 or more dyslexia positives on the Bangor Dyslexia Test or 4 positives plus under

20/30 on the Schonell spelling test. All four dyslexic subjects scored 4+ves and above on the Bangor Dyslexia Test and all scored under 20 out of 30 correct on the spelling test.
All members of the non-dyslexic group were given the two measures too in order to verify that none possessed dyslexic tendencies unbeknown to them. All nine control subjects scored 2.5+ves or below on the Bangor Dyslexia test (mean=1.72+ves). The mean spelling score for the control group was 22.56. Three subjects (Control 5, Control 8 and Control 9) produced scores of below 20 but these were coupled with scores of 1+ve, 2.5+ves and 2.5+ves respectively on the Bangor Dyslexia test, so the low spelling scores were put down solely to bad spelling and nothing else.

All subjects underwent a simple reaction time task. This involved subjects responding to a single stimulus (a cross) presented in the centre of the computer screen, by hitting a preassigned key (the spacebar) on the keyboard, as soon as they saw the cross appear. This task was presented on an AppleMac SE computer using a programme written on PsychLab (v 0.85 Miller, Dube 1988 - timing resolution 16.63ms). Subjects' responses were collated in a text file stored on the computer's hard disc.

This was considered to be an important measure as one of the experimental dependent variables was response latencies. This was an attempt to verify that neither of the two groups contained any remarkably slow or fast responders. Subjects' responses were measured on 10 trials and a median reaction time was calculated for each subject. Median latencies were used to eliminate any outliers or extreme values which often characterise reaction time measures. The mean of the medians for the control group was 262.72ms and for the dyslexic group 290.06ms. An unpaired t-test was performed on these data and no significant difference was found between the two groups on this measure (t=-0.992, df=16, ns).

Once it was assured that all subjects were either dyslexic or non-dyslexic according to the above criteria and that they did not respond overly fast/slow on the simple reaction time test, they were included in the study.

A pre-experimental questionnaire was given to each subject. This concerned details which may have been of relevance to the subjects' performances on the experimental task (specifically; age, gender, psychology background, occupation, and bilingualism).

**Apparatus and Materials**

The study took place in one of the School's research rooms which was quiet and free from distractions.

Experimental Task

The Matching-to-Sample task was presented, and responses recorded, on an AppleMac SE computer fitted with a touchscreen. The MTS v 9.32 programme generated the experimental trials (timing resolution = 1/60 seconds, rounded down to 1/100 in textfile; supplied by W.V.Dube, E.K.Shriver Centre, Massachusetts. May 1993).

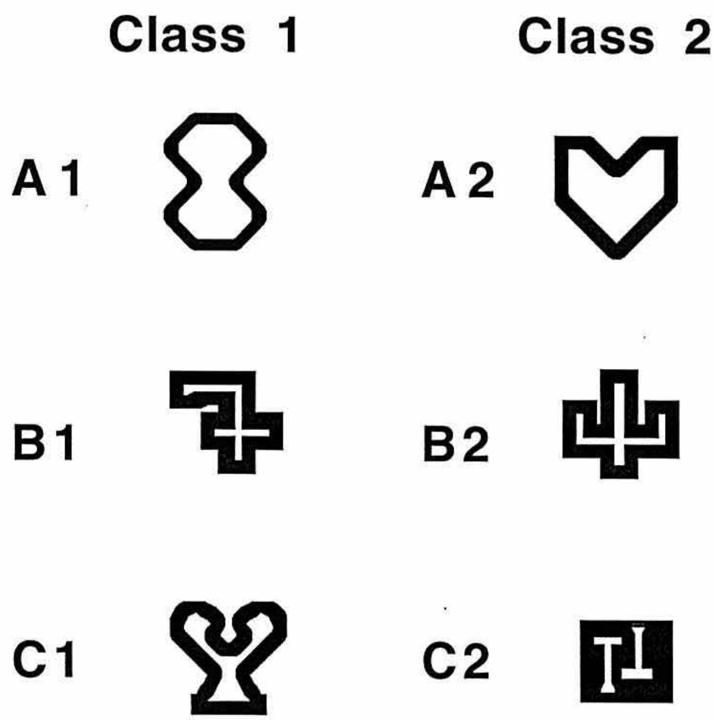
The experimental stimuli were chosen from the MTS shapes font, size 72 point. In selecting the stimuli, the experimenter tried to avoid those that represented 'real-life' objects and those that did not explicitly fall into groups (see Figure 6). The arbitrariness of the stimuli was confirmed prior to the actual running of the experiment. Six independent volunteers were chosen at random and presented with the six stimuli (printed in black on the centre of white cards, each measuring 3 inches by 2.5 inches). They were requested to sort the cards into two groups of three. None of the volunteers grouped the stimuli in the same way as they were to be grouped for the purpose of the experiment. This indicated that the stimulus classes chosen did not go intuitively together.

The subjects' responses were collated by the computer and stored in text files.

Post-experimental Task

A post-experimental questionnaire (see Appendix B) was presented to each subject which contained questions

**Figure 6:** **The six experimental stimuli arranged in two equivalence classes.**


Class 1
Class 2
A1
A2
B1
B2
C1
C2

concerning the subject's perceptions of the experiment, any strategies used during the task, and any names assigned to the stimuli.

## Experimental Design

This study employed a mixed design as it incorporated a *between-subjects* design (that is, comparing dyslexic and non-dyslexic groups of subjects) and a *within-subjects* design (each subject was measured on six tasks during the test session: AB/BC baseline tasks; BA/CB symmetry tasks; and AC transitivity/CA transitivity with symmetry or equivalence task). Such a design ensured that subjects in each group could be compared on the dependent variables (number of errors made and response latencies) and on any interaction between these and the two independent variables (groups and test task). To maintain experimental control, the two populations comprised of volunteers and the experimenter did not purposely pick out any subjects. The only subject selection which took place concerned the operational definition of dyslexia and the verification that the two groups did not differ on a simple reaction time measure (see Subjects section). To counteract any effects of practice or priming, trials were 'pseudo-randomly' presented, in the sense that they were random with the only criteria being that no more than three trials of the same trials type (for instance, A1-B1, A2-B2 and so on) could appear concurrently. The position of the correct comparison stimulus (right or left) was randomised in the same manner. Every subject received the same order of trials.

**Procedure**

The subjects came into the experimental room for a period of between one and one and half hours on average. Breaks were permitted between each experimental phase. The experimenter was present at all times.

Experimental Task

This was a standard matching-to-sample task involving visual stimuli.

Phase 1 - Train AB and BC Relations

Subjects were trained on the AB and BC relations simultaneously (see Figure 7). This session consisted of 48 trials broken down as follows:

12 A1-B1 (incorrect comparison B2)
12 A2-B2 (incorrect comparison B1)
12 B1-C1 (incorrect comparison C2)
12 B2-C2 (incorrect comparison C1)

These were presented in a pseudo-random order (see Experimental Design section for constraints).

A sample appeared in the top half of the screen, and when the subject touched it, two comparison stimuli appeared in the bottom half of the screen (see Figure 8).

The subject was required to select one of the comparisons by touching. In this first training phase reinforcement was given. This consisted of the screen flashing and the computer emitting an electronic noise (which lasted approximately 3000ms) if the subject responded correctly. If a subject selected an incorrect comparison there was no sound and the next trial was presented.

The experimenter read the following instructions to each

Figure 7: Examples of the training trials, where a sample stimulus is presented followed by two comparison stimuli. The black solid arrows denote the correct relations (for example, A1-B1, A2-B2). The grey arrows illustrate the incorrect choice.

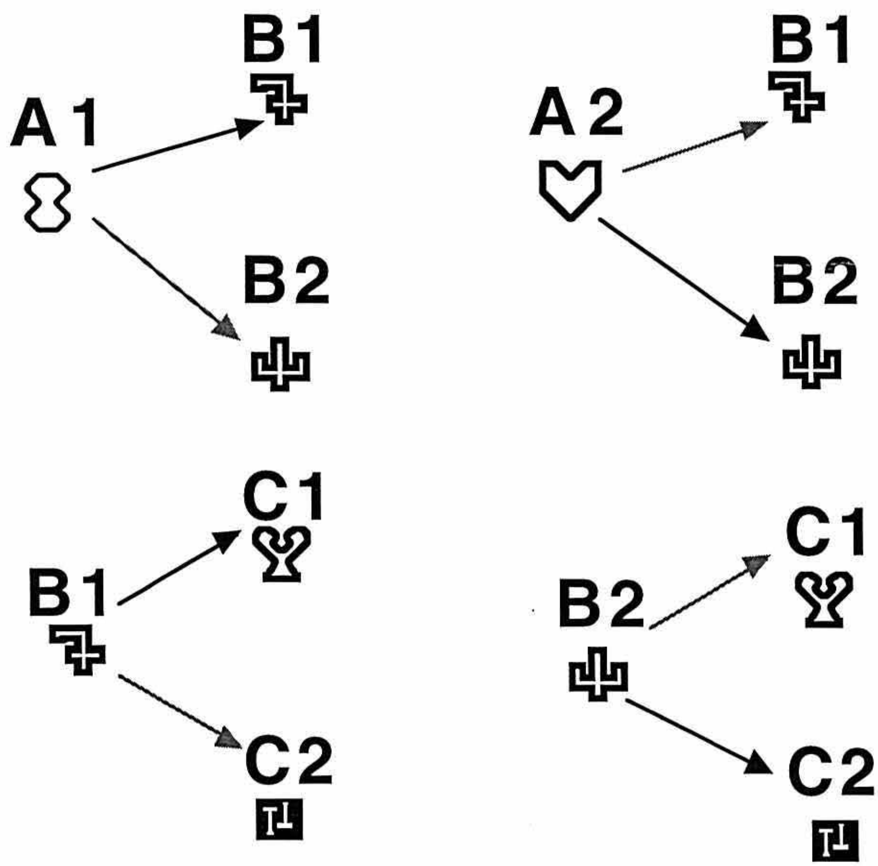
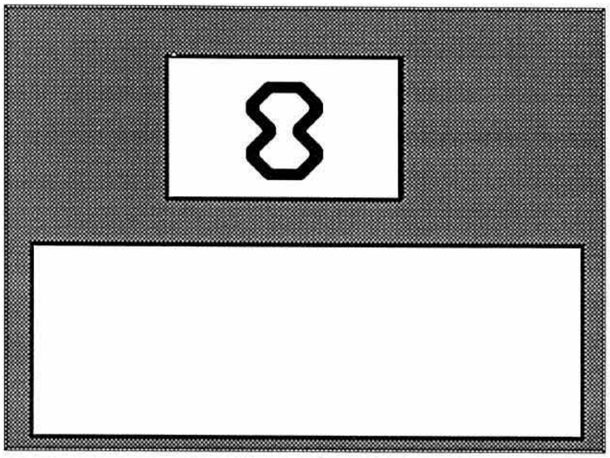
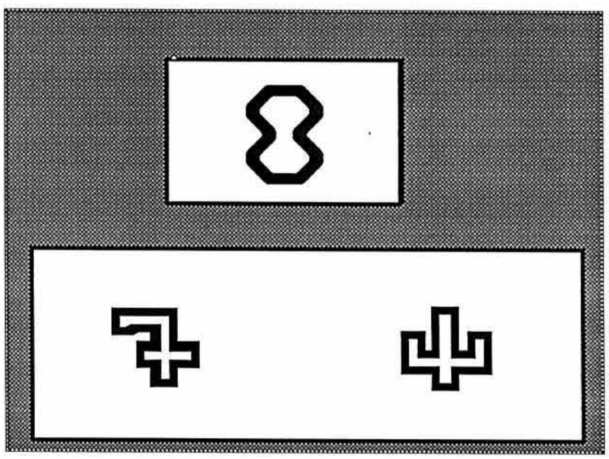

A1
B1
B2
A2
B1
B2
B1
C1
C2
B2
C1
C2

**Figure 8:** **An illustration of how each trial is presented on the computer screen.**



Sample
A1



Sample
A1

Comparisons
B1 and B2

subject:

> In front of you is a computer screen. When the experiment begins a visual image will appear in the centre of the screen. In order to carry on you must press the screen right in the centre of the stimulus. Two more images will appear towards the left and right corners of the screen. When these appear please press either of the images. If you press the correct image the computer will make a noise and the screen will flash. If you press the incorrect image nothing will happen and the next trial will begin. Feel free to talk aloud as you go along; it may be helpful. At the end of the session I will tell you how well you have done. Please try to get as many correct as you can. Do you understand or do you want me to repeat the instructions?

Training sessions continued until the subject made no more than two errors in the last twenty-four trials (i.e. approximately 90% correct).

Phase 2: AB and BC trials without reinforcement

Once the subject had learned the AB/BC relations to criterion, she/he progressed to the second phase. The same 48 AB and BC trials were presented (in a different random order) but this time no reinforcement was given for correct responses. This was in preparation for the test session and it served to assess the retention of learning and whether the trained relations could be maintained without reinforcement. The following instructions were given:

> The procedure is exactly the same as before but this time the computer will not tell you whether your response is right or wrong. I will tell you at the end how well you have done. As before try and make as few mistakes as possible.

Phase 3 - Test Symmetry (BA and CB), Transitivity (AC) and Equivalence (CA)

The subjects were tested on the following tasks: AB training (8 trials; 4 A1-B1, 4 A2-B2); BC training (8 trials; 4 B1-C1, 4 B2-C2); BA symmetry (8 trials; 4 B1-A1, 4 B2-A2); CB symmetry (8 trials; 4 C1-B1, 4 C2-B2); AC transitivity (8 trials; 4 A1-C1, 4 A2-C2); CA equivalence (8 trials; 4 C1-A1, 4 C2-A2).

The above trials were presented randomly with no reinforcement. The instructions were the same as given in the previous phase, so the subjects were unaware at the outset that novel stimulus combinations would be presented. Subjects responses (including latencies and errors) for all training and test sessions were recorded by the computer.

Post-experimental Measures

Each subject was asked to complete a short questionnaire concerning their perceptions of the experiment, any strategies used during the sessions, and any names assigned to the stimuli. The experimenter was aware that verbal reports can prove to be unreliable. It could be the first time the subjects have thought about the stimuli and under the pressure of the situation they may feel obliged to formulate some kind of strategy for the experimenter's sake. This could be completely different to what had actually occurred during the experiment. However, this was the only means available to assess the subject's thoughts on the experiment so, although it is unwise to rely solely on post-experimental reports in conjunction with the other measures they could provide useful data for shaping subsequent experiments.

After completing the questionnaire, subjects were told how well they did on the MTS tasks. The experimenter explained the aims of the experiment and subjects were encouraged to ask questions about the procedure and their results, or make any comments.

## Results

**Training - Phases 1 and 2**

**Table 1:** The mean number of errors per group during: Phase 1,train AB and BC; Phase 2, AB and BC without reinforcement; and the mean number of repeat training sessions required by each group. Standard deviations are given in brackets.

| | Phase 1 Number of Errors to Criterion | Phase 2 Unreinforced Errors | Repeat Number of Sessions |
|---|---|---|---|
| Dyslexic Group | 29.44 (17.33) | 7.11 (9.65) | 1.44 (1.13) |
| Control Group | 19.89 (18.16) | 0.89 (1.05) | 1.22 (1.20) |

The first question to be addressed was whether there was a difference between the number of errors to criterion made by each group in Phase 1. For this purpose the total number of errors on the reinforced trials (Phase 1, AB and BC with reinforcement) was calculated for each subject to assess learning rate. Table 1 presents the group means and it can be seen that the dyslexic group made more errors than the control group. However, an unpaired t-test revealed that this difference was not significant (t=1.142, df=16, ns).

There may have been no significant difference between the groups on their overall number of errors made but perhaps the removal of feedback during the unreinforced trials in Phase 2 affected the dyslexic group more than the control group (as

outlined in Chapter 3). The number of errors made by each subject during Phase 2 was isolated. Looking at the group means presented in Table 1 it can be seen that the dyslexic group made on average 7 errors during this phase and the control group less than one error. It was observed that the dyslexic group's data demonstrated significantly more variance as measured by an F-test (F=83.80, num.df=8, den.df=8, P<0.0001). To account for this irregular distribution the non-parametric Robust Rank-Order Test (see Siegel and Castellan 1988) was undertaken and a second non-significant difference between the groups was demonstrated (U`=1.32, n, m=9, ns). So, using statistical analysis no significant difference was observed between the groups despite the fact that the dyslexic group made seemingly more errors. On returning to the raw data it can be seen that the extreme difference in variability can be accounted for by three dyslexic subjects (Dyslexics 5, 8 and 9) who made 11, 23 and 23 errors respectively over the 48 trial unreinforced session, all of which were found on AB trials. All the remaining subjects (from both groups) made between 0 and 4 errors.
In other words a third of the dyslexic sample made more errors than any of the other subjects which is not enough to produce a statistically significant group difference but which is a trend found only on the dyslexic group.

The final issue to be investigated was whether the groups differed on the number of repeat training sessions required (the minimum number of sessions was 2, so this was a measure of how many sessions above this number subjects required). This was included as it was thought to measure something different from the number of errors made. It measures how much exposure each subject has to the reinforcement contingencies prior to testing, which is not evident from the two error measures. For example, two subjects could receive the minimum of two training sessions (one in Phase 1 with reinforcement and one in Phase 2 without reinforcement) but make differing numbers of errors; one could make a few errors and the other make many more but still reach the criterion, making the number of errors recorded high but the number of training sessions needed low. Conversely, one subject could make relatively few errors but miss the criterion and have to repeat the session. This would result in the total number of errors being low and the subject receiving extra exposure to the baseline relations.

The total number of repeat training sessions required by each subject was calculated and it was found that fourteen out of the eighteen subjects (seven control and seven dyslexic) required one or more extra sessions. Table 1 illustrates the mean number required by each group and it can be seen that the difference is minimal. An unpaired t-test confirmed that there was no significant difference between the groups on this measure ($t=0.404$, $df=16$, ns). So, both groups received equal amounts of exposure to the reinforcement contingencies.

**Test Session**

The dependent variables being measured were the number of errors made by each subject during the test session and the time taken to match the stimulus pairs (sample and comparison) on the various tasks. For the purpose of simplicity of analysis the tasks were collapsed into the following categories: *baseline* ( AB and BC tasks); *symmetry* ( BA and CB tasks); and *transitivity* ( AC transitivity and CA equivalence). The AC transitivity task and CA equivalence task were combined as there is no reason to believe that performances on these two tasks would significantly differ (see Bentall et al, 1993).

Errors made during the test session

The pass/fail rate criterion adopted stated that a subject could make no more than three errors per category (as defined above). Put another way, each subject was required to produce at least 14/16 correct responses on each of the above categories. This meant that each subject must be correct on 87.5% of the trials. As each trial comprised of a choice between just two comparison stimuli, the level of responding correctly by chance alone would be 50%, thus the criterion level was set well above this chance level.

**Table 2:** Pass/Fail rate for each group

| | **Pass** | **Fail** |
|---|---|---|
| **Dyslexic Group** | 4 | 5 |
| **Control Group** | 6 | 3 |

Table 2 illustrates the number of subjects from each group who passed or failed the test session in accordance with the above criterion. To ascertain whether there was a statistically significant difference between the two groups' performances a Fisher's Exact test was applied to the data. These data are categorical (in the sense that they represent frequency counts) and the Fisher's test was deemed to be more suitable to identify significant effects in small sample sizes. The difference between the two groups was non-significant (phi=0.224, df=1, ns).

Due to the strict pass/fail criterion imposed, the subjects who passed made very few errors. The dyslexic group made no more than a mean of 5% errors on any category and the control group no more than a mean of 3% errors. This meant that the differences between the errors made on each of the categories was minimal, so any statistical analysis of these data would serve no useful purpose.

Examining the data of the subjects who failed could prove to be more fruitful. For example, even though no significant difference was detected between the pass/fail rate of the two groups there could still be a difference between the number of errors each group made. Due to the fact the pass/fail rate is arbitrarily imposed one group could make a large amount of errors and thus fail badly but the other make relatively few errors and only just fail.

There is also a need to see how the errors were distributed across the various tasks. So these data were considered on an individual basis in order to detect any further trends. The first issue to be addressed was whether the group analysis obscured any important differences between dyslexic and control individuals.

In the dyslexic group 5/9 subjects failed to respond on the test trials in accordance with the relational properties of transitivity. In other words, they failed to form equivalence classes from the two stimulus sets. Despite being taught the AB and BC relations they did not link the stimuli as follows: A1-B1 and B1-C1, so A1-C1. Of the control subjects only two subjects (Controls 8 and 9 respectively) failed in this manner.

One subject, Control 7, produced 0% correct on the trials involving transitivity. This consistent responding suggests that she was demonstrating equivalence by exclusion (S-minus control) where it is the incorrect comparison which controls responding (see also Sidman 1987). This type of responding is observed only in protocols where there are only two comparison stimuli. In other words, when presented with an A1 sample the response is *not* to choose B2 but to select the alternative comparison regardless of what that might be. In this way, responding according to the properties of transitivity would proceed as follows: A1 *not* B2, B2 *not* C1, therefore, A1 *not* C1 which would result in the subject scoring 0% correct. Or alternatively, A2 *not* B1, B1 *not* C2, therefore, A2 *not* C2, again resulting in 0% correct on the test of transitivity.

This raises the issue of what should or should not be considered to be a pass. According to the Sidman definition of equivalence, Control 7 failed to demonstrate equivalence for the experimentally defined relations. However, as outlined in Chapter 1, this viewpoint maybe limited. Saunders and Green

(1992) reappraise the mathematical definition of equivalence and suggest that examples of S-minus control should be classed as evidence of equivalence. Therefore, Control 7, should strictly speaking also be considered to have passed the equivalence test.

This changes the original pass/fail rate. Previously, 4 dyslexic and 6 control subjects passed and 5 dyslexic and 3 control subjects failed. Now, although the dyslexic group's figures remain the same, 7 control subjects can be said to have demonstrated equivalence and only 2 failed. Therefore, there is a much greater difference between the groups with many more dyslexic subjects failing. A Fisher's test was performed on these reappraised data and again no significant difference was demonstrated (phi=0.342, df=1, ns). It should be noted that this test although suited to small sample sizes, is not a very sensitive one. In order for differences between the groups to be significant values of:

| | |
|---|---|
| 7 | 2 |
| 2 | 7 |

are needed.

Another question which arises is whether the distribution of errors changes at all during the test session. In other words, do subjects make less errors on the latter half of the session? Researchers (such as Bush, Sidman and DeRose, 1989; Lazar, Davis-Lang and Sanchez, 1984; and Sidman, Kirk and Willson-Morris, 1985) have reported instances of *delayed emergence* where equivalence has emerged only after repeated testing, despite the absence of reinforcement. Therefore, it is possible that group differences in equivalence might be revealed in the latter stages of the session rather than at the beginning. It could be the case that the control subjects who failed overall had passed by the second half of testing, whereas the dyslexic subjects who failed did so throughout.

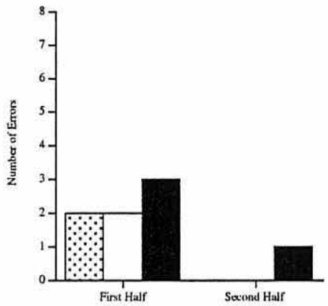
Figure 9 illustrates the number of errors made by each subject who failed to demonstrate equivalence. Each graph is divided into first and second half of the session. The vertical axis is numbered up to eight as this is the total number of trials for each category on one half of the session. This was calculated as follows: the first half comprises of 24 trials (the first 8 baseline trials, the first 8 symmetry trials, and the first 8 transitivity trials); the second half again consists of 24 trials (the last 8 baseline trials, the last 8 symmetry trials, and the last 8 transitivity trials).

The most notable improvements in performance were observed for Dyslexic 5 and Dyslexic 6 who both made minimal numbers of errors on the second half of the session. Control subjects 8 and 9 similarly only demonstrated marginal fails on the second half. Therefore, such improvements were observed in equal numbers in both groups. The three remaining dyslexic subjects 7, 8, and 9 who failed still performed well below criterion on the second half of the test session whereas the 2 control subjects who failed (Control 8 ands Control 9) were very close to reaching the criterion by this point in the session. Therefore, there is a trend for more dyslexic subjects to fail persistently.

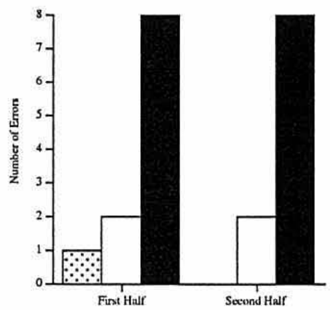
**Figure 9:** **Graphs illustrating the number of errors made by the subjects who failed on each category, on each half of the test session in Experiment 1.**

9

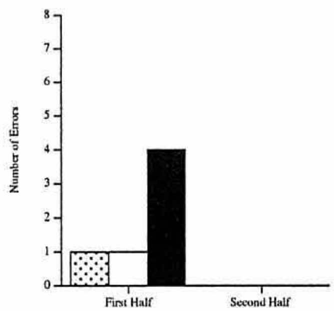
**Dyslexic 5**



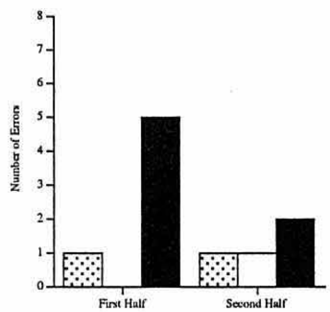

**Control 7**



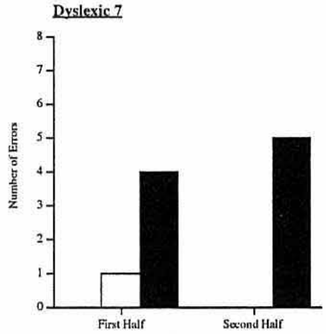

**Dyslexic 6**



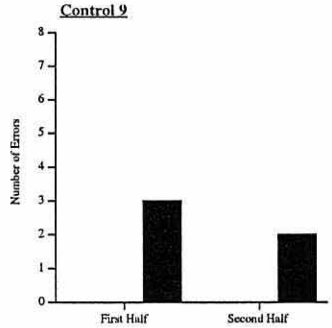

**Control 8**



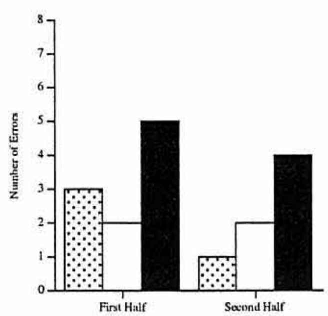

**Dyslexic 7**




**Control 9**




**Dyslexic 8**



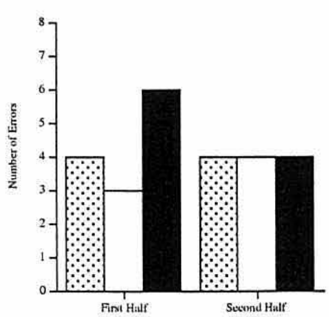

**Key**

Baseline

Symmetry

Transitivity

**Dyslexic 9**



Two dyslexic subjects (Dyslexic 8 and Dyslexic 9) showed evidence of baseline deterioration. However, Dyslexic 9's baseline performance did reach criterion on the second half of the test session yet she still failed. Dyslexic 9 produced responding which on the surface looked like baseline deterioration. She made four baseline errors on each half of the session, but she did respond consistently (be it erroneously) to some degree. She made all her errors because she mixed up the AB relations, in that she matched A1-B2 and A2-B1. If these relations were also equivalence relations her responding should have been as follows: A1-B2, B2-C2, therefore, A1-C2, resulting in 0% correct on the tests of transitivity. However, this was not found to be the case (as it was with Control 7) so she did not demonstrate equivalence despite consistent baseline responding. (It should be noted that it was Dyslexics 8 and 9 who demonstrated poor AB performances during the unreinforced training phase).

In this way it can be said that all the subjects who persisted in failing did so despite the fact that their baseline performances were maintained and thus in theory capable of supporting transitive responding and there were more dyslexic subjects than control subjects who were found to continuously fail throughout the session.

Test trial response latencies

The next issue to address was whether the groups differed in the amount of time taken to match the correct sample and comparison stimuli. The latency between the presentation of the sample stimulus and the subject choosing a comparison was recorded for each trial and a median response latency calculated for each test category. The median was chosen as being the most appropriate measure of central tendency as it remains unaffected by outlier responses which tend to characterise reaction time studies.

Only subjects who successfully demonstrated stimulus equivalence according to the criterion were included in the analysis. According to the Naming Hypothesis (as outlined in Chapter one) in order to successfully demonstrate stimulus equivalence subjects must name the stimuli (subjects who fail are either not naming or naming inappropriately). As a result, any response latency difference (which it is hypothesised may reflect any underlying naming latency differences) should be observed only in the performances of subjects who pass the test session. Although it can be seen from the error analysis that Control 7 responded in accordance with the properties of transitivity her data should not be included in this analysis as her strategy was fundamentally different from all the others (her responding was controlled by the S-minus comparison stimulus).

This reduces the sample sizes somewhat (4 dyslexic subjects and 6 control subjects) so that they fall below the traditionally recommended numbers for conducting parametric analyses and more importantly for detecting reliable significant effects. However, a 2X3 mixed design ANOVA was still applied to the data as this was the only analysis available by which between and within-subject group comparisons could be made. Any results can only be cautiously interpreted and an attempt was made to look for trends which if more data were available may suggest significant effects.

No group effect was demonstrated (F=0.394, df=1,8 ns), however, a significant difference was observed between the test categories (F=4.079, df=2,16 p<0.05). No significant interaction effect was observed (F=0.497, df=2,16 ns).

**Table 3:** The mean of the medians for each group on each of the test categories (in milliseconds). Standard deviations are given in brackets.

| | **Baseline** | **Symmetry** | **Transitivity** | **Total** |
|---|---|---|---|---|
| **Dyslexic Group** | 2635.00 (1704.12) | 2228.75 (1245.35) | 3841.25 (3777.75) | 2901.67 (2370.38) |
| **Control Group** | 2064.17 (854.83) | 2025.00 (443.25) | 2795.00 (1172.79) | 2294.72 (900.01) |
| **Total Mean** | 2292.50 (382.21) | 2106.50 (798.24) | 3213.50 (2411.05) | |

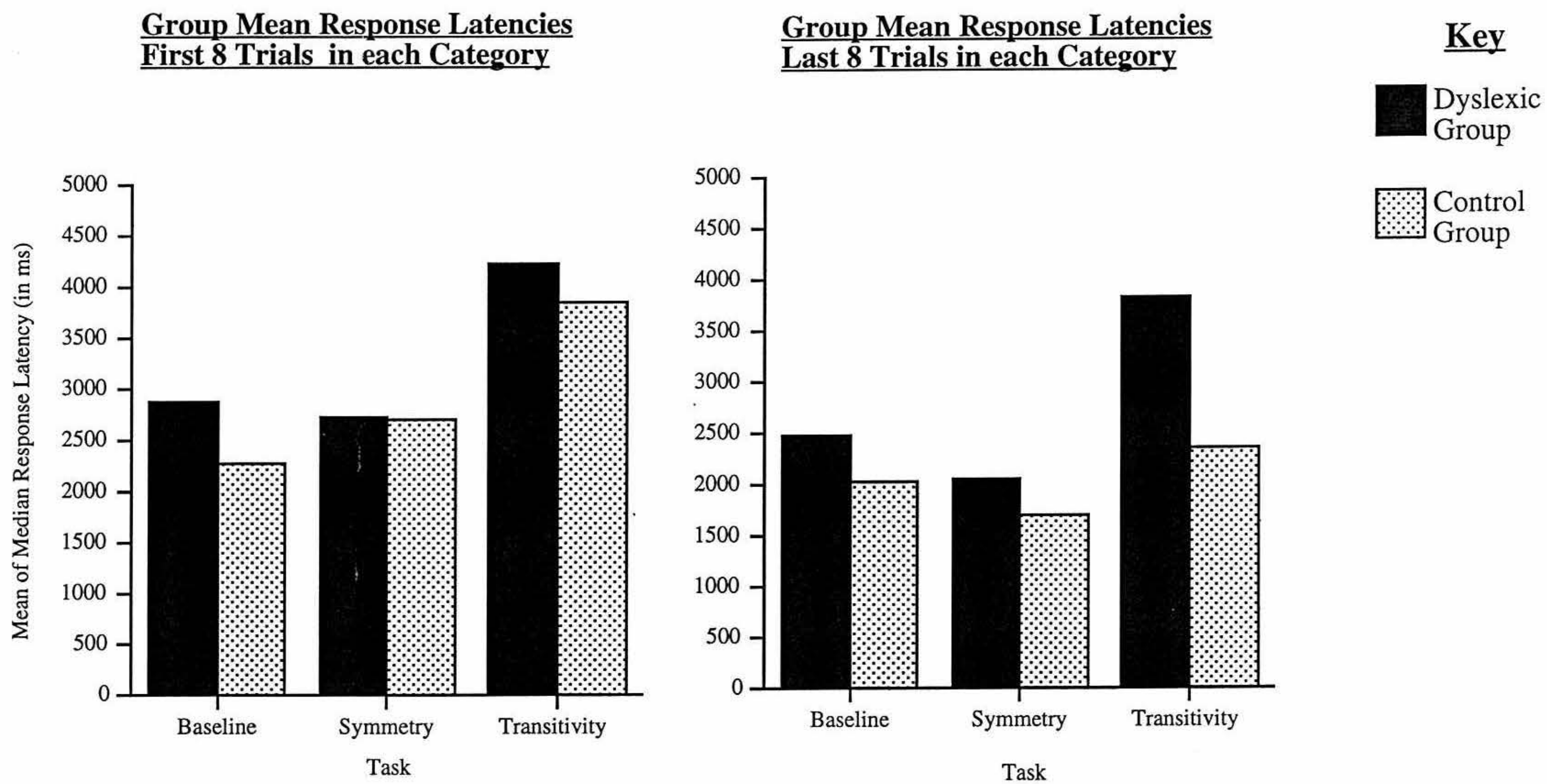
Looking at the means of medians in Table 3 it can be seen that the control group produced shorter latencies throughout although this difference was not statistically significant. The largest group difference was found on the tasks involving transitivity with the dyslexic group producing a mean which was 1046.25ms slower than the control mean. From the raw data it was noted that the dyslexic group's mean was inflated by the data from one subject, Dyslexic 4, who produced a transitivity latency of 9450ms which was much slower than all the others. If her data is omitted then the dyslexic transitivity mean becomes 1971.67ms which was <u>faster</u> than the control mean.

Both groups produced much longer latencies on the transitivity trials, the symmetrical trials were quicker than the baseline trials but this difference was only slight. To confirm where the significant within-subjects effect lay the category means were compared using Tukey's honestly significant difference (HSD) test. The mean difference between the symmetry tasks and the transitivity tasks was found to be the only comparison significantly different at the 0.05 level (q (3, 16) = 1071.23).

From the original ANOVA it was found that the two groups' latencies did not significantly differ on the test session. However, it could have been that they did differ on the first half of the session (specifically, when presented with the novel test combinations for the first time) but as the subjects became more practised the differences may have dropped on the latter half of the session leading to an overall non-significant latency difference.

To highlight such potential group differences graphs were drawn. Figure 10 presents both groups' mean of median response latencies on each category over the two halves of the session. Using visual inspection it can be seen that initially it appears that, contrary to the above prediction, the difference between the groups is greater on the second half, specifically with the dyslexic group taking longer on the transitivity task. However, this dyslexic transitivity mean (as noted earlier) was inflated by the extreme values of just one subject (Dyslexic 4). If this datum is omitted, the group difference

Figure 10: Graphs comparing each group's mean of median latencies on the first and second half of the test session in Experiment 1.



Group Mean Response Latencies
First 8 Trials in each Category
Mean of Median Response Latency (in ms)
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
Baseline
Symmetry
Transitivity
Task
Group Mean Response Latencies
Last 8 Trials in each Category
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
Baseline
Symmetry
Transitivity
Task
Key
Dyslexic Group
Control Group

diminishes to 398.34ms.

So, the differences between the two groups can be said to remain constant over both halves of the session.
Also, looking at Figure 10 it can be seen that the latencies, for both groups, decreased on the latter half of the session as would be expected as the subjects became more practised at responding to the stimulus combinations.

## Discussion

No significant group differences were found on any aspect of the training sessions. Despite the fact that the dyslexic group made noticeably more errors before reaching criterion than the control group, the difference between the groups was non-significant (t=1.142, df=16, ns). This indicates that the two groups mastered the baseline relations equally quickly. In addition, both groups received equal amounts of exposure to the reinforcement contingencies as there was no significant difference found between the groups on the number of repeat training sessions required.

In Chapter 3 it was hypothesised that the dyslexic subject's naming impairment might lead them to taking longer to learn the baseline relations because they name the stimuli inconsistently and thus make more rule-based errors. In contrast, if the control subjects consistently named the stimuli this should have a facilitative effect on their performance, lifting it above that of the dyslexic subjects. In this study no differences were found between the groups. This suggests that: (i) the dyslexic subjects do not display a deficit in naming and therefore, their performance is not appropriately impaired, and/or (ii) the dyslexic subjects are not naming but employing a strategy (that is, a nonverbal one) which is just as effective as that of the controls.
This could not be determined in this study as no direct measure of the subjects' naming ability was administered.

To consider the issue of how well the two groups retained the baseline matching relations, the number of errors made on the unreinforced trials were compared. Again no statistically significant difference was found between the two groups. However, it was observed that a third of the dyslexic subjects made many more errors than any of the other subjects. This was specific to the dyslexic sample and suggests a trend for more dyslexic subjects to be adversely affected by the removal of reinforcement.

So, the dyslexic subjects did not require more exposure to the training relations in order to achieve an error-free performance as predicted in Chapter 3. There was however, an indication that the dyslexic group was more reliant on feedback for a consistent performance, which could suggest that any names the subjects assigned to the stimuli were vulnerable to disruption. Or generally speaking, it could signify that dyslexic individuals are not as confident in their performance when feedback is withdrawn, whatever it entails.

Moving on to the data accrued from the test session, there was no significant difference in the pass/fail rate of the two groups. Even when the data was reanalysed (to include the the subject, Control 7 who responded accorded to the properties of transitivity) no significant difference was observed. However, a Fisher's test does not appear to be a sensitive one. So, although the difference between the groups is not significant, it could be concluded that there is a definite trend for more control subjects to pass. This would then support the hypothesis that the subjects are naming during the test session and that the dyslexic subjects are not naming consistently, given their phonological impairments. However, this supposition can be nothing more than tentative as there was no means of directly accessing what strategies the subjects employed throughout this session.

The error data of the subjects who failed to demonstrate equivalence were analysed on an individual basis (see Figure 9). Out of the five dyslexic subjects who failed the equivalence test, two (Dyslexics 5 and 6) reached criterion on the second half of the session. The three remaining subjects failed persistently across the session whilst responding consistently on the baseline relations meaning that in theory the relations were capable of supporting transitive responding.
Two control subjects (Controls 8 and 9) failed. A further subject, Control 7, failed according to the experimental criterion but she did demonstrate the properties of equivalence by exclusion, so she cannot, strictly speaking, be said to have failed. Dyslexic subjects 8 and 9 failed initially but had virtually attained the criterion by the second half of the session. These data (along with that of dyslexic subjects Peter and Kevin) support the notion of *delayed emergence* whereby equivalence can emerge as a result of repeated testing in the absence of feedback. This phenomenon, however, was observed equally in both the dyslexic and control samples.

Overall, from the error analysis it can be seen that there is a definite trend for more dyslexic subjects to persist in failing to demonstrate equivalence. Notably, it is in the sample which is characterised by a naming deficit where this persistent failing and more errors on the transitivity tasks are observed. Again this is tentative evidence to support the hypothesis that naming is related to the ability to successfully demonstrate equivalence and merits further investigation. It is tentative in that this evidence is only correlational, the only direct evidence with regard to the presence of naming on these trials comes from the latency analysis.

With regard to the test session latencies, no significant differences were observed between the groups on this measure. However, this analysis must be viewed with caution due to the small sample numbers employed. More specifically a non-significant result may indicate: (1) no difference between the groups; (2) or the presence of an effect which can not be detected because the sample numbers are so small. Therefore, it is necessary to note trends in the data which may reflect underlying effects.

The dyslexic group did appear to take longer on the transitivity tasks but these means were made up of only four data points, one of which (Dyslexic 4's) was an overly long latency for this category, with this data omitted the difference between the groups is minimised.

Therefore, when comparing the dyslexic and control subjects who successfully demonstrated equivalence there was no statistically significant difference on the test trial latencies, or on either half of the test session when visually inspecting Figure 10. This could indicate two possibilities: (1) the dyslexic subjects were not naming on these trials but still demonstrate stimulus equivalence which would challenge the assumptions of the naming hypothesis; (2) the dyslexic subjects were naming the stimuli but this does not affect their test trial latencies, possibly because there is no naming latency difference between the two groups. The hypothesis that there was a consistently observed naming deficiency in dyslexic adults was predicted from the existing literature. It was not empirically tested for using these samples of subjects.

The only evidence of whether the subjects were naming or not during the test session comes from the subjects' post-experimental questionnaires. Of the subjects who passed (and according to the naming hypothesis should be naming) eight out of the ten subjects reported that they had assigned names to at least three of the stimuli (which is all that was needed with this protocol as the other class can be formed by elimination).

The two other dyslexic subjects who passed reported that they remembered the stimuli, 'visually'. Dyslexic 2 named only one stimulus but said that one class all contained, 'curves'. Dyslexic 1 stated that she had named none of the stimuli, but put one class together because they were all, 'squarish'. So, although these two subjects were not assigning specific names to the stimuli, their descriptive labels could in fact function as common names which served to discriminate one class from the other. Similarly, Control 5 said that the stimuli in one class were all, "Ts"; again a common name. These data do support the naming hypothesis in that all the subjects who successfully demonstrated equivalence did name the stimuli in some way.

However, of the eight subjects who failed the test session, six reported that they too had assigned three or more names to the experimental stimuli. Only two subjects (Control 7, who actually 'passed' and Dyslexic 5) said that they did not use names to link the stimuli in any way. This would suggest that it is not sufficient just to assign names to the stimuli in order to demonstrate equivalence. It should be emphasised that the key factors, according to the naming hypothesis are to name consistently and appropriately according to the experimenter defined classes and to link the names relationally.

It must be noted that these data must be viewed with caution, in that it could, on one hand, indicate that the majority of subjects named the stimuli and yet some failed. On the other hand, this could be the first time the subjects have thought to name the stimuli, and only report doing so because it is being asked of them in the context of the post-experimental questionnaire. Therefore, subjects may not have been naming the stimuli during the actual session but still many of them passed, which could challenge the assumptions of the naming hypothesis.

The data were also considered as a whole (that is regardless of group). As reported in Chapter 3, Bentall et al (1993) predicted that a subject's pattern of responding (as predicted by their test trial latencies) may reflect their underlying naming strategy. They hypothesised that a flat pattern of responding reflects a common naming strategy. A pattern in which trials involving transitivity produced significantly longer response latencies and significantly more errors than baseline or symmetry trials was taken by the authors to indicate that another form of control was operating on behaviour, such as Field's and colleagues', "Associative Distance Effect". It was later argued that this latter pattern is indeed not incongruous with a naming hypothesis and may indicate that the stimuli are being named individually (see Dugdale and Lowe 1990; and Horne and Lowe 1996).

A significant difference was observed between the test session latencies on the various categories ($F=4.07$, $df=2,16$ $p<0.05$) with the tasks involving transitivity taking significantly longer than the symmetry tasks. The pattern found was not as clear cut as the one identified by Bentall et al (that is, baseline/symmetry > transitivity) as no significant difference was found here between the baseline and transitivity tasks (note that this difference only just failed to reach significance at the 0.05 level). Despite this, however, there is enough evidence to conclude that the trials which involved putting the relations together according to the properties of transitivity took longer, which is was what would be predicted if the subjects were assigning individual names to the stimuli. It must also be noted that the increased latencies are not due to increased errors because latency analysis was only performed using the subjects who passed, who, as noted earlier made a minimal number of errors.

Therefore, there exists a certain amount of correlational evidence (from the post-experimental remarks and pass/fail rate) to suggest that the subjects who pass are assigning names to the stimuli. Yet no significant group difference was observed between the groups on the test trial latencies, which would be predicted if the two samples significantly differed in their naming speed.

The logical conclusion to draw from this interpretation is that there is no significant difference in naming latencies between the dyslexic and non-dyslexic groups. This is, however, pure speculation as no empirical measure of naming speed was employed. However, it seems unlikely as the evidence from existing research (as cited in Chapter 2) suggests a persistent naming difficulty in dyslexic individuals and the dyslexic subjects in this study were seen to produce an inferior performance to the control subjects on various aspects of the experimental tasks (for example, on the unreinforced training session; and the pass/fail rate.

It is more probable that a non-significant difference between the groups on the test trial latencies signifies that the dyslexic subjects are not naming during the test session, which challenges the assumptions of the Naming Hypothesis. The post-experimental reports, that subjects who passed were assigning names to the stimuli, is not conclusive evidence that this is what they were doing during the actual experiment or that they named appropriately in order to demonstrate equivalence.

In conclusion, this first experiment provided much valuable evidence concerning the performance of dyslexic and non-dyslexic samples on the matching-to-sample task.

Procedurally, it was found that there is a need to include a measure of naming in the test battery in order to empirically determine whether a naming difference exists between the two samples. Also, there is a need to increase the number of comparisons presented to the subjects. This is to eliminate the possibility that the subjects responses are controlled by exclusion strategies (as argued by Sidman 1987) as was the case for (Control 7).

The stimuli utilised in this study were also deemed to be too nameable. In other words, were too easy to name, in particular for the dyslexic adult subjects whose performances involving familiar stimuli may have been compensated for over the years. If the stimuli were harder to name this may put the dyslexic subjects under more pressure and as a result more markedly disrupt their performances especially resulting in them producing longer latencies on the equivalence test trials. What can be gleaned from this study is that there was a trend for the dyslexic subjects to perform worse than the control subjects on the experimental task. They were more adversely affected by the removal of feedback. More dyslexic subjects persistently failed to demonstrate equivalence and there was a tendency for these subjects to make more transitivity errors. Given the procedural limitations of this study it can be tentatively concluded that there are indications that the sample of subjects which is characterised by a naming deficit did not perform as well on tests of stimulus equivalence. This points to a need for more empirical investigation into the relationship between naming and stimulus equivalence using such subjects.

## Experiment 2

### Background

From Experiment 1 a trend was seen for the dyslexic subjects to perform worse than the control subjects on certain aspects of the equivalence test. For instance, more control subjects passed the tests of equivalence and more dyslexic subjects persistently failed; and this is something which would be expected, by the naming hypothesis, if the subjects were not assigning names to the experimental stimuli or not naming the stimuli appropriately (that is according to the experimentally defined classes or by not naming the stimuli consistently). No significant differences, however, were demonstrated between the groups on the test session latencies. It was hypothesised that a significant difference would be observed on this dependent variable if the subjects were assigning names to the stimuli during this session as predicted by the Naming Hypothesis.

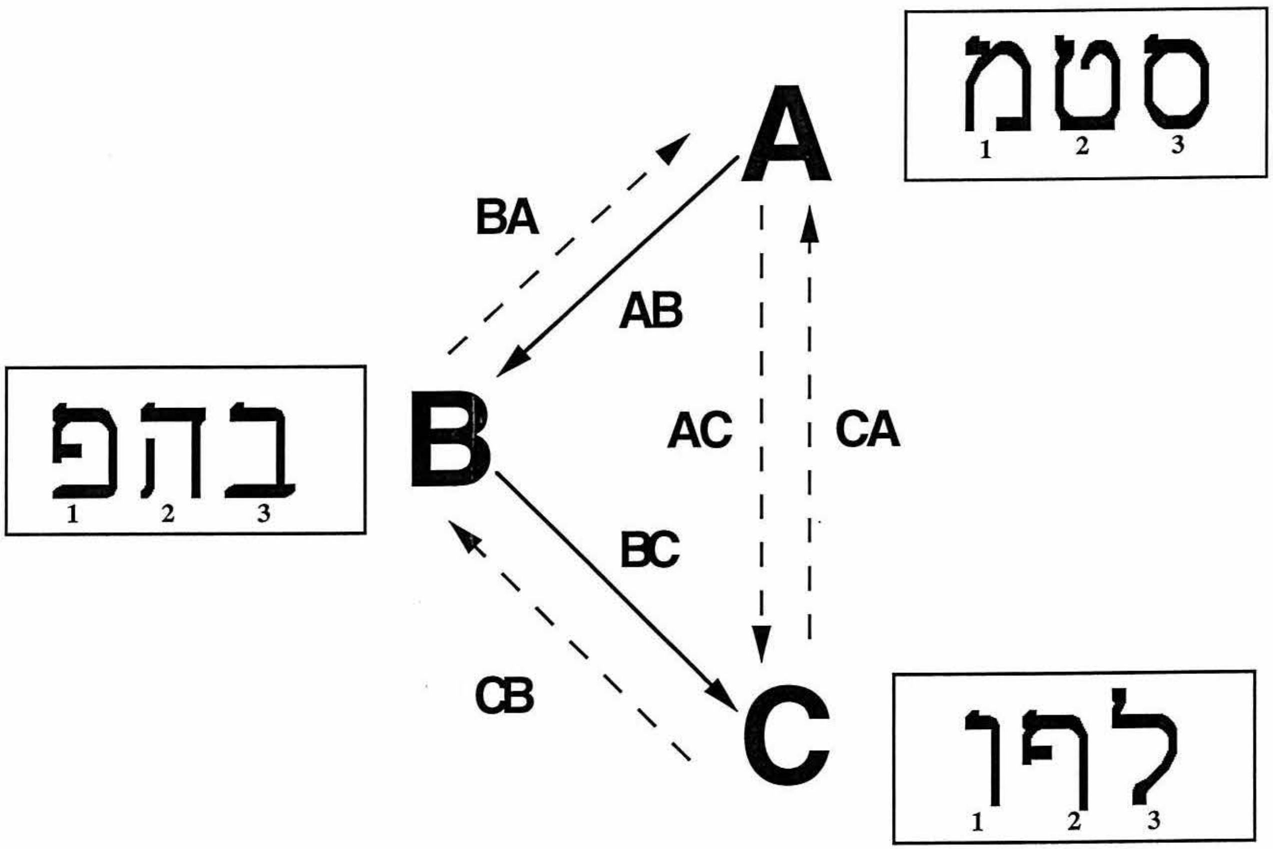
No firm conclusions could be drawn from Experiment 1 for two main reasons:

(1) The matching-to-sample protocol was too simple, meaning that if a subject scored 0% on any of the tests this constituted as a pass (as was the case with Control 7). Therefore, having three comparison stimuli reduces the likelihood of control by negative relations and therefore the likelihood of miscategorisng the pass/fail rate.

The actual paradigm employed is illustrated in Figure 11. This shows the relations trained and tested, and the stimuli used.

(2) It was not known whether the dyslexic and non-dyslexic

Figure 11: The matching-to-sample paradigm employed in Experiments 2 and 3. Solid arrows show the trained relations. Dotted arrows show the emergent relations. Arrows point from samples to comparisons.



סטמ
1 2 3
A
BA
AB
בהפ
1 2 3
B
AC
CA
BC
CB
C
לפן
1 2 3

group did produce significantly different naming latencies. As a result any observed difference or lack of difference between the groups on the test session trials could not be confidently attributed to whether the subjects were naming or not. So, a measure of the subjects' naming speed was introduced.
A discrete-trial format was adopted for the naming test because: (i) this was thought to mirror most accurately the naming processes which were expected to occur on the matching-to-sample trials. More specifically, the stimuli were presented individually or in small groups, and subjects were not under any constraints of time. (ii) Watson and Brown (1992) reported adult dyslexic/non-dyslexic naming speed differences employing a discrete-trial format; as did Fawcett and Nicolson (1994b) with adolescent dyslexic subjects. So, it was thought to be a reliable means of detecting the dyslexic subjects' naming difficulties.

Abstract, pictorial stimuli were employed once again, but the stimuli chosen were thought to be more difficult to discriminate between. One reason abstract, pictorial stimuli were employed was to make the stimuli harder to name and so, if naming is necessary for stimulus equivalence, they would be more likely to reveal differences between dyslexic and control subjects. The stimuli in Experiment 1 were, with hindsight, too easy to name as some resembled, quite closely, real-life objects. Given that the dyslexic subjects were adults and practised in naming such stimuli may not have placed their naming resources under enough strain to reveal any significant impairment.

Finally, a more in depth post-experimental interview was devised using a combination of questionnaire and interview. This was an attempt to access as accurately as possible the strategies subjects used throughout the experiment.

## Method

**Subjects** (see Appendix C for full subject details)

Sixteen subjects took part in this study (eight dyslexic subjects and eight non-dyslexic subjects). Each group comprised of seven males and one female, whose ages ranged from 19 years to 28 years (mean=22.12 years) for the dyslexic group; and 20 years to 32 years (mean=23.62 years) for the control group. Seven members of each group were students from various departments in the University of Wales, Bangor; the remaining two members were recruited from the general public.

One subject was not included in the final analysis. This was a dyslexic subject who failed to reach the criterion required during training despite two intensive sessions. Unfortunately, time restraints prevented any further attempts at training sessions.

Three members of each group were psychology undergraduates who received course credits for taking part in the study. The remaining ten subjects were paid a fixed sum for attending the two experimental sessions, regardless of how long they actually took to perform the experimental task.

None of the subjects had any prior knowledge of the theories or methodologies which were about to be tested.

Subject selection

Seven out of the eight dyslexic subjects were recruited via the Dyslexia Unit and had received an independent dyslexia assessment, which in the majority of cases included the administration of the Bangor Dyslexia Test (Miles, 1982) and the Schonell spelling test, and in all cases comprised of standard reading, spelling and intelligence assessments. Dyslexic 17 had not previously received an official diagnosis elsewhere and as a result was given the Bangor Dyslexia test and Schonell spelling test pre-experimentally. She scored 7 dyslexia positives on the Bangor Dyslexia test and 5/30 on the Schonell spelling test. This satisfied the dyslexia criteria outlined in Experiment 1.

Every member of the control group underwent the Bangor Dyslexia test and the spelling test to ensure that none of them possessed any dyslexic tendencies. None scored over 3.5 dyslexia positives (mean=2.25+ves). The mean spelling score for the control group was 20.13. Three subjects (Controls 10, 14 and 17) scored below 20 but this was paired with 2.5+ves, 3.5+ves and 3.5+ves, respectively and so the low score was attributed to bad spelling alone.

All subjects were: orally informed of their rights as participants in a psychology experiment (see Experiment 1); asked to sign a consent form; and required to fill out a pre-experimental questionnaire.

**Apparatus and Materials**

Once again the experiment was conducted in a quiet research room away from any distractions. The experimental setup is illustrated in Figure 12.

Identity Matching task - this task was run on an AppleMac LCii with a colour monitor. The programme was written and executed using the SuperLab 1.5.9 Beta application (Cedrus Corporation 1989-1992; timing resolution 1ms). The subject was required to respond using the specially designed keypad which fitted over the 4, 9, 6 and 3 keys on the number pad at the end of the keyboard (see Figure 12 for description). The stimuli used were chosen from the MTS Shapes font size 72 point (see Figure 13) because they were abstract, pictorial stimuli but not the same as the ones to be used in the Experimental task. Subjects' results (errors and response latencies) were recorded in a text file.

Naming Test - again this task was presented on the AppleMac LCii. Subjects were required to respond to the stimuli by saying its name aloud in to a microphone which was set up in front of them on a stand (see Figure 12). The subject's responses were recorded using a MacRecorder Sound System Pro (Macromind, Paracomp) and stored on a separate AppleMac SE computer (see Figure 12). Subjects' vocalisations were stored in the form of sound patterns so further analyses of vocalisations onset and total articulation times were possible using the SoundEdit Pro application (Macromind, Paracomp). Thirty-three stimuli were used. These were arranged in four subtests as follows:

Figure 12: A diagram to illustrate the experimental setup employed in Experiments 2, 4 and 5. S refers to the subject. E refers to the experimenter.

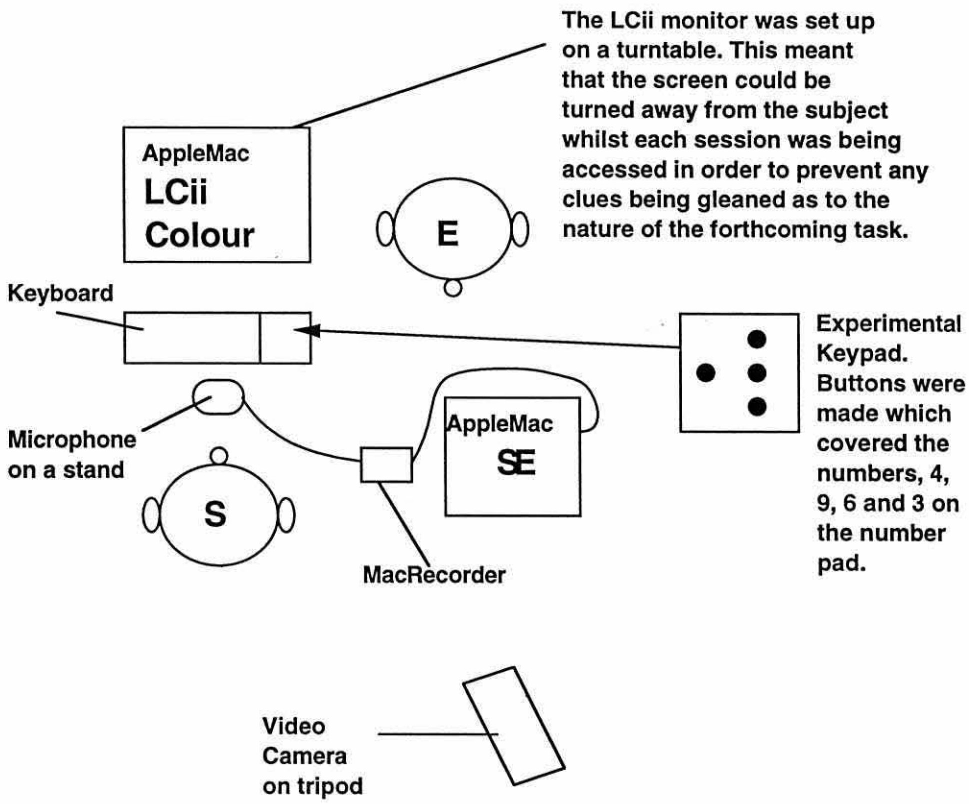
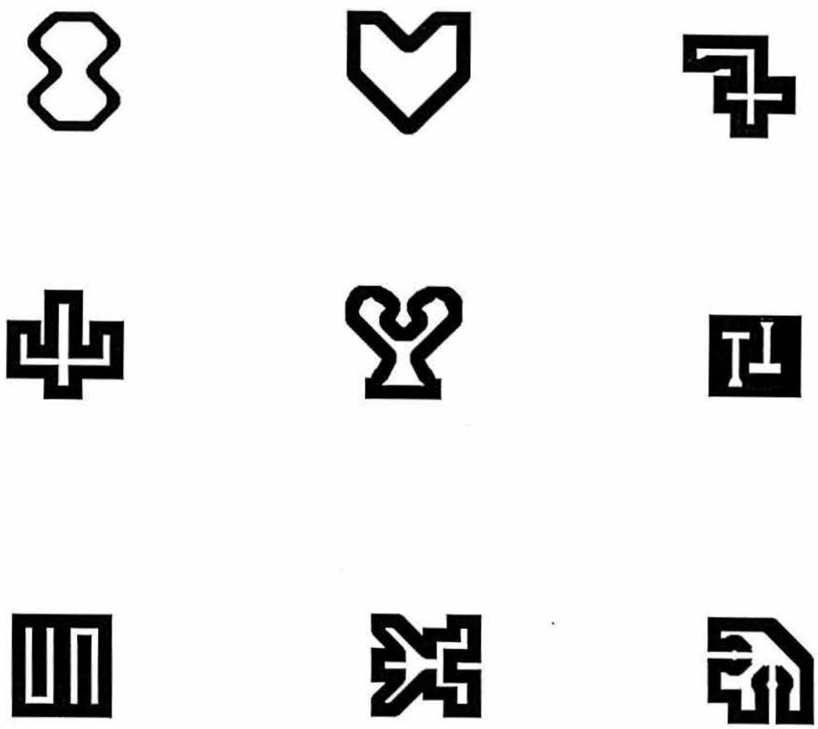

The LCii monitor was set up on a turntable. This meant that the screen could be turned away from the subject whilst each session was being accessed in order to prevent any clues being gleaned as to the nature of the forthcoming task.
AppleMac
LCii
Colour
E
Keyboard
Experimental Keypad. Buttons were made which covered the numbers, 4, 9, 6 and 3 on the number pad.
Microphone on a stand
AppleMac
SE
S
MacRecorder
Video Camera on tripod

**Figure 13: The nine experimental stimuli used the Identity Matching task for Experiments 2, 3, 4, and 5.**



| *9 pictures:* | *6 colours:* | *9 numbers:* | *9 lower case letters:* |
|---|---|---|---|
| Boat | Blue | 9 | t |
| Tree | Black | 1 | d |
| Fish | Green | 7 | s |
| Pig | White | 3 | c |
| Nose | Yellow | 5 | b |
| Dog | Red | 8 | o |
| Hand | | 2 | w |
| Frog | | 6 | g |
| Bed | | 4 | m |

The pictures, numbers and letters were taken from a naming test authored by Rod Nicolson (see Fawcett and Nicolson, 1994, for details). They were drawn in black on a white background and measured approximately 1 inch X 1.5 inches (2X3 cm). The 6 colours were drawn in KidPix (Craig Hickman and Broderman Software Inc. Version 1.1, 1991) and consisted of an appropriately coloured circle which measured 3"X2" (8cmX5cm) drawn on a white background. The programme was designed and run using the SuperLab 1.5.9 Beta application.

Experimental Task - this task was run on the LCii AppleMac. The programme was written and presented on SuperLab 1.5.9 Beta. As in the identity the identity matching session the subjects were required to respond using the keypad and their results were collated in a text file. Individual subjects responses were also recorded as being correct/incorrect manually on handwritten response sheets (Appendix D). This was in order to enable the experimenter to ascertain at a glance whether or not the subject had reached the session-by-session criterion. This was instead of accessing the computer files which would have been both time consuming and possibly distracting for the subject.

The nine experimental, abstract stimuli to be used in the matching-to-sample task were selected from the Hebrew font, size 72 point (see Figure 14).

Post-experimental interview - each subject was asked to complete a post-experimental questionnaire requiring them to relate any strategies used during the task and any perceptions they had concerning the experiment (see Appendix E). Subjects were also given a sheet with the experimental stimuli printed on and asked if they assigned names to any of them to help them during the sessions. Each subject was also presented with examples of the various trial types they experienced during the training and testing phases and asked to describe to the Experimenter, how they had responded when presented with these particular stimuli, and what went through their minds. This was all written and presented on SuperLab 1.5.9 Beta.

All sessions were video taped using a cam recorder mounted on a tripod in the corner of the room (see Figure 12).

**Figure 14:** **The nine experimental stimuli arranged in the three equivalence classes.**

| | Class 1 | Class 2 | Class 3 |
| --- | --- | --- | --- |
| | A1 מ | A2 ט | A3 ס |
| | B1 פ | B2 ת | B3 ב |
| | C1 ו | C2 ף | C3 ל |

**Experimental Design**

As in the previous experiment a mixed design was employed. The between-subjects factor was group (dyslexic vs non-dyslexic) and the within subject factor was task/category (on the test session: AB/BC baseline task; BA/CB symmetry task; AC transitivity and CA transitivity with equivalence task). This time nine stimuli were presented grouped into three equivalence classes (see Figure 14). Therefore, on each trial one sample stimulus was presented with three comparison stimuli (one from each equivalence class).

The trial presentation was 'pseudorandom' (to counteract any priming or practice effects) in the sense that trials were presented randomly within the constraint that no more than three trials of the same type could appear consecutively. Neither could three samples or comparisons appear together in the same position (top, middle, bottom) three times in a row. Every subject received the same order of trials on each session.

The experiment took place over two sessions. Session one consisted of the administration of: the Bangor Dyslexia test and Schonell spelling test where appropriate; the discrete trial naming test; and the identity matching task. Session two comprised of the experimental task and the post-experimental interview and debriefing session.

**Procedure**

The experiment ran over two sessions on separate days. The first session lasted about one hour and the second session between one and two hours. Subjects were allowed breaks between experimental sessions as required. An experimenter was present throughout.

Naming test - a practice session preceded the real test in order to familiarise the subject with the forthcoming procedure. This comprised of the nine picture trials presented individually in a different order than in the actual experimental task. Each stimulus remained on the screen for 5000ms. The instructions read to the subject were as follows, "This is a practice session. You must say the name of the following pictures. I will tell you whether you're right or wrong."

Oral feedback was given by the experimenter if any problems were encountered and to ensure that the subjects used the correct name. Next followed the actual naming test which consisted of four subtests of stimuli (9 pictures, 6 colours, 9 numbers and 9 letters) presented in a discrete-trial format. Subjects were asked to name the stimuli as quickly as possible into the microphone which stood in front of them. Before each subtest a screen appeared for 5000ms stating what was coming up next (for example, 'Pictures' or 'Colours'). Simultaneously presented with each stimulus was a tone (sounded at 22000 Hz). This was to signify to the subject and more importantly to the MacRecorder, which was recording the whole of the test, when the stimulus was presented on the screen. Each stimulus remained on the screen for 3000ms.

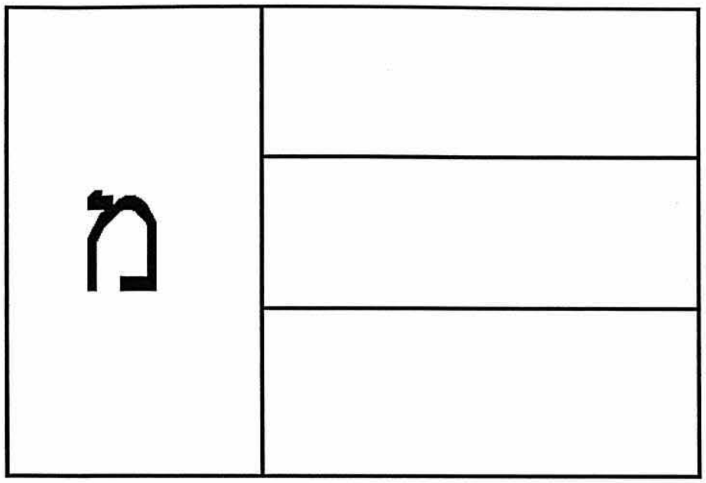
Subjects were given the following instructions:

> You will be shown a series of pictures, numbers, letters and colours. As each appears on the screen, you must say its name as fast as possible. Try and sit as still as possible; don't lean towards or away from the mic. Try and speak loudly and clearly. This time a tone will sound when the next picture is about to appear. Any questions?
> Please say the names as quickly and as accurately as you can. Are you ready?

Identity Matching - this served two purposes: (i) to provide a non-verbal reaction time measure (that is, by requiring the subject to match according to physical properties; and (ii) to familiarise the subjects with the matching-to-sample procedure. This followed the same M-T-S format as the subsequent experimental task. A sample stimulus appeared on the left side of the screen (see Figure 15). The subject had to respond by pressing the sample key (in the same position) on the keypad (see apparatus and materials). Three comparison stimuli then appeared down the right of the screen (see Figure 15). The subject received reinforcement for selecting the physically identical stimulus. The subject registered her/his selection by pressing the button on the keypad which corresponded to the position of the desired stimulus (that is top, middle or bottom). So, if the subject wished to select the top comparison they pressed the top button and so on. Feedback consisted of the word, 'correct' or, 'wrong' appearing appearing on the screen. This feedback message remained on screen for 3000ms, after which the next trial was immediately presented.

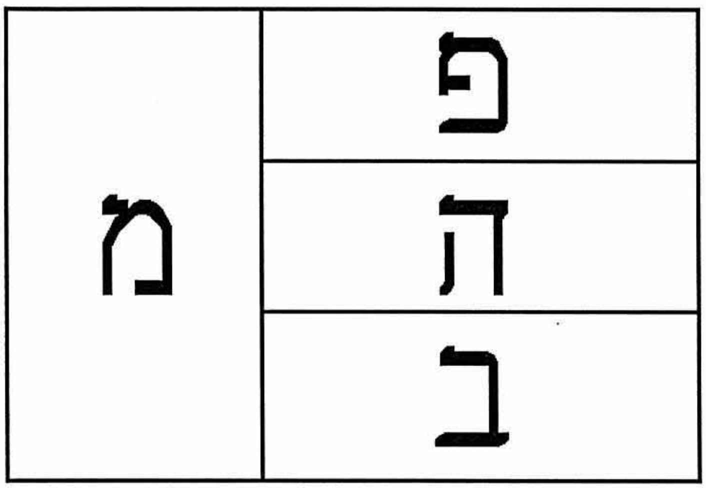
Eighteen trials were presented with different combinations of the nine stimuli (see Figure 13). If a subject made more than two errors in the last sixteen trials or else the session was

Figure 15: An illustration of the procedure for a typical training trial (A1-B1). This procedure is the same for the Identity Matching task (although the stimuli used are different).




Sample A1 appears.

The subject must press the corresponding button on the keypad.




3 comparisons (B1 B2 B3) appear. The subject must select one by pressing the appropriate button (top, middle or bottom) on the keypad.

repeated. The following instructions were read aloud to each subject:

> In front of you is the computer which you are going to use. You will respond by using the four black buttons at the end of the keyboard (experimenter points to where they are). When the experiment begins a visual image will appear on the left of the screen in the same position as the button here (experimenter shows). In order to carry on you must press this button. Three more images will then appear in the same position as these three buttons to the right (experimenter shows). When these appear you must press one of the three buttons. Each button corresponds to the image in the same position on the screen. If you select the right image the word, 'correct' will appear on the screen. If you select the wrong image the word, 'wrong will appear instead. Please try and get as many correct as you can. Feel free to talk aloud as you go along. Do you understand or do you want me to repeat the instructions?

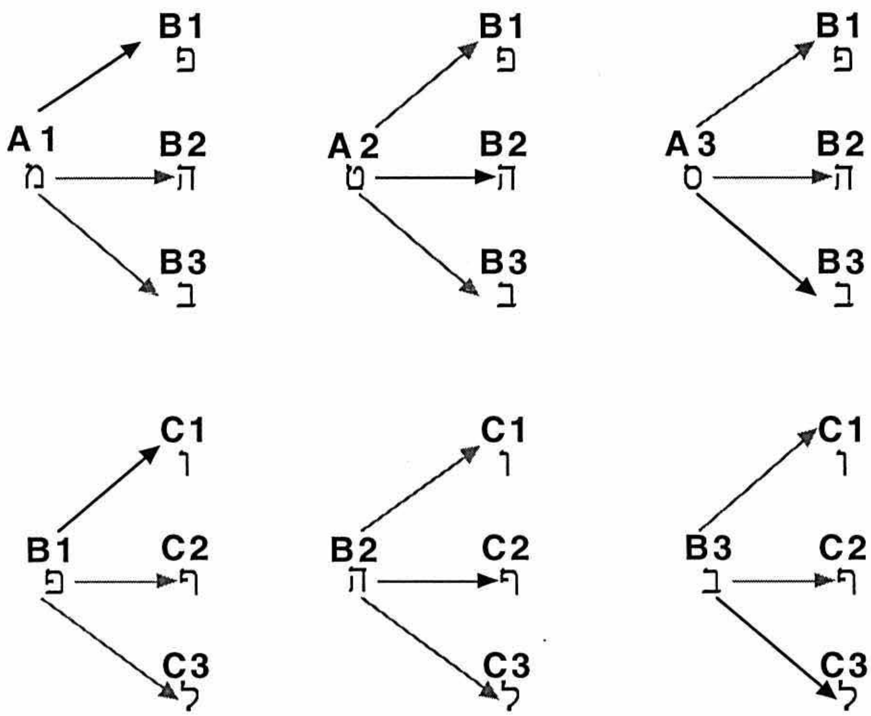
Experimental task - this followed the same M-T-S procedure as outlined under identity matching. However, instead of matching physically identical stimulus pairs, subjects were reinforced for matching samples and comparisons according to their classes (A1 B1 C1; A2 B2 C2; A3 B3 C3 - see Figure 16). An example trial is as follows:

A1 sample appears this leads to comparisons, B1, B2, and B3 being presented. The subject's selection of the B1 comparison is reinforced. Each subject received the same instructions and was required to respond in the same manner as in the identity matching phase.

Phase 1: Train AB and BC relations

Subjects were first trained on the three AB relations starting with the trial type A1-B1. Eighteen A1-B1 trials were presented with the different possible comparison

**Figure 16:** **Example training trials illustrating how a sample is presented (A1), followed by the comparison stimuli (B1 B2 B3). Black arrows represent correct choices. Grey arrows represent incorrect choices.**


A1 מ
B1 פ
B2 ה
B3 ב
A2 ט
B1 פ
B2 ה
B3 ב
A3 ס
B1 פ
B2 ה
B3 ב
B1 פ
C1 ו
C2 ף
C3 ל
B2 ה
C1 ו
C2 ף
C3 ל
B3 ב
C1 ו
C2 ף
C3 ל

combinations (for instance, A1-B1 B2 B3, A1-B1 B3 B2, A1-B2 B1 B3, A1-B2 B3 B1, A1-B3 B1 B2, A1-B3 B2 B1 and so on) repeated three times each. Trials were presented pseudorandomly (see experimental design). Subjects were required to make no more than two errors in the last fifteen trials (87% correct). If a subject fell below this criterion the session was repeated. If the subject reached criterion the next session was presented which was trial type A2-B2 followed by A3-B3. Once the three AB relations had been trained the subject moved on to an AB session which consisted of the three trial types mixed up. Thirty-six trials were presented pseudorandomly (12 A1-B1, 12 A2-B2, 12 A3-B3). Subjects could make no more than two errors per trial type which corresponded to 83% correct or else the session was repeated. If one particular trial type produced noticeably more errors than the others the individual trial type session was repeated. Once the AB criterion had been achieved the three BC trial types were trained in the same manner and to the same specifications (for instance, B1-C1, B2-C2, B3-C3, then BC - all the three trial types combined).

When the criterion had been reached on both the AB and BC tasks individually the various trial types were combined into a AB/BC reinforced training session. This consisted of thirty-six trials (18 AB - 6 A1-B1, 6 A2-B2, 6 A3-B3; and 18 BC - 6 B1-C1, 6 B2-C2, 6 B3-C3). Subjects proceeded to the next phase provided they made no more than one error per trial type (83% correct).

Phase 2: AB and BC without reinforcement

The latter session (AB/BC with reinforcement) was repeated with the trials arranged in a different order and in the absence of reinforcement to verify that the learned AB and BC relations could be maintained without feedback. After each trial a blank screen was presented for 3000ms where the reinforcement screen would have been. This was to ensure that the pacing of the trials was not affected by the dropping the reinforcement. Again the subject could not make more than one error per trial type or else the session was repeated. The following instructions were read to the subject:

> The procedure is exactly the same as before but this time the computer will not tell you whether your response is right or wrong. I will tell you at the end how well you have done. As before try and get as many correct as you can.

Phase 3: Test symmetry (BA & CB), transitivity (AC) and equivalence (CA)

Once the subject had reached the criterion required in Phase 2, the test session was presented without reinforcement. A blank screen was presented instead of the reinforcement screen as in the previous phase. The session consisted of the AB and BC baseline tasks pseudorandomly presented along with the symmetry, transitivity and equivalence tasks. One hundred and eight trials were presented (divided into two halves of fifty-four trials consisting of identical numbers of each trial types) which comprised of 18 trials per task and 6 per trial-type as follows:

18 AB trials (6 A1-B1, 6 A2-B2, 6 A3-B3)
18 BC trials (6 B1-C1, 6 B2-C2, 6 B3-C3)
18 BA trials (6 B1-A1, 6 B2-A2, 6 B3-A3)
18 CB trials (6 C1-B1, 6 C2-B2, 6 C3-B3)
18 AC trials (6 A1-C1, 6 A2-C2, 6 A3-C3)

18 CA trials (6 C1-A1, 6 C2-A2, 6 C3-A3)

The trials were presented without reinforcement and the same instructions as in the previous phase were read to the subjects. In addition the subject was told that this session was longer than the previous sessions but there was no indication given as to its composition.

Post-experimental Interview/Debriefing session

The subject was asked to fill out a written questionnaire (see Appendix E) asking her/him about their perceptions of the experiment or any strategies they employed to help them during the session. They were next shown the experimental stimuli printed on a sheet of paper and asked whether they had assigned any names to the stimuli to aid in remembering them. Subjects' responses were noted by the experimenter. Following this the subject was shown a representative selection of the trials they saw during training and testing and asked what went through their minds at various stages of the experiment and how they associated the stimuli (see Appendix F for transcript).

# Results

## Naming Test

From each subject's performance on the naming test came a series of sound pattern recordings stored in the MacRecorder application. These comprised of the signal which represented the presentation of the visual stimulus on the screen and the subsequent vocalisation. Two measures were extracted from these recordings: (i) the latency from the stimulus presentation time to the vocalisation onset; and (ii) the time from the stimulus presentation to the end of the total articulation (total articulation time). Both measures were taken in order to ascertain exactly where any difference lay. These measures were subjective up to a point in that they involved visually judging where measures should begin and end, so a second observer was asked to make the same measurement on the same data. This reliability check was performed on 25% of the data. No more than 8.5% disagreement was observed between the two observers' measurements. The data from the colours subtest were not included in the final analysis due to procedural difficulties.

**Table 4: Group means (presented in milliseconds) for each measure on each of the three naming test subtests (standard deviations given in brackets)**

| Vocalisation Onset | Pictures | Numbers | Letters |
|---|---|---|---|
| Dyslexic Group | 790.12 (106.22) | 828.25 (182.37) | 1028.43 (388.64) |
| Control Group | 804.52 (118.38) | 775.05 (126.76) | 846.92 (107.61) |
| Total Mean | 797.32 (108.91) | 801.65 (154.19) | 937.68 (290.99) |
| *Difference* | *14.4* | *53.2* | *181.51* |
| **Total Articulation** | | | |
| Dyslexic Group | 1038.26 (122.87) | 1080.89 (189.82) | 1263.19 (406.99) |
| Control Group | 1066.18 (132.78) | 1052.54 (131.07) | 1100.72 (116.42) |
| Total Mean | 1052.22 (124.42) | 1066.72 (158.26) | 1181.96 (301.11) |
| *Difference* | *27.92* | *28.35* | *162.47* |

Table 4 shows that the biggest difference between the groups was found on the letters subtest, regardless of the measure; and the smallest difference was found on the pictures subtest. The control group produced longer latencies than the dyslexic group on the pictures subtest. On the other subtests, the dyslexic group took the longest. A mixed design 2X3 ANOVA was performed on each set of data. On the vocalisation onset measure no significant difference was found between the two groups' data ($F=0.726$, $df=1,14$ ns). The same was also found using the total articulation time ($F=0.352$, $df=1,14$ ns) However, for both measures a significant difference was found between the time taken on each of the subtests: vocalisation onset ($F=6.9$, $df=2,28$ $p<0.005$); total articulation ($F=5.047$, $df=2,28$ $p<0.01$). In order to locate where the significant differences lay pairwise comparisons were made between the subtest means using Tukey's HSD tests at the 0.05 level of significance. On the vocalisation onset measure there was no significant difference between the pictures and numbers subtest, but a significant difference was found between pictures and letters, and numbers and letters ($q\ (3, 28) = 106.44$). In other words, the letters subtest took significantly longer than the other two. Using the total articulation time measure a similar picture emerged. Again there was no difference found between the pictures and numbers subtest; but the letters subtest took significantly longer than the other two (according to Tukey HSD test at the 0.05 significance level, $q\ (3, 28) = 110.85$).

No interaction effects were found on either measure (vocalisation onset, $F=2.681$, $df=2,28$ ns; total articulation, $F=2.389$, $df=2,28$ ns).

## Identity Matching

Each group was measured on their time taken to match pairs of physically identical abstract shapes. A median response

latency was calculated for each subject.
The dyslexic group's mean was 1490.31ms (sd=838.11) and the control group's mean was 1128.62ms (sd=345.04). An F-test predicted that there was a significant difference in variance between the two samples (F=5.9, num.df=7, den.df=7, p<0.05). Therefore, the data were analysed using the non-parametric Robust Rank-Order Test, which demonstrated that there was a non-significant difference between the two groups (U`=0.59, m,n=8, ns).

**Training - Phases 1 and 2**

The groups were compared on their performances during the matching-to-sample training sessions. The number of errors to criterion made by each subject over the reinforced training sessions was calculated; followed by the number of errors made on the unreinforced training sessions alone; and finally the number of repeat sessions each subject required (the minimum number was 10).

**Table 5:** The mean number of errors per group during: Phase 1, train AB and BC; Phase 2, AB and BC without reinforcement; and the mean number of repeat training sessions required by each group. Standard deviations are given in brackets.

| | Phase 1 Number of Errors to Criterion | Phase 2 Unreinforced Errors | Repeat Number of Sessions |
|---|---|---|---|
| Dyslexic Group | 93.63 (87.95) | 1.25 (1.83) | 13.00 (13.23) |
| Control Group | 21.63 (21.27) | 0.50 (0.76) | 2.25 (1.67) |

Table 5 shows that the dyslexic group made more errors before reaching criterion than the control group. The two standard deviations show that there is a difference in distribution between the two groups. This was confirmed by an F-test which demonstrated that there was a significant difference in variance between the two groups (F=17.092, num.df.=7, den.df=7, p<0.001). As a result the Robust Rank-Order Test was applied to the data and a significant difference between the groups was observed (U`=2.95, m,n=8. sign. at the 0.05 level). Therefore, the dyslexic group made significantly more errors then the control group during Phase 1.

The next question which arose was whether this difference was maintained in the training sessions once the reinforcement had been dropped (Phase 2). Table 5 shows that the dyslexic group produced a mean of 1.25 errors and the control group a mean of 0.5 errors. Once again an F-test confirmed that there was a significant difference between the amount of variance in the two groups (F=5.875, num.df=7, den.df=7, P<0.05). Therefore, a non-parametric test was used. The Robust Rank-Order Test showed a non-significant difference between the groups on this measure (U`=0.66. m,n=8, ns).

The final comparison was made on the number of repeat training sessions each group required. Table 5 demonstrates that the dyslexic group needed a mean of 13 extra training sessions and the control group only 2.25. However, it must also be noted that there is a large difference in the standard deviations of the two samples. An F-test showed that the variance of the groups' data differed significantly ($F=62.769$, num.df=7, den.df=7, $p<0.0001$). Therefore, a Robust Rank-Order Test was used and a non-significant difference was found ($U`=1.88$, m,n=8 ns)).

Looking at the raw data it can be seen that the dyslexic group's high mean is inflated by three subjects' data (Dyslexic 15, 26 sessions, Dyslexic 15, 23 sessions; and Dyslexic 17, 35 sessions) which is nearly half of the group. Such high numbers of sessions are only found in the dyslexic sample and suggests a strong trend that the dyslexic group required more exposure to the reinforcement contingencies than the control group.

**Test Session - Phase 3**

As in Experiment 1 the test session trials were collapsed into three categories: *baseline* (AB and BC tasks); *symmetry* (BA and CB tasks); and *transitivity* (AC and CA tasks).

Errors made during the test session

The test session consisted of 108 trials - 36 trials per category. The pass criterion adopted was that a subject had to make no more than eight errors per category. Therefore, each subject had to attain 28/36 per category which corresponded to 78% correct (there were three comparisons so chance level was 33.3%).

**Table 6: The pass/fail rate for each group**

| | Pass | Fail |
|---|---|---|
| Dyslexic Group | 5 | 3 |
| Control Group | 6 | 2 |

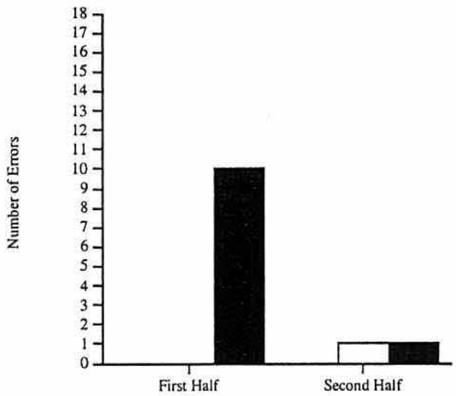
Table 6 shows the numbers of subject from each group who met the criterion. Only one more control than dyslexic subject passed. A Fisher's Exact test (for small samples) was applied to the data, and this demonstrated that there was no significant difference between the groups (phi=0.135, df=1, ns).

As in Experiment 1 the subjects who passed made very few errors due to the strict pass/fail criterion enforced. The dyslexic group made a mean of no more than 8% errors on any category. For the control group it was no more than 7% errors. So, no further statistical analysis was conducted.

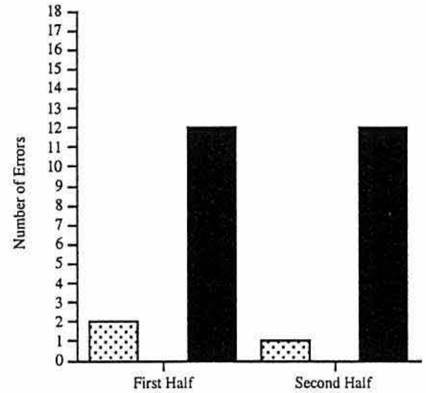
However, the errors of the subjects who failed to demonstrate equivalence were examined on an individual basis. Figure 17 presents error graphs for the subjects who failed to reach criterion on the test session (n=2 control subjects; n=3 dyslexic subjects).The horizontal axis is split into first and second halves of the test session. The whole session comprised of 108 trials which could be divided into two halves consisting of 56 trials of identical types but presented in a different psuedorandom order. The vertical axis is numbered up to 18 as this is the total number of trials in each category for

**Figure 17: Graphs illustrating the number of errors made by the subjects who failed, on each category, on each half of the test session in Experiment 2.**

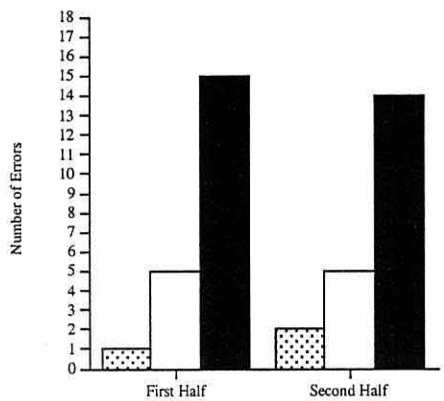
**Control 16**



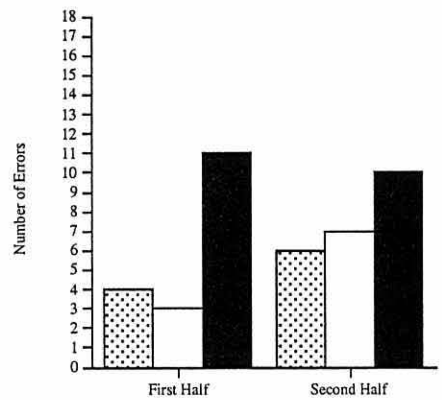

**Dyslexic 15**



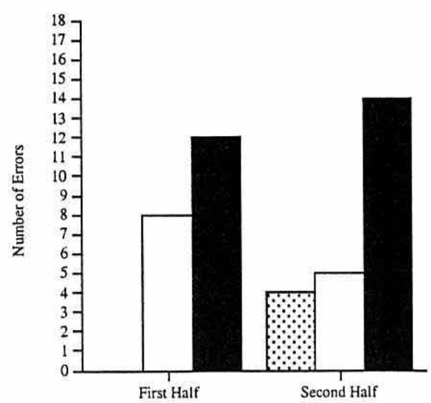

**Control 17**




**Dyslexic 16**




**Dyslexic 17**




Key

- Baseline
- Symmetry
- Transitivity

each half of the session (for example, first half = 18 baseline, 18 symmetry, and 18 transitivity). With the exception of Control 16) whose performance improved on the second half, all subjects first and second half performances were approximately the same. Thus the three dyslexic subjects who failed were still doing so by the second half of the session. This is reflected in the fact that Dyslexic 16 and Dyslexic 17 performed slightly worse on the baseline trials on the second half, and this was also the case on the symmetry trials for Dyslexic 16. The only subject to show evidence of baseline deterioration was Dyslexic 16.

In comparison only one control subject, Control 17, persisted in failing. Once again there is some evidence that the dyslexic subjects who fail are more likely than control subjects to persist in failing.

Test trial response latencies

The median response latency on each test session category was calculated for each of the subjects who successfully demonstrated stimulus equivalence (n=5 dyslexic subjects; n=6 control subjects). Once again it must be noted that these are small sample sizes so any parametric analyses must be viewed with caution.

**<u>Table 7: The mean of median response latencies for each group on the three test session categories (standard deviations given in brackets).</u>**

| | Baseline | Symmetry | Transitivity | Total |
|---|---|---|---|---|
| Dyslexic Group | 2370.5 (854.17) | 2820.7 (969.91) | 4491.4 (2078.65) | 3227.53 (1613.55) |
| Control Group | 2150.17 (1075.33) | 2719.83 (907.45) | 4229.08 (3161.62) | 3033.03 (2082.52) |
| Total Mean | 2250.32 (939.81) | 2765.68 (889.27) | 4346.32 (2597.12) | |

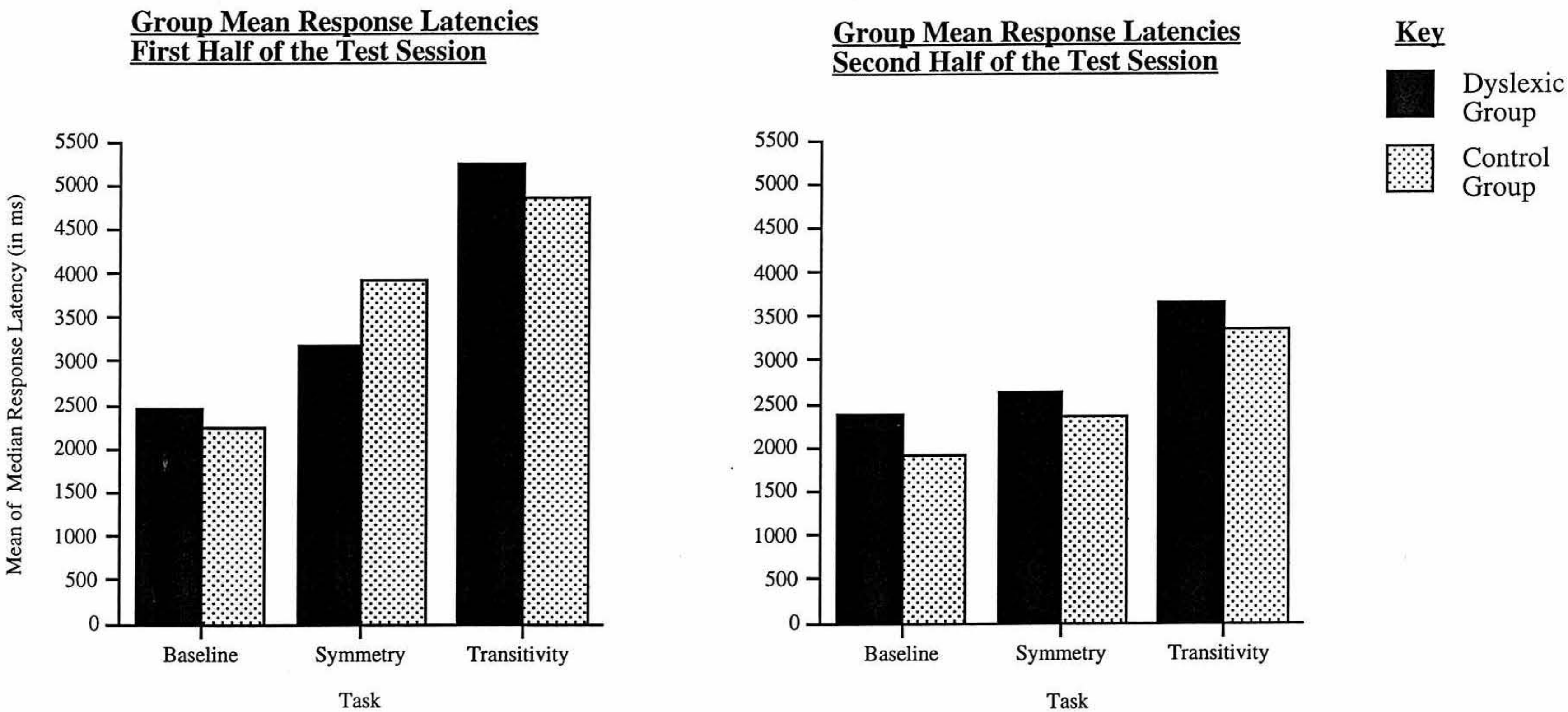
Looking at Table 7 it can be seen that there is little difference between the group means, although the dyslexic subjects took slightly longer on all three test categories. A 2X3 mixed design ANOVA confirmed that there were no significant differences (F=0.047, df=1,9 ns). There was also no significant interaction effect (F=0.014, df=2,18 ns). However, looking at the total means for each category both groups produced much longer mean latencies on the transitivity trials. The above ANOVA confirmed a significant within-subject effect (F=9.537, df=2,18 p<0.001). This effect was investigated further using Tukey's HSD tests. No significant difference was observed between the baseline and symmetry tasks but the transitivity tasks did significantly differ from the other two (at the 0.05 level of significance, q (3, 18) = 1278.07). Therefore, the transitivity trials were taking significantly longer.

As stated previously a non-significant main effect of group or groupXtask interaction is possibly not due to the absence of an

effect but in this instance could be down to there not being enough data points to detect any difference. Therefore, an alternative way of looking for potential effects was to represent each groups data graphically. Figure 18 illustrates each group's mean of median response latencies for each of the test session categories on the two halves of the session.
It can be seen that the mean latencies drop over time, and that this was especially the case for the symmetry and transitivity tasks, so the subjects get faster as the session progresses as they become more practised. The group difference, which is the variable of interest, remains the same over both halves of the session. So, there is no reason to hypothesise that the groups would statistically differ on one half rather than the other.
It should be noted that the control subjects took longer than the dyslexic subjects on the first half symmetry task, which is not what would have been predicted if the subjects were naming (that is, the dyslexic subjects should take longer due to their naming weakness).

Finally, the data of the subjects who passed the test session was split according to their performance on the naming test. A mean naming latency was calculated for each subject, placed in nummerical order, and assigned to either the fast or slow group depending on its rank order (fast, n=5, mean=754.61ms; slow, n=6, mean=1053.89ms). An unpaired t-test showed that these two samples were signifcantly different ($t=-3.69$, $df=9$, $p<0.005$). The test session latencies of these two groups were then compared using a 2X3 mixed design ANOVA. No significant group difference was found ($f=0.286$, $df=1, 9$, ns).

**Figure 18: Graphs comparing each group's mean of median latencies on the first and second half of the test session in Experiment 2.**



Group Mean Response Latencies
First Half of the Test Session
Mean of Median Response Latency (in ms)
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Baseline
Symmetry
Transitivity
Task
Group Mean Response Latencies
Second Half of the Test Session
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Baseline
Symmetry
Transitivity
Task
Key
Dyslexic Group
Control Group

## Discussion

The dyslexic and non-dyslexic group were compared on the discrete-trial naming test using familiar stimuli (pictures, letters and numbers). No group differences were found using either the vocalisation onset measure (F=0.726, df=1,14 ns); or the total articulation time (F=0.352, df=1,14 ns). This was a surprising finding given the existing research outlined in Chapter 3 which predicts that naming differences should be observed under these circumstances due to the fact that developmental dyslexic is characterised by phonological deficits which persist into adulthood. What is apparent from these data is that these deficits, most typically reported in childhood, may not exist into adulthood to such an extent.

No evidence was found that adult dyslexic and non-dyslexic samples differ in their production of the names of familiar stimuli which are presented individually despite significant differences being found in childhood (Denckla and Rudel, 1976b; Perfetti, Finger, and Hogaboam, 1978; and Stanovich, Freeman, and Cunningham, 1983).

Fawcett and Nicolson (1994) reported significant naming differences between dyslexic and non-dyslexic samples which included 17 year old subjects, using a test and stimuli which were almost identical to the one implemented in this study. So, the data presented here challenge their assumption that discrete-trial naming deficits persist into adulthood. However, from the evidence presented in their paper it is impossible to ascertain whether the 17 year old dyslexic and control samples significantly differed as the authors did not make this direct comparison. The significant group difference they report was only found when these subjects were grouped together with the younger samples suggesting the possibility that a difference did not exist between the older subject groups but only between the younger ones.

So, the data presented here are of value in that they establish that adult dyslexic and non-dyslexic samples do not significantly differ in their discrete-trial naming of familiar stimuli and that this naming deficit, reliably observed in childhood, can be compensated for. This compensation is doubtless as a result of years of experience in producing the names of familiar pictures, numbers and letters. The important implication is that caution should be taken when generalising from findings established from childhood studies to adult subject groups.

How does this finding relate to the experimental hypothesis? The rationale behind using familiar stimuli in the naming test was that if there is a significant naming difference between the groups using the familiar stimuli then there should be a significant difference using the abstract stimuli which are presumably harder to name. However, a non-significant difference on the discrete-trial naming test utilising familiar stimuli does not necessarily mean that the same non-significant difference will be reflected in the equivalence test session latencies. Naming differences could still be produced on the equivalence test trials because the stimuli employed here are unfamiliar and thus harder for the dyslexic subjects to name.

This all suggests that the discrete-trial naming test employing familiar stimuli may not be the most pertinent measure to detect any dyslexic and non-dyslexic differences and may not provide much useful information when examining subject's naming on test session trials.

However, a significant difference was found between the various subtests on both measures of articulation onset and total articulation time. Whatever the measure taken, both groups took significantly longer to name the letters subtest, and it was on this subtest where the largest group differences were found see Table 4. The dyslexic group took 181.51ms longer according to the vocalisation onset measure which is consistent with the notion that dyslexic individuals are deficient in producing the spoken representations/names of letters and words (for example, Perfetti et al 1978 - words; Stanovich et al 1983 - letters). Also, looking at the mean values Fawcett and Nicolson (1994) present for the dyslexic and non-dyslexic 17 year old subjects, the largest group difference there was found on the letters subtest too.
The dyslexic group took the shortest time on the pictures subtest whereas the control group took the least time on the numbers subtest. Therefore, different subtests had a differing effect on the groups. The dyslexic group seemed to take longer when the stimuli were graphological and involved symbolic processing (for example, numbers and letters) as opposed to non-graphological (pictures or colours). The similarity of the findings on the two measures suggested that there is no need to take both measurements in order to assess naming speed.

Moving on to the matching-to-sample data, no significant difference was found between the two group's performances on the nonverbal Identity Matching task. Therefore, any subsequent significant group latency differences were not a result of the samples containing overly fast or slow responders.

On the training sessions the dyslexic group made significantly more errors before reaching criterion (U`=2.95, m,n=8 sign. at the 0.05 level). However, no significant differences were found between the groups regarding the number of errors made during the unreinforced sessions alone; or regarding the number of repeat training sessions needed. On the latter measure, however, there was a strong trend which suggested that the dyslexic group were more likely to require extra exposure to the baseline relations before reaching criterion as nearly half of the group needed many more training sessions than any of the other subjects.

So, the dyslexic group made significantly more errors and tended to require more training sessions before reaching criterion when the stimuli used were difficult to name. This is in line with the predictions outlined in Chapter 3 whereby dyslexic subjects have difficulty learning verbal material (Done and Miles, 1978; Vellutino, 1987). The fact that there was no significant difference between the groups on the unreinforced sessions suggests that both groups were able to maintain the learned relations without feedback. No deterioration in performance was observed once the reinforcement was dropped. Therefore, even though the dyslexic subjects took longer to learn the relations once they were established their performance was on a par with that of the control subjects. This is in contrast to Experiment 1 where the dyslexic group did not differ from the control group in learning the initial relations but did tend to be more adversely affected by the removal of feedback. This suggests that maybe 'over-training' was needed in order for the dyslexic subjects to maintain their learning as well as the control group as seen in Experiment 2.

Looking at the equivalence test session data, there was no significant difference between the pass/fail rate of the two groups. However, examining on an individual basis the data of the subjects who failed (see Figure 17), it was seen that more dyslexic subjects persistently failed over both halves of the session. Three dyslexic subjects were still failing by the second half in comparison to just one control subject. Therefore, although there was no statistical difference between the groups there was a tendency for more dyslexic subjects to fail more unequivocally.

So, it can be seen that the dyslexic subjects make more errors during training and also make more errors during the test session. These findings are consistent with the assumption that dyslexic subjects demonstrate a naming impairment in that it was predicted that if the subjects do not name consistently then they would fail to demonstrate stimulus equivalence.

It was hypothesised in Chapter 3 that any naming latency difference between the groups should be reflected in the test trial latencies of the subjects who passed if they were naming the stimuli (as they should be according to the naming hypothesis). Even though no significant difference was detected between the groups on the discrete-trial naming test it was important to examine the test trial latencies given that different stimuli were used here. It has been argued that no significant difference between the groups on the discrete-trial methodology using familiar stimuli does not necessarily reflect no significant latency difference on the test session trials employing unfamiliar stimuli. However, no significant differences were found between the groups when the test session latencies were analysed.

Looking at Figure 18 it was seen that the groups did not differ over either half of the session. It was seen, however, that for first half symmetry task the control group took 721.08ms longer, which is not what would be predicted if the subjects were naming, that is, the dyslexic subjects should take longer. Why should the control subjects take longer on the symmetry task and not the other tasks? Of the two groups it is more likely that the control group would name the stimuli (the naming hypothesis states that this is the most likely strategy adopted by verbally competent adults). If this were true then the symmetry task would initially be the less familiar task as a human subject is not practised in reversing name-name relations (for instance, A1B1, "green-house" to B1-A1, "house-green") and such relations may take longer to establish.

The overall finding was that no significant group differences were observed on the test session latencies. As in Experiment 1 this signals two lines of argument:

(1) The dyslexic subjects do not demonstrate a greater difference in naming speed. In other words, if there was a significant difference between the group's naming of abstract stimuli (which can not be empirically tested here) it was not reflected on the test session trials suggesting that there was no difference between the groups on this measure.

(2) The subjects are not naming on the test session trials.

An underlying assumption of the experimental rationale is that if the groups significantly differ in naming speed and if they are naming during the test session then this will result in a significant difference on the test session latencies. In order to partly address this issue the data of the subjects who passed was split into two groups (fast and slow namers) who significantly differed on the measure of naming. These two fast/slow groups were then compared on the test session latencies and no significant difference was found. So, here are two groups who definitely differ on the naming measure but not on the test session latencies. As a result it could be argued that there is a strong possibility that the subjects are not naming. However, it should be noted that these two groups were also compared on the identity matching task (that is, a nonverbal task) and were found to again significantly differ ($t=-2.673$, $df=9$, $p<0.05$). So it could be that one group merely contained generally fast or slow responders who would signifcantly differ from a control group regardless of the nature of the task.

One way to assess what the subjects were doing over the course of the experiment was to examine the subjects' post-experimental interviews.

All the subjects who passed either reported that they assigned names to the stimuli or described the physical features of the stimuli in some way. Many subjects reported that due to the fact that a particular stimulus may have resembled a real life object/letter for them, this suggested that they should assign names to the other stimuli. This illustrates, perhaps that a stimulus can never be totally abstract as a stimulus can always be subjectively likened to anything in the individual's experience.

Physical features also provided the subjects with some basis for discriminating the stimuli. For example, A1 and B1, share a gap at the bottom and A2 and B2, share a gap at the top. Some subjects reported that they realised that the stimuli fell into three equivalence classes before testing and named the stimuli accordingly. For example, Dyslexic 11 reported that A1 looked like a, 'house' and this linked to B1 which he, as a result, called, 'garden' and C1 became,'I' which represented indoors. Control 10 linked the stimuli in a short sentence. For example, A3B3C3 was hanged man-gallows-noose. Control 11, gave a common name to the stimuli in each class: class 1 was gap at the bottom; class 2 was gap at the top; and class 3 was no gap. Similarly, Control 14 divided the stimuli as follows: class 1 was named shapes; class 2 was outcast shapes; and class 3 was desired shapes. This evidence again supports the Naming Hypothesis in that all the subjects reported to have assigned names to the stimuli passed the equivalence test.

However, the five subjects who failed also reported that they assigned names to the stimuli. Yet only Control 16 (whose performance notably improved on the second half) could verbalise what the task requirements were. He grouped the stimuli according to physical features. This seems to be the difference between the subjects who passed and failed. The subjects who passed linked the stimuli together either as rules (for instance, A goes with B) or by common names and features. Such strategies were not evident in the subjects who failed.

The issue which remains is that despite the post-experimental reports it can never be said definitively what the subjects were doing during the sessions. The subjects may have been naming, but not consistently. Or merely reported that they named post-experimentally. The question is, was the verbal strategy a cause or a product of stimulus class formation. This can not be established conclusively from this data alone. For instance, for the subjects who pass, their verbal strategies could have been a result of them demonstrating equivalence in that during the post-experimental interview, they were verbalising what they had been doing during the test session for the first time. So, what on the surface looks like support for the naming hypothesis may not be.

The successful subjects' patterns of responding were also examined. Significant differences were found between the latencies on the test session categories. It was demonstrated for both measures that: baseline/symmetry < transitivity which is one pattern observed in the performances of subjects who passed the test of equivalence in Bentall et al (1993). This pattern according to Dugdale and Lowe (1990) and Horne and Lowe (1996) is compatible with the subjects assigning individual names to the stimuli.

All this evidence suggests that it cannot be said unequivocally whether the subjects were or were not naming on the equivalence test session. There is therefore, a need for a shift of tactic: (i) to utilise stimuli for which the naming latencies can be measured: and (ii) to employ a procedure on which it can be said that the subjects are definitely naming the stimuli, rather than relying on the subjects' post-experimental remarks.

## Experiment 3

### Background

In the first two studies no significant group differences were observed between the subject's response latencies on the naming or the equivalence test session trials. It was hypothesised that if there existed a significant difference in naming latencies between the dyslexic and non-dyslexic subjects, then if naming was required on the equivalence test session categories for the subjects to successfully demonstrate equivalence (as argued by the naming hypothesis) then this difference should be reflected in the equivalence test session latencies. It could be argued that if no significant differences were observed on the test session latencies then this implies that many subjects demonstrated equivalence independently of naming.

This hypothesis is founded on two assumptions:
(1) There exists a significant naming latency difference between dyslexic and non-dyslexic adult subjects.
Experiment 2 included a comparison of the two groups on a discrete-trial naming test using familiar stimuli. It was argued that a significant difference in the naming of familiar stimuli would point to a significant difference in the naming of abstract stimuli, given that the latter task is more difficult.
No significant difference was found between the groups on the task involving familiar stimuli so it could not be concluded whether the subjects would or would not significantly differ when naming the experimental abstract stimuli. Therefore, if

anything is to be said concerning the presence or absence of group differences on the test trial latencies there is a need to compare each group's naming latencies on the actual experimental stimuli.

(2) The second assumption lies in the notion that the subjects are naming the experimental stimuli throughout the test session.

So far, the only evidence available concerning this issue has been gleaned from subjects' post-experimental remarks which can be notoriously unreliable as they require subjects to verbalise strategies which may not necessarily have been occurring during the actual experiment. In addition, the subjects' patterns of responding were also examined. For instance, if the trials which involve transitivity take significantly longer this could reflect the underlying control of the subject's behaviour (for example, the associative distance effect, Bentall et al 1993); or subjects assigning individual names to the stimuli (Dugdale and Lowe 1990; Horne and Lowe 1996).

It can be seen that there is only correlational evidence to support the hypothesis that subjects who name the stimuli successfully demonstrate stimulus equivalence. Therefore, what is needed is a protocol which ensures that the subjects are consistently assigning names to the stimuli throughout the matching-to-sample sessions.
In order to empirically test the two assumptions set out above this third study adopted a 'name-aloud' procedure whereby the

subjects were shown the experimental stimuli beforehand (randomly, that is, not grouped according to the equivalence classes) and asked to assign names to them. Then, during the matching-to-sample trials each subject was required to say his/her name for the stimuli out loud before selecting them as samples or comparisons. In this way it was ensured that the subjects were naming the stimuli consistently throughout the test session. Also, a post-experimental test could be implemented to compare the subjects' naming speed using the experimental stimuli.

The procedure adopted here enforces on the subject what, according to the naming hypothesis, should naturally occur in the test session if the subject is to successfully demonstrate equivalence. That is, the subject has to come up with their own name for the abstract stimuli, produce these names consistently throughout the session and if they are individual names, link them intraverbally or via a verbal rule. Adding naming to the procedure in this way should, according to the naming hypothesis, enhance the pass rates on the tests of equivalence. However, another scenario could be that the dyslexic subjects do not rely on naming to pass the equivalence test. This might be why no significant group differences have been observed so far on the test session latencies. If that were the case then, requiring the dyslexic subjects to name the stimuli might even be detrimental to their performance.

Apart from these changes the procedure was almost identical to that employed in Experiment 2.

## Method

**Subjects** (see Appendix G for full subject details)

Eighteen subjects took part in this study (nine dyslexic subjects and nine control subjects). The dyslexic group consisted of two females and seven males, whereas the control group comprised of eight females and one male[1].

The age range for the dyslexic group was 18-55 years (mean=27.56 years) and for the control group 19-45 years (mean=27.78 years). All subjects were undergraduate students at the University of Wales, Bangor. Any psychology students received course credits for participating in departmental studies providing that a written report of their experiences as a subject was submitted. The dyslexic subjects were volunteers recruited from the student dyslexia panel. This meant that they had been previously contacted via the Dyslexia Unit, were asked whether they were willing to be approached with the view of taking part in experimental studies, and subsequently allowed their name and address to be stored on a confidential database.

No subject possessed any prior knowledge concerning stimulus equivalence theories or methodologies.

1This uneven distribution of males/females was unavoidable in this study for two reasons: (i) the control subjects were recruited from the School of Psychology's student subject pool which meant that certain procedures had to be adhered to. The subjects themselves, 'signed up' for a particular experiment on a first come first served basis, until all the slots stipulated by the experimenter were filled. Therefore, the male/female ratio was hard to determine exactly. (ii) This experiment was chronologically the last of the five studies to be undertaken. This meant that dyslexic volunteers who had not previously taken part in any other of the studies were becoming scarce . Therefore, any dyslexic volunteers no matter what gender were gratefully included in the experiment. However, it was not thought that a gender difference between the groups would unduly affect the experimental hypothesis.

Subject selection - the dyslexia criterion used in this study was the same as previously adopted but procedures were tightened up further. In the previous studies, all subjects assessed in Bangor received the Bangor Dyslexia Test (Miles 1982) and the Schonell spelling test (Schonell and Schonell 1952) as part of their assessment (and this was by far the majority of the sample), but not all subjects had undergone diagnosis in Bangor, so the same could not be said with certainty of them. As a result it was decided to verify (for the purpose of the experiment) any previous diagnoses made by stipulating that all subjects (dyslexic and non-dyslexic) should be tested on the Bangor Dyslexia test; the Schonell spelling test; and on the Advanced Progressive Matrices (Raven, 1962). The matrices task was added as a non-verbal intelligence test to ensure that the two groups were matched for intelligence (previously it had been assumed because the two groups consisted of all students; or of people from similar educational backgrounds). It is usual when diagnosing developmental dyslexia to contrast poor reading or spelling scores with average or above average intelligence. This is to demonstrate that the below average reading/spelling performance is not a result of some general cognitive shortcoming.

Miles (1994b) makes the case for measuring a dyslexic individual's intelligence as it is often cited as being one of the indicators of dyslexia (for instance, dyslexics are often noted to be of above average intelligence) and should therefore, not be overlooked. However, he stresses the importance of using an appropriate measure of intelligence. Standard intelligence tests (for example, the Wechsler Intelligence Scale for Children - Wechsler 1976; or the British Ability Scales - Elliott, Murray and Pearson 1983) contain items which would be problematic for the dyslexic participants. More specifically, the items which involve verbal processing or a short-term memory load (note for example, the ACID - Arithmetic, Coding, Information and Digit - profile of subtests the dyslexic would have difficulties with on the WISC) would lead to a lower overall score. Miles argues that there is a need for, "an uncontaminated measure of intelligence" (p.75), and says this can be achieved by looking at reasoning ability where the verbal requirement is at a minimum. Such a test which is appropriate for adults is the Advanced Progressive Matrices (Raven, 1962) where subjects are required to analyse patterns and choose the missing component from a choice of eight possible contenders. Miles (1993) claims that this is a more accurate measure of intelligence in the case of dyslexic subjects because, "What is being tapped is the skill at reasoning - the ability to recognise relationships - and not skill at finding the right words for communicating one's ideas to others" (p.230).

The test was administered as follows: each subject is firstly given a booklet containing twelve examples and has about five minutes to work through as many as possible in the time available; the answers are then checked and any discrepancies are addressed. Next, set two, comprising of thirty-six problems, is presented to the subject. Standardised instructions are read to the subject and responses are made by the subject noting her/his choice on a blank piece of paper.

Subjects are given forty minutes to work their way through as much as they can.

To recap then, the dyslexic group was required to demonstrate the previously outlined dyslexic profile on the Bangor Dyslexia Test and the Schonell spelling test but not to differ from the control group on the reasoning test. This was found to be the case. No control subject scored more than 3.5 dyslexia positives on the Bangor Dyslexia Test (mean=1.17); the mean for the dyslexic group was 4.89 positives. One dyslexic subject (Dyslexic 26) scored only 3 positives, but he was still included in the sample due to his previous diagnosis of dyslexia. The mean score on the Schonell spelling test was 20.89 out of 30 for the control group and 13.56 for the dyslexic group. Five control subjects scored below 20 on the spelling test but as these were coupled with a low number of dyslexia positives, the scores were put down to poor spelling and nothing more. One dyslexic subject (Dyslexic 23) scored over 20, this was however alongside a score of 5+ves on the Bangor Dyslexia test and a previous diagnosis of dyslexia. On the matrices task the dyslexic group scored a mean of 22.3 out of 36 and the control group 24.3. The norm score for university students cited by Raven (1962) is 21 with a standard deviation of 4. All subjects fell within this distribution, except Dyslexic 19. The norms however cited in Raven (1962) are for subjects under the age of 25 years; as Dyslexic 19 was 44 years old and her score was only 3 points outside she was kept in the sample.

As in previous studies all subjects were informed of their rights as a subject and asked to sign a consent form and to fill out a pre-experimental questionnaire which concerned details such as age, occupation, psychology background.

**Apparatus and Materials**

Identical stimuli, programmes, and equipment were used as in Experiment 2 for the identity matching test, experimental matching-to-sample task, and the post-experimental interview.

A short session was introduced before the matching-to-sample task and a discrete-trial naming test was implemented after the post-experimental interview. Both sessions were written and presented on SuperLab 1.5.9 Beta and utilised the Hebrew font 72 point. For the naming test the subject's vocalisations activated a voice onset key in SuperLab. This was channelled through the LCii's external microphone which was fixed to the front of the monitor directly below the screen. In Experiment 2 no differences were found between the vocalisation onset measure and the total articulation time measure (they demonstrated the same group differences and showed the same pattern of responding), therefore, just the vocalisation onset times were measured in this study.

Also in Experiment 2 the subjects' median naming latencies were calculated using a small number of data points (nine at the most for each subtest). In order to produce a more representative median it was decided to introduce more instances of each stimuli (that is the nine stimuli were repeated five times at random). This was also in an attempt to replicate the RAN style test (as described in Chapter 2 - where subjects are required to name stimuli repeated 10 times, as quickly as possible, and then an overall time is measured). The discrete-trial format was still adhered to but an overall median of 45 naming episodes was calculated for each subject to produce a more representative measure of central tendency. All sessions were videotaped.

**Experimental Design**

This was on the whole identical to that undertaken in Experiment 2, with the exception of the addition of the 'name-aloud protocol' to the matching-to-sample sessions.
The study was run over two sessions. The first session consisted of: The Bangor Dyslexia Test; the Schonell Spelling test; and the Advanced Progressive Matrices. This lasted approximately sixty minutes. The second session comprised of: the identity matching test; the experimental task; the post-experimental interview; and the discrete-trial naming test. This session lasted between one and two hours

**Procedure**

Identity Matching task - this consisted of identical procedure, stimuli, and instructions as implemented in Experiment 2.

Experimental Task - the same abstract, pictorial stimuli (Hebrew font - see Figure 14) from Experiment 2 were used, and formed the same three equivalence classes. However, prior to the running of this task the subjects were presented with the nine experimental stimuli individually (in a random order) and were asked to assign to each a name which could be easily recalled. The only stipulation was that a different name must be assigned to each stimulus. Subjects' names were recorded manually by the experimenter for reference throughout the experiment. Subjects were then tested to ascertain that they could recall the names reliably. Specifically, the nine stimuli were presented randomly until the subject was able to recall correctly the names of all the stimuli twice.

<u>Phase 1: Train AB and BC relations</u> - the parameters of the experiment were the same as in Experiment 2. The only difference was the instructions read to each subject. These were as follows:

> In front of you is the computer which you are about to use. You will respond by using the four black buttons at the end of the keyboard (experimenter points to where they are). When the experiment begins a visual image will appear on the left of the screen in the same position as this button here (experimenter shows). You must first of all say its name aloud and then in order to carry on you must press this button. Three more visual images will then appear on the right of the screen in the same position as these three buttons here (experimenter shows). When these appear you must choose one, say its name aloud, and then press the corresponding button. If you select the right image, the word, 'correct' will appear on the screen. If you select the incorrect one, the word, 'wrong' will appear instead. Please try and get as many correct as you can. Do you understand or do you want me to repeat the instructions?

In other words, the subjects were instructed to 'name-aloud' the stimuli as they selected them.

Phase 2: AB and BC without reinforcement

Phase 3: Test symmetry (BA & CB), transitivity (AC), and equivalence (CA)

These were identical to Experiment 2 with the exception of the instructions given at the beginning of each phase. These were as follows:

> The procedure is exactly the same as before, but this time the computer will not tell you whether your response is right or wrong. Remember to keep saying the names aloud before you press a button. You will be told at the very end how well you have done. As before try and get as many correct as possible.

In this way, the subjects were reminded to continue to name the stimuli throughout the sessions.

Post-experimental Interview/debriefing Session - the subjects were presented with examples of the trials they encountered during the training and test phases. They were asked by the experimenter to recall how they matched the pairs together; what strategies they used; what went through their minds; and whether assigning names to the stimuli beforehand helped or hindered them.

Discrete-trial Naming Test - subjects were finally presented with the nine experimental stimuli in discrete-trial format and asked to produce their names for them. Forty-five trials were presented consisting of the nine stimuli repeated five times randomly. The following instructions were read to each subject:

> You will be shown the nine experimental stimuli repeated randomly five times each. As each appears on the screen, you must say its name out loud as fast as possible. Try and speak loudly and clearly. Any questions? Please say the names as quickly and as accurately as you can. Are you ready?

## Results

### Naming Test

A median naming latency (over 45 trials) was calculated for each subject. The mean latency for the dyslexic group was 786.11ms (sd=115.81) and for the control group 766.89ms (sd=96.02). The data were compared using an unpaired t-test and a non-significant difference was demonstrated ($t=-0.383$, $df=16$, ns).

### Identity Matching

The groups were firstly compared on the nonverbal identity matching task. The median response latency was calculated for each subject, and then a group mean. The control group produced a mean of 1082.33ms (sd=270.69) and the dyslexic group a mean of 1428.28ms (sd=383.10). These data were analysed using an unpaired t-test which showed a significant difference ($t=-2.12$, $df=16$, $p<0.05$).

### Training Sessions

As in the previous experiments the total number of errors made to criterion; the number made on the unreinforced sessions alone; and the number of repeat training sessions required was calculated for each subject.

**Table 8: The mean number of errors per group during; Phase 1, train AB and BC; Phase 2, AB and BC without reinforcement; and the mean number of repeat training sessions required by each group. Standard deviations are given in brackets.**

| | Phase 1 Number of Errors to Criterion | Phase 2 Unreinforced Errors | Repeat Number of Sessions |
|---|---|---|---|
| **Dyslexic Group** | 17.11 (18.39)) | 1.22 (2.95) | 1.33 (1.87) |
| **Control Group** | 9.33 (7.68) | 0.78 (1.64) | 0.78 (0.97) |

Table 8 shows that the dyslexic group made more errors before reaching criterion. Looking at the dyslexic group standard deviation it can be seen that the sample distributions differ. An F-test confirmed that the variance of the two groups differed significantly ($F=5.735$, num.df=8, den.df=8, $p<0.05$). Therefore, a Robust Rank-Order Test was performed. This produced a non-significant difference ($U`=0.74$, m,n =9 ns).
Table 8 shows that on average there was little difference between the groups as measured on the number of errors made during the unreinforced sessions. This was confirmed using an unpaired t-test ($t=0.395$,df=16, ns).
The same applied to the number of extra training sessions required. Once again a non-significant difference was observed ($t=0.791$, df=16, ns).

**Test Session**

The test trials were divided into the same categories as applied in experiments 1 and 2 (*baseline, symmetry, and transitivity).*

Errors made during the test session

The pass/fail criterion adopted was identical to the one used in Experiment 2.

**Table 9: Pass/fail rate for each group**

| | Pass | Fail |
|---|---|---|
| **Dyslexic Group** | 4 | 5 |
| **Control Group** | 6 | 3 |

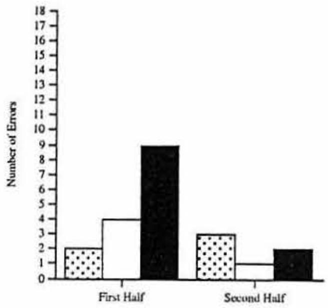
It can be seen from Table 9 that once again more control subjects successfully demonstrated equivalence. This difference, however, was non-significant as measured by a Fisher's Exact test (for small sample sizes), (phi=0.224, df=1, ns).

Once again the subjects who passed made a minimal number of errors. The control group made no more than a mean of 7% errors on any category and the dyslexic group no more than 8%.

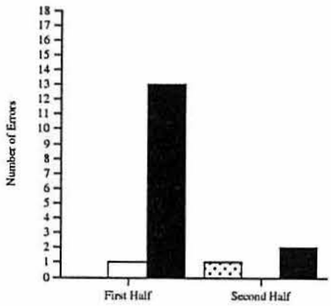
The distribution of errors made by the subjects who failed (n=3 controls; n=5 dyslexics) was examined (Figure 19). The subjects who passed made very few errors. Here the test session is broken down into two equal halves. For two subjects (Control 24 and Dyslexic 22) performances improved

**Figure 19:** **Graphs illustrating the number of errors made by the subjects who failed, on each category, on each half of the test session in Experiment 3.**

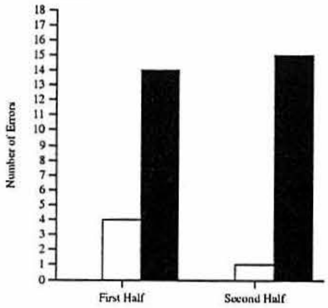
Control 24



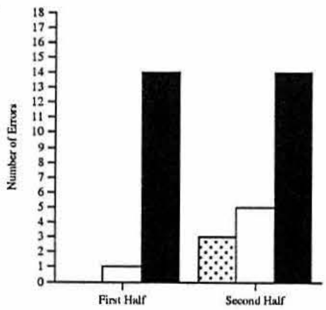

Dyslexic 22



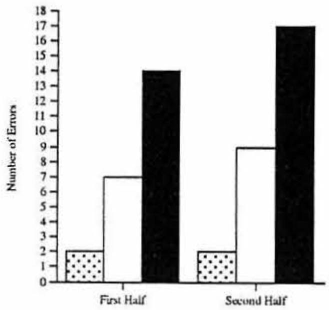

Control 25



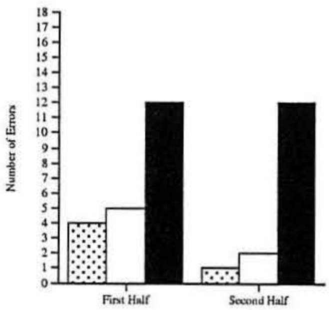

Dyslexic 23




Control 26



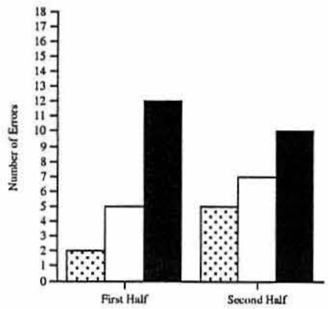

Dyslexic 24



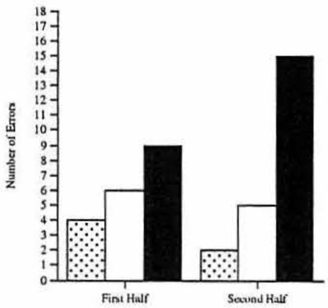

Key

- Baseline
- Symmetry
- Transitivity

Dyslexic 25




Dyslexic 26



on the second half of the session. For the remaining subjects on the whole their performances (especially on the transitivity tasks) remained constant over time, with some performances getting worse on the second half. No subject showed evidence of baseline deterioration on this task.

Therefore, just 2 control subjects compared to 4 dyslexic subjects persistently failed to demonstrate equivalence, again showing that there is a trend for more dyslexic subjects to fail more dramatically.

Test trial response latencies

For each subject who passed (n=6 controls; n=4 dyslexics) a median response latency was calculated for each test session category.

**Table 10: Mean of median response latencies for each group on the three test session categories (standard deviations are presented in brackets)**

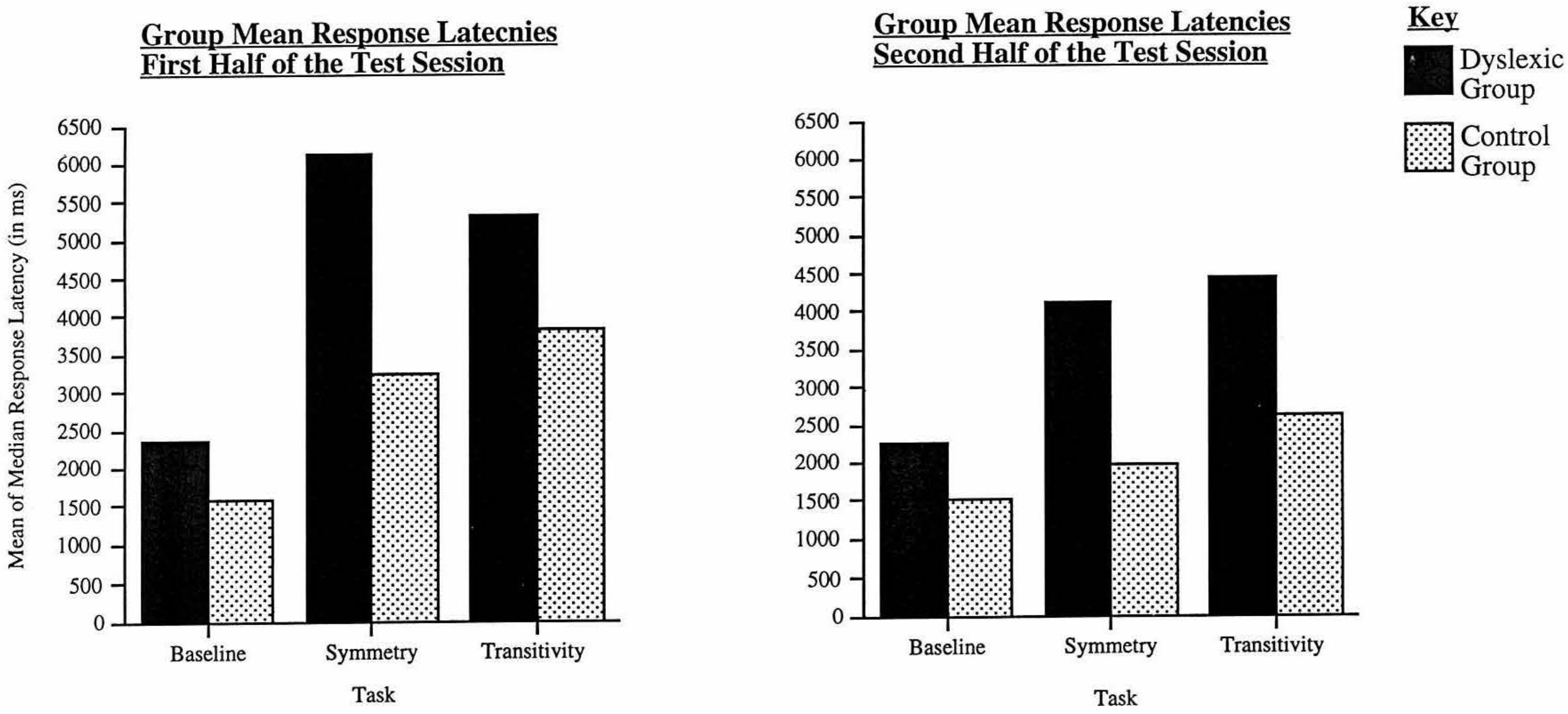
| | Baseline | Symmetry | Transitivity | Total |
|---|---|---|---|---|
| **Dyslexic Group** | 2356.50 (1027.25) | 5242.62 (4294.42) | 4655.00 (2162.73) | 4084.71 (2878.36) |
| **Control Group** | 1559.00 (584.66) | 2341.75 (876.14) | 3003.83 (1227.18) | 2301.53 (1067.00) |
| **Total** | 1878.00 (843.36) | 3502.10 (2969.48) | 3664.30 (1767.15) | |

From Table 10 it can be seen that on all categories the dyslexic group took the longest. Both groups produced the shortest latencies on the baseline trials. The control group produced the next longest on the symmetry tasks, followed by the transitivity tasks. The dyslexic group, however, took longer on the symmetry trials than on the transitivity trials (note that, this must be viewed with caution. This mean constituted only 4 data points one of which at 11339.5ms (Dyslexic 21) may have artificially skewed this mean). With this datum omitted this mean diminished to 3210.33ms which is below that for the transitivity tasks.

A 2X3 mixed design ANOVA was performed on the data and showed a non-significant difference between the two groups (F=2.792, df=1,8 ns). No significant interaction effect was demonstrated (F=2.147, df=2,16 ns). A significant difference was observed, however, between the test categories (F=7.801, df=2,16 p<0.005). Further analysis employing Tukey's HSD tests (q (3,16) = 1290.89 at the 0.05 level of significance) demonstrated that there was no significant difference between the symmetry and transitivity tasks but the baseline trials were significantly shorter than both the symmetry and transitivty trials.
This pattern was different to the one observed in Experiments 1 and 2, therefore, it seemed possible that the dyslexic subject 21's high symmetry median may be inflating the data, so the same analysis was performed without his data (F=15.364, df=2, 14, P<0.0005). The resulting pairwise comparisons, however, identified the same pattern as above: baseline < symmetry/transitivity (q (3, 14) = 751.65 at the 0.05 level of significance).

Graphs of each group's median response latencies over the two halves of the session were also drawn (Figure 20). Not only do

**Figure 20:** **Graphs comparing each group's mean of median latencies on the first and second half of the test session in Experiment 3.**


Group Mean Response Latecnies
First Half of the Test Session
Mean of Median Response Latency (in ms)
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Baseline
Symmetry
Transitivity
Task
Group Mean Response Latencies
Second Half of the Test Session
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Baseline
Symmetry
Transitivity
Task
Key
Dyslexic Group
Control Group

these highlight any potential effects which may not be detected due to small sample sizes, but also establishes whether the groups' response latencies differed over the course of the session. From Figure 20 it can be seen that each group's latencies drop over time apart from on the baseline tasks which remain stable for both groups. Large differences were seen between the groups on the symmetry tasks and to a lesser extent on the transitivity tasks. This was possibly due to Dyslexic subject 21's data which inflated the dyslexic group's mean. The difference between the groups (which is the variable of interest) remains stable over both halves of the session.

The data of the subjects who successfully demonstrated equivalence was split according to their performance on the naming test and two groups were identified: fast namers, n=5, mean=685.8ms; slow namers, n=5, mean=799.4ms. These two groups were found to significantly differ on the naming measure ($t=-3.713$, $df=8$, $p<0.005$). Finally, a comparison was made between the two group's response latencies on the test session trials. No significant group effect was observed ($f=0.132$, $df=1,8$, ns).

## Discussion

A discrete-trial naming test was devised which measured the subjects' naming latencies on the stimuli employed on the Experimental task. Although the mean latency for the dyslexic group was slightly larger than that of the control group the difference was non-significant.

So, no significant differences between the dyslexic and control group on the discrete-trial naming test have been found using familiar stimuli (see Experiment 2) or novel, newly-acquired stimuli (see Experiment 3). In the latter instance this was thought to be a novel task for the subjects and one for which the dyslexic group could not have compensated. The stimuli themselves were novel, but the names assigned to them by the subjects were not. Even though the subjects chose certain names because the stimuli were thought to resemble real-life objects/figures they still had to recall the resemblance. To all intents and purposes the task was similar to that of picture naming but not so well-practised; a task on which dyslexic/non-dyslexic naming differences were observed in childhood (Denckla and Rudel 1976b) yet importantly not here with adults.

Therefore, no naming differences were observed and thus equivalence test session latency differences should not be observed if the same kind of naming is involved. However, according to the naming hypothesis the subjects who pass should be naming the stimuli intraverbally which is a more complex task than the single stimulus presentation which occurs in the discrete-trial naming test, so significant differences might still occur on the test session latencies.

A significant difference was found between the two groups on the Identity Matching task using abstract, pictorial stimuli ($t=-2.212$, $df=16$, $p<0.05$). So, here is a nonverbal task on which the two groups significantly differ. In the previous study the same task was employed and no significant differences found, so it is likely that in this instance the dyslexic group contained generally slow responders. If this is true, however, then the dyslexic group should be equally slow on the other measures. It has already been shown that the dyslexic group although slower was not significantly slower on the discrete-trial naming task, so the difference was not reflected there which would suggest that this is just an anomalous finding.

No significant differences were found between the two groups on any of the training measures, although the dyslexic group did make more errors before reaching criterion during the sessions. Notably, in comparison with Experiment 2, fewer errors were made and the subjects required fewer sessions to maintain the baseline relations. This was chiefly due to the dyslexic group's performance being greatly improved using this 'name-aloud' procedure. Therefore, imposing naming on the subjects did not adversely affect the dyslexic subjects' performances in any way perhaps because this procedure required the subjects to name consistently whereas in the previous studies there was no such requirement. In Experiment 2 it was concluded that the dyslexic subjects' impaired performance on the training session was compatible with the theory that the subjects were naming inconsistently. In this study the subjects were, without question naming reliably and their performances improved.

On the test session, two more control than dyslexic subjects successfully demonstrated equivalence. Although this difference was not significant it does perhaps suggest a trend of more dyslexic subjects failing when the task involves naming.

Looking at Figure 19, two subjects' performances improved notably over time (Control 24; and Dyslexic 22). The performance of the remaining subjects was stable or got slightly worse on the second half of the session. In other words, there was no sign of any learning taking place during the test session. Therefore, 4 dyslexic subjects versus 2 control subjects persistently failed. So once again there was a trend whereby the dyslexic subjects failed more dramatically. There was no evidence of baseline deterioration in any of the subjects' test session performances, so the relations learnt in training remained robust despite the introduction of the novel stimulus combinations. In the previous studies some amount of baseline deterioration had been seen suggesting that the, 'name-aloud' protocol has made the learned relations more stable and also meant that fewer errors to criterion were made during training.

Three subjects (Control 26; Dyslexic 25; and Dyslexic 26) did not reach criterion on the symmetry trials despite the fact that there was no evidence of deterioration on the reverse of these relations (that is the baseline trials). This suggests that the relationship between the tasks was not recognised by these subjects. This finding was peculiar to this study and suggests that if subjects name-aloud then the relations can become rigid and unidirectional (that is they are not obviously reversible as it is not a familiar task to reverse word-word relations) because of the naming imposed on them. So, in this respect intraverbal naming sequences may impede stimulus equivalence which goes against the assumptions of the naming hypothesis which states that intraverbal naming sequences should facilitate equivalence formation.

Looking at the subjects' equivalence test session response latencies in the subjects who passed the equivalence test, no significant group differences were observed despite the fact that overall the dyslexic group was 1783.18ms slower. The dyslexic group was notably slower on all three test categories, especially the symmetry task but this particular mean was inflated by one subject (Dyslexic 21). With his data removed the difference was minimised.

Therefore, no significant differences were observed between the groups on the discrete-trial naming test and this was reflected by again no significant differences being found on the test session trials where it was known that the subjects were naming the stimuli. This suggests no significant naming difference existed between the groups for these particular stimuli.

However, two groups were identified which did significantly differ according to their naming latencies but again no significant difference was found between the fast/slow groups on the test session latencies. The same analysis was performed on the data of the subjects who passed in Experiment 2, similarly no significant was seen between the groups on the test session trials, but the two groups did significant on the nonverbal identity matching measure meaning that one group may just contain generally fast/slow responders. The above fast/slow groups did not differ on the identity matching measure (t=0.168, df=8, ns) so the same caution need not be applied here.

So, two groups have been established who do actually significantly differ in naming speed, who name intraverbally during the test session, but who do not significantly differ on the test trial latencies. This suggests that the naming measure on which they differ is not analogous to the type of naming which occurs on the typical matching-to-sample test session trial. On the other hand, the significant naming latency could be swamped on the test session trial by other cognitive processing (such as, reasoning or attentional factors).

The results of this experiment present a potential problem for the naming hypothesis. It states that if the subjects are intraverbally naming the stimuli consistently then equivalence between the stimuli should emerge. This was the case for all the subjects in this study yet eight subjects still failed to demonstrate equivalence.

Why should this be the case? The post-experimental remarks of the subjects who failed were examined. Two of the subjects who failed (Control 24; and Dyslexic 22) knew what the task required, and this is reflected in the fact that their performances dramatically improved on the second half of the session. It cannot be established, though, whether the strategies reported occurred at the time of testing or came about as a product of testing.

Three of the remaining subjects who failed (Control 25; Control 26; and Dyslexic 23) reported that they believed that their task was to remember the trained relations and that the novel stimulus combinations were presented in order to interfere with their learning. As a result, they deliberately matched new stimulus pairs in a way which could be easily recalled but which would not interfere with the relations they had previously learned (for example, they put together pairs which began with the same initial letter or pairs which rhymed). In other words, they had totally misconstrued the task requirements.

Two of the dyslexic subjects 24 and 25 said that they concentrated on the sounds of the words and put together pairs of words which sounded right together. Dyslexic 24 said that he remembered the trained relations by the sound of the two names together and not by what the stimuli looked like. As a result he did not realise that the symmetry trials were the baseline relations reversed. In this case naming interfered with the emergence of equivalence. Dyslexic 26 said that he was matching the stimuli visually and assigning names to the stimuli had hindered him as they confused him because some of the letters he had chosen as names sounded alike.

It could be argued that the actual protocol (specifically requiring the subjects to assign names to the stimuli) resulted in some of the subjects failing. Stipulating that the subjects had to choose names for the stimuli may have placed emphasis on the name and the task of remembering the names as opposed to how the stimuli were related to one another. This could explain why some of the subjects thought that the purpose of the task was to remember the trained relations at all costs.

This seriously challenges the naming hypothesis in that the bidirectionality of the intraverbal name relations is not sufficient to bring about equivalence between stimulus-stimulus relations as predicted. Lowe and Beasty (1987) reported that all their subjects who repeated the names of the baseline stimulus pairs went on to pass the test of equivalence. The subjects in this study did exactly the same but 8 subjects failed to demonstrate equivalence. So, merely repeating the names intraverbally is not enough for equivalence to emerge.

However, realistically, these conclusions must be tempered when the notion of the statistical power of this study is considered, given that the sample sizes employed in this study were relatively small. No significant difference was found between the groups' test session latencies. Yet looking at the group means in Table 10 it can be seen that the mean difference is large in cognitive processing terms (1783.18ms). So, there does appear, on visual inspection, to be an effect there, so it could be that the experimental design is not statistically powerful enough to detect a difference. Put another way, if the design has low power then there is a high probability of making a Type II error; that is not detecting a significant difference when there actually is one present.

Cohen (1988, 1992) presents a methodology whereby the power of any given experiment can be calculated. Conversely, it is also possible to determine the exact sample size needed in order to establish a high level of experimental power. The premise being that the larger the sample size, the more likely it is that an effect will be detected. A power analysis was performed on the test session latency data from Experiment 3 (following the method presented by Howell, 1985) and it was found that given a sample size of mean n=5 per group, an effect size of 0.67, the power was 0.17. In other words, there was only a 17% chance of correctly rejecting a false null hypothesis.

One purpose of actually performing power analysis is to justify the adoption of the null hypothesis. That is if an experiment has a high level of power then it can be safely argued that the study was powerful enough to have detected an effect if there had been one. Experiment 3 had a low power level and therefore it can not be unequivocally determined whether the null hypothesis had been correctly adopted or not. As already mentioned one way to increase power is to increase the sample size (note that Experiment 3 had a small sample size and this probably is the reason for its low level of power. Cohen recommends setting the power level to be attained at

0.80 (a compromise between relatively high power and practicality in that 0.95 power usually entails massive sample sizes unless the effect to be detected is also extremely large) in the same way as historically the alpha significance level is set at 0.05 (95%). As a consequence, in order to have an 80% chance of correctly rejecting the null hypothesis in Experiment 3, forty-one subjects would be needed for each group. If no significant difference between the groups was still observed then it could be safely concluded that there is no notable effect.

So, because Experiment 3 possessed such low power it can not be argued unequivocally that there existed no significant group effect just because the null hypothesis was adopted. Alternatively, it must be noted that just because there existed a large difference between the means that it follows that this represents a large effect. It would appear large in cognitive terms where simple reaction time measures are usually employed. However, the matching-to-sample tasks involve complex response latencies such as scanning the multiple stimulus presentations, and making a choice reaction time. In such a context the difference may not be considered to be large.

Taking the data as a whole, a different pattern of responding was observed to the one shown in Experiments 1 and 2 (minus the data of Dyslexic 21), that being: baseline < symmetry/transitivity. Specifically, there is a distinction made between the baseline and symmetry tasks. This raises doubt over the assumption that the pattern, reliably found by

Bentall et al: baseline/symmetry < transitivity is compatible with subjects assigning individual names to the stimuli. In this experiment the protocol was such that it was certain that this is what the subjects were doing and yet a different pattern was produced which can not be explained by common naming (which would itself predict a flat pattern of responding) or the 'Associative Distance Effect' (Fields and colleagues) which would not predict a difference between baseline and symmetry tasks given they possess the same nodal distance.

What can be said is that a subject's pattern of responding on matching-to-sample is not as clear cut as predicted by Bentall et al (1993). These data suggest that more than two patterns of responding may be reliably seen. It could be that this different pattern is only seen when subjects apply intraverbal naming strategies. The symmetry task takes longer because it is not as practised a skill in that verbal humans are usually only required to say words in one order and not to reverse them.

Similarly, if this is the pattern demonstrated when subjects employ a verbal strategy this implies that subjects in Experiment 1 and 2 were not naming the stimuli and hence the different pattern.

## General Discussion

The predictions set out in Chapter 3, that the dyslexic group (which is chiefly characterised by phonological processing deficits) will significantly differ from the non-dyslexic group on the matching-to-sample task, if naming is necessary on the equivalence test session trials (as stipulated by the naming hypothesis), rely on two assumptions:

(1) that the two adult samples do demonstrate a significant naming latency difference;

(2) that naming is indeed necessary in order for adult subjects to successfully demonstrate equivalence.

Has any evidence been gathered from these first three studies to support either of these two assumptions?

The first issue concerns the dyslexic/non-dyslexic naming latency difference. Evidence was presented in Chapter 2 to suggests that dyslexic and non-dyslexic children differ in discrete-trial naming (Denckla and Rudel 1976b; Perfetti et al 1978; and Stanovich et al 1983) and there was evidence to indicate that this difference persisted into adulthood (Fawcett and Nicolson 1994; Watson and Brown 1992). As a result no measure of the subjects' naming speed was included in Experiment 1. No significant differences were found between the groups' test session latencies and this presented an interpretational problem. In other words, were no significant test trial latency differences observed because the two groups did not differ significantly in naming speed? Or was it because the dyslexic subjects were using an equally effective strategy to pass the equivalence test which meant that their performances were on a par with those of the control subjects? What was needed was an independent measure of naming, if any unequivocal conclusions were to be made regarding subjects' performances on the experimental task.

Experiments 2 and 3 employed such a measure. From Experiment 2 it was seen that the adult dyslexic and non-dyslexic subject groups did not significantly differ when compared on a discrete-trial naming test involving familiar stimuli, despite such differences being reliably observed in childhood, which is important given that Fawcett and Nicolson (1994) report that dyslexic naming differences continue to be found in adulthood. It was concluded, however, that from this non-significant difference it did not necessarily follow that a non-significant difference would also be observed in the naming of the experimental stimuli which were abstract, rather than familiar.

In Experiment 3 it was possible due to the 'name-aloud' procedure to measure the subjects' naming latencies for the actual abstract stimuli employed during the experimental task. Once again no significant difference was observed between the two groups when compared using a discrete-trial naming test.

These findings are of value because they suggest that dyslexic subjects' naming impairment (which is well documented in childhood) can and is (going by the data gleaned from Experiments 2 and 3) compensated for in adulthood. It is reasonable to believe that this would be the case for familiar stimuli (Experiment 2) as an individual would be likely to have had extensive exposure to such stimuli. Yet it is more surprising that no naming difference was observed for newly-acquired name pairs (Experiment 3) which could not have been as fully compensated for. It seems that a few hours of practice and no pressure of time (that is, discrete-trial presentation) is enough for the adult dyslexic subjects to perform as well as the control group.

The question which arises is whether the dyslexic naming impairment is fully compensated for under all conditions? In Chapter 2 it was reported that adult dyslexic and non-dyslexic populations were seen to differ significantly on RAN tests (Wolff, Michel and Ovrut 1990) which require the subjects to name continuous lists of stimuli under the pressure of time. It is possible that this procedure would place additional strain on the dyslexic subjects' naming resources and may not be as receptive to compensation.

Similarly, a deficit in naming in adult subjects may be dependent on the stimuli used. For example, Watson and Brown (1992) reported adult dyslexic/non-dyslexic naming differences when the stimuli used were words which could be argued to require more complex naming. This is in comparison to letter, number or picture naming where single, stimuli are presented rather than a sequence of letters, as in a word.

So, on one hand it can be said that from the evidence presented in Experiments 2 and 3 no significant naming differences were observed between adult dyslexic and non-dyslexic subjects and as a result the first assumption underlying the experimental hypothesis is not valid. Yet it does not necessarily mean that there exists no naming impairment in dyslexic subjects under all conditions. Evidence presented in Chapter 2 suggests that the deficit may only become apparent in adult subjects only under certain conditions which places strain on the subjects' naming ability either through time constraints, or the complexity of the stimuli presented.

The type of naming which may occur under matching-to-sample conditions does not merely require the subjects to name an individual stimulus under discrete-trial conditions. The most likely type of naming to occur in a visual-visual task (according to Horne and Lowe 1996) is intraverbal naming which involves pairs of stimuli being named together, which is an altogether more complex task than the one measured by the naming tests in Experiments 2 and 3. Strings of words (or continuous lists) are produced if the name relations are to become bidirectional (see Chapter 1) and this type of naming may be more analogous to the RAN test methodology.

So in summary, there is still a possibility that the two populations significantly differ in naming and importantly that this difference will be reflected in the test session trials if naming is occurring there. It can be argued that even though no significant group differences were observed on the naming measure it does not necessarily follow that no significant differences will be observed on the experimental task, if naming is indeed necessary for the successful demonstration of equivalence. What is needed is a more appropriate measure of naming speed to reflect more accurately the naming which is predicted by the naming hypothesis to occur on the equivalence test session trials. In addition, stimuli need to be presented for which the dyslexic group could not have compensated for.

The second assumption relies on the notion that, according to the naming hypothesis, naming is necessary for the emergence of stimulus equivalence. In order, to establish whether this was or was not a reasonable prediction the dyslexic and non-dyslexic group's performances were compared on the matching-to-sample tasks using abstract, pictorial stimuli. The hypothesis was that if naming was occurring during these tasks the dyslexic subjects' performances would be worse than those of the control subjects in all aspects of the experiment; specifically in errors made during training; the pass\fail rate on the test session; and the time taken on the test session trials .

Experiment 1 was procedurally more simple than Experiments 2 and 3 (only two comparison stimuli were presented to the subjects and the stimuli although abstract were fairly easy to discriminate) and as a result can not be directly compared to the other two. However, it was a most valuable study in that trends were detected in the data which shaped the design of the later experiments. Experiments 2 and 3, on the other hand, could be compared as they both employed identical stimuli and procedures. The major difference was that in Experiment 3 it was stipulated that the subjects must name the stimuli throughout the experimental sessions and this provided a useful comparison in that there was no doubt in this study that subjects were reliably naming the stimuli which is what the naming hypothesis stipulates is necessary if stimulus equivalence is to emerge.

The first matching-to-sample task on which the groups were compared was the nonverbal Identity Matching task. This measure was not implemented in Experiment 1, instead a simple, reaction time task demonstrated that the two groups did not significantly differ on this task. In Experiment 2 no significant difference was observed between the groups on the choice reaction time task whereas using the identical task a significant difference was found in Experiment 3. This significant difference was not reflected in a significant difference on any of the other latency measures. As a result, two views of this discrepancy can be taken: (i) the Identity Matching task is a nonverbal task whereas the other measures are proposed to contain a verbal element. So, these two groups differ on nonverbal measure, but this does not affect their performance on verbal tasks. (ii) Or this result, as it was not replicated in Experiment 2 was just an anomalous finding.

Looking at the data from Experiment 1, there was a trend during the training sessions for the dyslexic subjects to be more affected by the removal of feedback, which is in line with the predictions made in Chapter 3 where it was indicated that the dyslexic sample would be more reliant on feedback for an error-free performance. This may have been as a direct result of the subjects' naming which is more vulnerable to disruption or may signify a more general defect whereby dyslexic subjects are less confident in their performances regardless of the skills it entails. However, it is logical to assume that when a dyslexic individual is required to employ a skill which is known to be impaired, the subject would be more unsure about his/her performance. In Experiment 2 the dyslexic subjects found it harder to learn the baseline relations and as result made significantly more errors and there was a tendency for these subjects to require more training sessions. No significant differences were found between the groups during the unreinforced sessions so both groups were equally able to maintain the relations once they had been well learnt. In Experiment 3 no significant differences were found between the groups on any of the training measures.

So, a picture emerges as follows, in the first study where only four baseline relations were taught (2 AB and 2 BC) there was no difference between the groups in learning the relations but there was in maintaining them without feedback. When the task was made more demanding in Experiment 2 differences emerged between the groups in learning the relations. The obvious reason for this would be that this is a result of there being more relations to learn (3 AB and 3 BC) and the stimuli were more difficult to discriminate between or put another way, assign names to. These factors affected the dyslexic subjects more than the control subjects. The fact that the stimuli were difficult to name could have meant that initially the dyslexic subjects were reluctant to name them (as they know this is one of their weaknesses) and so tried to remember the stimuli, for instance, visually but found this to be an unreliable strategy and had to resort finally to naming which would lead to significantly more errors being made and a tendency for some of the subjects to need more training sessions.

Alternatively, the fact that the stimuli were hard to discriminate between may have meant that the dyslexic subjects found it difficult to assign a consistent name to them again resulting in more errors. However, this mere speculation as it could not be definitively said what the subjects were doing during these sessions. In Experiment 3 this was not a problem because the procedure explicitly required the subjects to name the stimuli consistently and when this happened no significant differences were observed between the groups on any aspect of training. In addition, the number of errors to criterion dramatically reduced as seen in Table 11.

Table 11: The mean number of errors to criterion made by each group during training. Standard deviations are given in brackets.

| | **Experiment 2** | **Experiment 3** |
|---|---|---|
| **Dyslexic Group** | 93.63 (87.95) | 17.11 (18.39) |
| **Control Group** | 21.63 (21.27) | 9.33 (7.68) |

Requiring subjects to name the stimuli aloud imposed on the dyslexic subjects, a consistent procedure which dramatically improved their performance, indicating that in Experiment 2 their problem may have been in finding a consistent strategy (as outlined earlier). So, it seems that the subjects do not have a problem learning the name relations (Experiment 3) but do in producing consistent names (Experiment 2), or indeed hitting upon a reliable strategy in the first place maybe because of their reluctance to use names. In Experiment 2 the majority of the dyslexic subjects post-experimentally reported that they assigned names to the stimuli, so they did eventually use naming but they required more training to be able to produce an error-free performance. Once the baseline relations had been learnt subjects in Experiment 2 and 3 showed no difficulty in maintaining these relations suggesting that in Experiment 1 the dyslexic subjects, although they were able to reach criterion had not had enough experience of the relations in order to maintain them without feedback. Whereas in in Experiment 2, and to a lesser extent in Experiment 3, because the dyslexic subjects were making more errors this resulted in them having more exposure to the relations which seemed to be beneficial to their performance.

How did the subjects' training performances affect their performances on the test session?

In Experiment 1 there was a trend for the dyslexic subjects to be more affected by the removal of feedback and this was carried through to their test session performance with a tendency for more dyslexic subjects to fail to demonstrate equivalence. Of the eight subjects who failed in this study, more dyslexic subjects persistently failed. This was despite the fact that the majority of dyslexic subjects reported that they had assigned names to the stimuli. This indicates that either the names were not consistent or that merely assigning names to the stimuli is not enough to bring about stimulus equivalence. Or that this was a strategy formulated post-experimentally. This was also the case for Experiment 3.

Therefore, the 'name-aloud' protocol seems to have helped the subjects to learn the initial relations (see Table 11) but does not seem to have facilitated the emergence of equivalence. This is reflected in the overall pass/fail rate. In Experiment 2, 5 dyslexic and 6 control subjects passed and 3 dyslexic and 2 control subjects failed; whereas in Experiment 3, 4 dyslexic and 6 control subjects passed; and 5 dyslexic and 3 control subjects failed. So, generally more subjects failed in Experiment 3 and specifically, this worsening in performance was greater for the dyslexic subjects. In Experiment 2, 37.5% of dyslexic subjects fail, whereas in Experiment 3, 55.5% of subjects fail a difference of 18%. In comparison, 25% of control subject fail in Experiment 2 and 33.3% fail in Experiment 3 a difference of only 8.3%.

The first issue that arises from these findings is why did more subjects fail when the 'name-aloud' protocol was adopted? It could be that the procedure of requiring the subject to name the stimuli aloud makes the baseline relations more rigid and less likely to be equivalence relations. This is possibly the result of years of language practice. For instance, in the English language sequences of words are usually unidirectional and so an English speaker has little practice in manipulating words or grouping them into equivalence classes. This would not be the case if the classes of stimuli were, for example, combinations of words, pictures, or sounds. Put another way, it is the name-name (response-response) relations which the subjects seem to have become focused on and not the stimulus-stimulus relations. Therefore, this evidence suggests that it is harder to demonstrate the properties of equivalence for name-name relations.

The other finding was that when the 'name-aloud' procedure was introduced proportionally more dyslexic subjects failed. Therefore, despite the fact that they are naming consistently they still fail. This is in line with the notion described in Chapter 2 that dyslexic individuals are less effective at manipulating verbal stimuli. In other words, when naming is imposed on the procedure the performance of the dyslexic subjects deteriorates. If this is true it suggests that not as many dyslexic subjects were naming in Experiment 2 (despite their post-experimental reports) and as a result not as many failed to demonstrate equivalence. Importantly with regard to the experimental hypothesis this indicates that naming is not always necessary for the successful demonstration of equivalence.

The overall finding from these three studies regarding errors made on the test session is that more dyslexic subjects persistently fail than control subjects. The fact that the dyslexic subjects made more errors on the test session indicates that (given the majority of dyslexic subjects from all three reported that they had assigned names to the stimuli throughout the experiments) they were the subjects who were inclined not to name as reliably or not to relate the names appropriately. The latter is an important point to reiterate. In Experiment 3 the subjects were effectively required to name the stimuli intraverbally (which according to the naming hypothesis should bring about equivalence) and consistently but this did not result in the demonstration of equivalence for eight subjects (5 dyslexic and 3 control subjects). So, it is not enough for subjects just to name the stimuli, something more is needed in order for equivalence to emerge. On the other hand, it could be that the subjects' verbal strategies have little to do with them passing or failing the equivalence test. It certainly did not facilitate the subjects' performances on the test session.

So there is a trend to suggest that the subjects differ on the pass/fail rates but what about on the test trial latencies which were hypothesised to reflect any naming latency differences between the groups if naming is occurring on these trials?

In Experiment 1 no significant difference was found between the groups on the test session latencies. Neither were any significant group differences found in Experiments 2 and 3, on either half of the session. Table 12 shows how introducing the 'name-aloud' protocol affected the data when compared to Experiment 2.

Table 12: The overall mean latencies made by the subjects who passed on the test session. Standard deviations given in brackets.

| | **Experiment 2** | **Experiment 3** |
|---|---|---|
| **Dyslexic Group** | 3227.53 (1613.55) | 3407.28[1] (2878.36) |
| **Control Group** | 3033.03 (2082.52) | 2301.53 (1067.00) |

1 This mean was calculated minus the data of Dyslexic 21, whose median latency skewed the symmetry mean.

The largest difference in latencies between the groups is seen in Experiment 3 with the dyslexic group taking longer than the control group when subjects were required to name the stimuli. So, this procedure has slowed the dyslexic group down but notably quickened the control subjects. This was despite the fact that no significant difference was observed between the groups on the discrete-trial naming test using the same experimental stimuli. This, as argued earlier, does not necessarily mean that this difference can not be attributed to a naming difference as the matching-to-sample task involves more complex naming than the discrete-trial naming measure. Especially given the fact that the dyslexic group were significantly faster than the control group on the Identity Matching task, therefore, despite being faster on the simpler, physical matching procedure, the dyslexic group are slower when the verbal element is introduced. In addition, it can be clearly seen how the 'name-aloud' procedure assisted the control subjects more than the dyslexic subjects whose performance seems to have been depressed by the introduction of naming.

Even though no significant latency differences were found further analyses were carried out in an attempt to establish whether there were any correlations between subjects naming speed and performances on the matching-to-sample tasks.
A naming latency was calculated for every subject who passed from each group. In Experiment 2 each subject's naming latency comprised of the mean of the numbers, letters, and pictures median. In Experiment 3 the naming speed was each subject's naming median over 45 discrete trials. Using

Spearman's Rank correlation coefficient naming speed was correlated with:

(1) The transitivity median to establish whether producing a slow naming speed correlated with producing a slow transitivity median (given that subjects who demonstrate equivalence should be naming here). This was chosen over the symmetry tasks as it was on the transitivity trials where the biggest difference was found between the groups. Experiment 2 showed a non-significant correlation (N=11, Rho=0.055, Z=0.172, ns); as did Experiment 3 (N=10, Rho=0.188. Z=0.564, ns). Therefore, no significant relationship was found between these two performances.

(2) The baseline median, where again the subjects should be naming, but which was a simple, more practised task. There was a non-significant correlation for the Experiment 2 data (N=11, Rho=0.009, Z=0.029, ns); similarly for the Experiment 3 data (N=10, Rho=0.127, Z=0.382, ns). Again there was no significant relationship.

(3) Finally the Identity matching task median. Would there be a correlation between a task where the subjects should not be naming? For Experiment 2 a significant relationship was found (N=11, Rho=0.864, Z=2.75, $p<0.005$); but for Experiment 3 there was no significant correlation (N=10, Rho=0.006, Z=0.018, ns).

Therefore, generally speaking naming speed was not reliably correlated with performance on any other tasks. More specifically, in the context of the experimental hypothesis, naming speed is not related to latency on the baseline or transitivity trials, such that a large naming latency was not reflected by a large transitivity latency. Put another way, the subjects who successfully demonstrated equivalence should, according to the naming hypothesis, be assigning names to the stimuli (as indeed they were in Experiment 3), yet their naming speed was not related to the transitivity trial latencies. This challenges the hypothesis that naming latency differences will be reflected in the test trial latencies. Or it possibly indicates (as argued earlier) that the two measures are not accessing the same type of naming, or not extracting the measure in the same context. This is further support for the implementation of a different naming measure which involves more complex naming and puts the subjects under more pressure, for example, the RAN test.

A significant correlation was observed in Experiment 2 where naming speed was significantly correlated with the nonverbal Identity Matching latency. However, a significant correlation was not observed in Experiment 3 using an identical task. The fact that there is a discrepancy between the two findings suggests that this may be an atypical result. If anything is to be said concerning the significant relationship it is that the two tasks of naming and identity matching could be viewed as relatively simple measures of reaction time whereas the baseline and transitivity tasks involve more complex processing.

The latency data from the three studies were also considered as a whole with respect to the within-subjects effects and whether they supported the patterns of responding found in the Bentall et al (1993) analyses.
In Experiment 1 the pattern was not as clear cut as the ones

identified by Bentall et al but the symmetry tasks took significantly less time than the transitivity tasks (the baseline tasks only just missed significance). In Experiment 2 both the baseline and symmetry tasks were significantly quicker than the tests which involved transitivity. So, there is evidence for a pattern in these data which Bentall et al ascribe to the response latencies of subjects who are not assigning common names to the stimuli. The authors conclude that this pattern can be explained by Field's, 'Associative Distance Effect' but it was argued in Chapter 3 that this pattern was also congruous with the subjects assigning individual names to the stimuli. Therefore the demonstration of these patterns could indicate that this is what the subjects in these two experiments are doing.

However, when the protocol changed for Experiment 3 to ensure that subjects were assigning individual names to the stimuli a different pattern was observed for the test session latencies of the subjects who passed (baseline < symmetry/ transitivity). For the first time a distinction is made between the baseline and symmetry tasks. This indicates that the trained, baseline relations are not obviously symmetrical which may be the case if the subject was focusing on the name-name relations and not the stimulus-stimulus relations, due to the unidirectionality of word-word relations.
This pattern is at odds with the one observed by Bentall et al and can not be explained by Field's, 'Associative Distance Effect' in that the number of nodes which separates the baseline and symmetry tasks is the same and therefore, should exert the same amount of control. It would suggest that if this

latter pattern is seen for subjects who are consistently employing a verbal strategy, then the pattern observed in Experiments 1 and 2 reflects that a nonverbal strategy is being employed by the majority of subjects. This implies that verbally competent adult subjects do not necessarily choose to assign names to the stimuli as is predicted by Horne and Lowe's hypothesis.

In conclusion, it was found that the dyslexic and non-dyslexic groups did not significantly differ on the discrete-trial naming test using familiar stimuli or novel, abstract stimuli which challenges the first assumption of the experimental hypothesis. It was, however, argued that these discrete-trial naming test may not reflect the same type of naming which is predicted by the naming hypothesis to be occurring during the experimental task. Therefore, there is a need for a different measure of naming such as the RAN test and different stimuli such as words where significant differences have been found in adult subjects using a discrete-trial format (Watson and Brown 1992).

It was seen during training that applying consistent names to the stimuli (Experiment 3) improved the performance of the dyslexic subjects but led to a deterioration in test session performances for all subjects but especially the dyslexic subjects. This is contrary to what is predicted by Horne and Lowe's naming hypothesis whereby intraverbal naming is only ever said to facilitate subjects' performances on the test session. The fact that the dyslexic subjects performed more badly during Experiment 3 reiterates the notion that these subjects suffer an impairment in naming and in the manipulation of verbal material. It was observed that the naming hypothesis needs qualification. Naming the stimuli is not sufficient to bring about stimulus equivalence. In some instances it can be detrimental to subjects' performances. No significant group differences were found between the test session latencies, although the 'name-aloud' procedure slowed down the dyslexic subjects (as predicted) and quickened up the control subjects. No relationships were found between naming speed and test trial latencies even in Experiment 3 were the subjects were definitely naming. This suggests that these two procedures are not measuring the same latent variable. Across the three studies two distinct pattern of responding have emerged. One (found in Experiments 1 and 2) were no distinction is made between the baseline and symmetry tasks and one (found in Experiment 3) where the baseline tasks take sifnificantly less time than the symmetry and transitivity tasks. It has been tentatively suggested that the former reflects a nonverbal underlying strategy and the latter a verbal one.

Further investigation is needed into the effects of subjects assigning consistent names to the stimuli on their matching-to-sample performances as it was when this was occurring that the largest differences between the groups were observed. In Experiment 3 the actual protocol employed may have affected the subjects' perceptions of the experimental task. Requiring the subjects to assign names to the stimuli beforehand and then asking them to say the names out loud throughout the experiment may possibly have led to an overemphasis on the name-name relations. It could be that this would not have occurred if the subjects' behaviour had been spontaneous. This was seen in the subjects' post-experimental remarks where some subjects thought that the purpose of the task was to remember the trained relations at all costs, or to put together pairs which, 'sounded right together'.

Therefore, there is a need for a protocol which implicitly rather than explicitly focuses on naming. In other words, instead of instructing the subjects to say the names aloud (which may shift the focus from the real nature of the task) it makes sense to present stimuli which are highly likely to be named spontaneously. For example, printed nonsense words.

# Chapter Five

## Experiment 4

Background

Method

Results

Discussion

## Experiment 5

Background

Method

Results

Discussion

## General Discussion

## Experiment 4

### Background

Experiments 1, 2 and 3 detailed one possible means of assessing the role of naming in the successful demonstration of stimulus equivalence, that being the presentation of abstract, pictorial stimuli. The rationale behind this was twofold:

(1) The stimuli are novel and thus have no obvious preestablished names/associations, and therefore the relations which come about are due to equivalence and no other previously learnt behaviour.

(2) The stimuli are visually difficult to discriminate between and thus name and therefore, this should accentuate the differences between the dyslexic and non-dyslexic subjects.

If subjects name the stimuli (which according to the naming hypothesis they must do in order to successfully demonstrate equivalence) then dyslexic/nondyslexic differences should be observed on the matching-to-sample performances.

Data from Experiments 1 and 2 illustrated that it was difficult to say definitively what the subjects' strategies were throughout the sessions and whether the subjects were or were not assigning names to the stimuli. No significant test-trial latency differences were found between the two groups but it could not be said that this was because, for example, the dyslexic subjects were not naming but instead using an equally effective strategy because there was no means of knowing exactly what the subjects were doing during the sessions. Therefore, it seemed that the key to this research was to ensure that the subjects were consistently naming. If the dyslexic subjects had been previously using a different strategy because of their naming weaknesses then imposing naming on them may lead to a detriment in performance which did appear to be the case (Experiment 3).

Also, ensuring that the subjects name the stimuli should lead to the successful demonstration of equivalence (i.e. the naming hypothesis), but was this the case? In Experiment 3, eight subjects failed to demonstrate equivalence despite naming the stimuli intraverbally. This suggests that simply naming the stimuli is not sufficient for equivalence to emerge. It was argued that the protocol involved in Experiment 3 over-emphasised the name-name relations (that is, naming the stimulus pairs aloud) instead of the actual relations between the stimuli.

Experiments 4 and 5 sought to explore these issues further. In Experiment 4 a different approach was taken. Naming was assured by virtue of the stimuli themselves being readily nameable, in that they were, 'wordlike'. Nonwords (for example, three letter consonant-vowel-consonant combinations) were therefore chosen. As these stimuli encouraged the subjects to read unfamiliar graphical combinations it was predicted that the dyslexic subjects (because of their phonological processing deficits) would find this task more difficult than the control group and would lead to the dyslexic/non-dyslexic differences being accentuated. The same format as in Experiments 2 and 3 was to be readministered but using nonsense words as stimuli.

The second issue which arose from the first three studies was that of the dyslexic/non-dyslexic naming difference. In Experiment 1 it was assumed (from the literature outlined in Chapter 2) that dyslexic individuals reliably demonstrate a naming deficiency, characterised by longer response latencies and more error-prone performances. It was hypothesised that this difference would be reflected in the test session trials if naming is taking place. In this first study no significant differences were seen between the two groups on the test session measures. Without evidence that the two groups significantly differed on naming speed this could indicate two things:

(1) The two groups are significantly different on the naming measure but not on the matching-to-sample task, therefore, it is questionable whether naming plays a role in the formation of equivalence classes.

(2) There is no significant difference between the groups on the test session because they demonstrate no significant naming differences.

Therefore, in Experiment 2 it was thought crucial to include a test to measure the magnitude of the naming difference between the two groups, if any useful conclusion concerning the test session latencies was to be made. A discrete-trial naming test using familiar stimuli was introduced. A discrete-trial methodology was chosen as it was believed to represent most closely the naming which would occur under matching-to-sample conditions. No significant differences were observed between the two groups. However, this did not necessarily rule out the possibility of a difference in naming being observed for the experimental stimuli.

A change in procedure (that is the subjects being required to assign names to the abstract stimuli) for Experiment 3 meant that subjects' naming using the actual stimuli used throughout the experiment could now be measured. Once again no significant group differences were observed using the discrete trial methodology. It was argued that it was possible that these measures did not represent the same type of naming which occurs on a typical matching-to-sample trial. Intraverbal naming should be occurring which involves continuous naming, as in the RAN test. It was also hypothesised that the introduction of more complex stimuli (that is words) could lead to the observation of group differences in adult subjects (Watson and Brown 1992) as these place more of a strain on the dyslexic subjects' naming resources. The fact that Experiment 4 employs nonwords as stimuli means that a measure of the subjects' naming latencies for the experimental stimuli can be included, and indeed the discrete-trial naming issue is pursued using these more complex stimuli.

In addition, a RAN test was introduced to assess whether dyslexic/nondyslexic differences are observed under these conditions. This should result in evidence to determine whether the dyslexic naming deficit, reported in childhood, can be compensated for (as in the discrete-trial naming) or whether it persists but is only revealed under specific circumstances, for example, when complex stimuli are used or the test is administered under the pressure of time. A comparison can also be made between the two methodologies to determine whether they both reliably measure the same latent variable (that is naming).

## Method

**Subjects** (see Appendix H for full subject details)

This study involved sixteen subjects (eight dyslexic subjects and eight control subjects). Each group was comprised of five males and three females. The age range for the dyslexic group was 18-35 years (mean 27.13 years); and for the control group 18-30 years (mean 24.37 years).

The dyslexic group consisted of members of the general public who had during their childhood received assessment at the University of Wales' Dyslexia Unit. They were recontacted via the director of the Unit and asked whether they would be willing to participate in a psychology study. Affirmative responders were then telephoned or written to by the experimenter and told in detail what the experiment entailed. These subjects were paid a lump sum of £7.50 to cover their expenses.

Subjects in the non-dyslexic group were recruited from the School's General Public Subject Pool which is a database of volunteers who are willing to be approached to act as subjects. According to the standardised regulations of the panel, each subject was paid £2.50 for each hour of their participation plus £1.50 travel expenses if any were incurred.

None of the subjects had any familiarity with stimulus equivalence theory or procedures.

Subject Selection

The same pretests employed in Experiment 3 were presented. That is every subject was given the Bangor Dyslexia Test; the last thirty items of the Schonell spelling test; and the Advanced Progressive Matrices. The latter test was of particular importance in this study as the subjects all came from differing backgrounds (that is, they were not all university students). So, the Matrices task served to establish that each subject produced an average or above average score on this test. On the Bangor Dyslexia Test the dyslexic group's scores ranged from 3.5 positives to 8.5 positives (mean=5.56 positives); the control group's scores ranged from 0 positives to 3.5 positives (mean=1.69 positives). Two dyslexic subjects (Dyslexic 31, and Dyslexic 31) scored only 3.5 positives but they also produced less than 20/30 correct on the spelling test. It could be argued that these are 'borderline' scores according to the experimental operational definition, but, it must be remembered that these subjects (and indeed all the dyslexic subjects from all five studies) had been previously diagnosed as being dyslexic at some time in their lives. As a result it can be reasonably assumed that any 'borderline' scores are a result of the subjects having compensated for their difficulties.

On the spelling test the mean score for the dyslexic group was 6.125; and for the control group, 21.37. No dyslexic subject scored over 20/30 but four control subjects (Control 27; Control 28; Control 31; and Control 34) did score under 20 correct, however, as these scores were coupled with less than 4 dyslexia positives, their low scores were attributed to a specific weakness in spelling.

On the Matrices task the dyslexic group scored a mean of

16.75; and the control group a mean of 24. The dyslexic group's scores were on the whole lower than those of the control group but six out of the eight subjects scored above 13 which according to the standardised norms (Ravens 1962) is in the expected range for their profession of technical/commercial workers - mean 18, standard deviation 5. The remaining two subjects (Dyslexic 34; and Dyslexic 32) scored 10 and 12 respectively, which is in the range for manual workers for their profession - mean 13, standard deviation 6. So, it can be said that all dyslexic subjects required at least average ability according to their job description. The same can be said of the control group, which is what the experimental criterion demands.

To summarise, none of the control group demonstrated any dyslexic signs; whereas all the dyslexic group did. Both groups showed average or above average ability on the Raven's Matrices task.

Two dyslexic subjects took part only in the first session of the experiment and were subsequently not included in the analysis. One was excluded because he failed to meet the subject selection criterion, and the other because he wished to be withdrawn from the study.

As in previous experiments all subjects: were informed of their rights as a subject; asked to sign a consent form; and filled out a pre-experimental questionnaire concerning details such as age, occupation, and psychology background.

**Apparatus and Materials**

The experiment took place in a quiet research room. The set up was identical to that illustrated in Figure 12.

The identity matching task was identical to that used in Experiments 2 and 3 with the exception of the stimuli used. In this instance, nonwords were employed as stimuli. These were as follows:

| | | |
|---|---|---|
| **BEP** | **GEQ** | **YEM** |
| **GOK** | **BOF** | **TOV** |
| **DAX** | **JAT** | **QAP** |

These were drawn in Times font, size 72 point. The stimuli were chosen to emulate the experimental stimuli, but present a task whereby the subjects were not required to name the stimuli. Matching can occur according to physical similarities alone, producing a nonverbal reaction time measure for similar stimuli to the ones used in the experimental task.

The equipment used for the Naming Test were also identical to those used in Experiment 2. The stimuli were the same too, but just five were used for each subtest in order to standardise the number of presentations in each subtest. They were as follows:

| **Pictures** | **Colours** | **Numbers** | **Letters** |
|---|---|---|---|
| Boat | Blue | 9 | d |
| Tree | Red | 7 | s |
| Fish | Yellow | 2 | o |
| Dog | Green | 6 | a |
| Bed | Black | 4 | m |

The pictures, numbers, and letters were taken from R. Nicolson's Naming test (see Fawcett and Nicolson, 1994). These were drawn in black (measuring 1"X1.5" - 2X3cm) and presented on a white background. A black border framed the screen. The five colours were drawn using Kidpix (1991 Craig Hickman and Broderman Software Inc. Version 1) and were in the form of a coloured circle (diameter 3" or 7cm) on a white background and a black border as before. All stimuli were presented in the centre of the screen.

A nonword discrete-trial naming test was also introduced in the same format as detailed above, but consisting of the nine experimental stimuli presented in the following order: MIB, ZIQ, YIM, ZEG, KEB, WEF, YOF, MOX, and KOJ. These were drawn in Times font, size 72 point.

The RAN Naming test consisted of the nine experimental stimuli presented in black on a white sheet of A4 (see Appendix I). The stimuli were randomly repeated five times (in a 9X5 grid) with the constraint that no two identical stimuli should appear consecutively. A standard microphone and tape recorder were used to record each subject's performance and a stopwatch to post-experimentally measure each subject's naming speed.

The format for the Experimental task was identical to that employed in Experiments 2 and 3. Only the stimuli differed. These were:

| | | |
|---|---|---|
| A1 | A2 | A3 |
| **KOJ** | **MOX** | **YOF** |
| B1 | B2 | B3 |
| **WEF** | **KEB** | **ZEG** |
| C1 | C2 | C3 |
| **YIM** | **ZIQ** | **MIB** |

These were taken from lists formulated by Glase (1928) and Krueger (1939) (Appendix A: in Underwood and Schulz, 1960 - "Meaningfulness and Verbal Learning") and were chosen for their low association and low frequency. The stimuli were arranged so that subjects could not form the classes according to the first, middle, or last letters alone (they were all sufficiently repeated to counteract this). All stimuli were written in Times font, size 72 point.

The Post-experimental Interview was the same as employed in Experiment 3 with the stimuli appropriately altered.

**Experimental Design**

The experiment was run over two sessions. Session one comprised of: the Bangor Dyslexia Test; the Spelling Test; the Advanced Progressive Matrices; the Discrete-trial Naming Test; the Nonword Discrete-trial Naming Test; and the Nonword RAN. This lasted one to one and a half hours. Session two consisted of: the Identity Matching task; the Experimental task; and the Post-experimental Interview. This lasted one to two hours.

The design of the Identity Matching task and the experimental matching-to-sample task was identical to that employed in Experiments 2 and 3.

**Procedure**

Discrete-trial Naming Test

This consisted firstly of a practice session. The 25 stimuli were divided into the four subtests with a screen title page separating each (for example, stating: "Pictures"; "Colours"; "Numbers"; and "Letters"). Each stimulus remained on the screen for 5000ms, the subject was required to say its name out loud. The following instructions were read to each subject:

> This is a practice session. You must say out loud the names of the following pictures, colours, numbers and letters. I will tell you whether you are right or wrong. Ready?

Oral feedback was provided by the experimenter stating whether each response was correct or not.

The actual Naming Test followed exactly the same procedure as the one employed in Experiment 2. The instructions were also identical, however, the stimuli were presented in a different order. In between each stimulus presentation there was an inter-trial interval (ITI) of 1000ms which consisted of a white screen. So, one trial proceeded as follows: ITI of 1000ms; the tone is sounded and the stimulus is presented simultaneously; the subject is required to name the stimulus as quickly as possible.

This was followed by the nonword discrete-trial naming test. The practice session consisted of the nine stimuli being presented (in a different random order to the actual test) for 5000ms and the subject was required to produce its name. The following instructions were given, "This is a practice session. You must read aloud the following nonsense words. I will tell you whether you are right or wrong. Ready?"
Oral feedback was given by the experimenter. Then followed the naming test which was procedurally the same as the previous discrete-trial naming test. The instructions were amended as follows:

> You will be shown the same series of nonsense words. As each appears on the screen you must read it out loud as quickly as possible. Try and sit as still as possible; don't lean towards or away from the mic. Try and speak loudly and clearly. This time a tone will sound when the next nonsense word is about to appear. Any questions? Please read aloud as quickly as possible. Are you ready?

<u>Nonword RAN</u>
The sheet containing the stimuli was placed in front of the subject. The experimenter held the microphone in front of the subject so that it was kept in a constant position. The instructions given were:

> You must read all the following nonsense words going from left to right (Experimenter shows the subject) and down the

page as quickly as possible. I will say, "go" after which you must start.
Are you ready? GO!

The whole task was tape recorded.

Identity Matching Task

Procedurally this was identical to the task in Experiments 2 and 3. The instructions were the same except the words, 'visual image' were replaced with, 'nonsense word' in line with the change of stimuli.

Experimental Task

This followed the same procedure as utilised in Experiments 2 and 3. The instructions given were the same as used in Experiment 2 with, 'nonsense word' replacing, 'visual image'. The instructions, "Please pay attention to the whole of the word as it will be helpful to you later in the session." were added to the end of the routine instructions in an attempt to counteract any obscure strategies.

Post-experimental Interview

This involved presenting the subjects with a representative sample of trials from each phase of the experiment. They were asked to describe what they did and what went through their minds during the sessions, or to report any strategy they used. If a subject successfully demonstrated equivalence they were asked how they connected the stimuli and whether they did so before or during the test session. This session was video recorded.

# Results

## Discrete-trial Naming Tests

This measure was first implemented in Experiment 2 where two measures were taken: (1) vocalisation onset; (2) total articulation time. No group differences were found on either measure; neither were any differences found when using one measure as opposed to the other. Both demonstrated the same pattern of responding. As a result, for economy's sake, just the vocalisation onset measure was utilised in this study.

The median of each subtest was calculated for each subject. Table 17 shows the group means.

**Table 13: Group means (presented in milliseconds) for each subtest on the Discrete-trial Naming tests (standard deviations are presented in brackets)**

| | Pictures | Colours | Numbers | Letters | Nonwords |
|---|---|---|---|---|---|
| **Dyslexic Group** | 1075.58 (109.92) | 985.71 (101.24) | 1045.35 (82.14) | 1064.78 (124.29) | 1261.05 (374.20) |
| **Control Group** | 1070.54 (121.83) | 1002.24 (102.78) | 1029.55 (115.65) | 1088.48 (196.46) | 1098.58 (152.91) |
| **Total Mean** | 1073.06 (112.12) | 993.98 (98.93) | 1037.45 (97.24) | 1076.63 (159.27) | 1179.81 (272.41) |
| ***Difference*** | *5.04* | *16.53* | *15.8* | *23.7* | *162.47* |

From Table 13 it can be seen that the biggest group difference was found on the nonwords subtest. The question which follows is, is this a significant difference?

The data from the Nonword subtest were analysed separately because it was administered separately to the others. A mixed

design, 2X4 repeated measures ANOVA was used to analyse the four subtests which were presented together. Here, no group difference was found (F=8.801, df=1,14 ns) nor was a significant interaction effect (F=0.223, df=3,42 ns) but a significant difference was observed between the subtest factor (F=3.901, df=3,42 p<0.05). Pairwise comparisons of the means were made using Tukey's HSD tests (q (4, 42) = 73.88 at the 0.05 level of significance) and it was found that the subtests significantly differed as follows:
Pictures and Letters > Colours.

Due to the fact that the nonword subtest was presented alone, the resulting data were analysed in an unpaired means comparison. An F-test showed that the variance of the two groups was significantly different (F=0.194, num.df=7, den.df=7, p<0.05). So, a non-parametric Robust Rank-Order Test was implemented and showed a non-significant difference (U`=1.27, m,n=8 ns).

**Nonword RAN**

Each subject underwent a Rapid Automatised Naming test which resulted in a measure of the overall time taken for each subject in milliseconds. The dyslexic group's mean was 34281.25ms (sd=9860.70); and the control group's mean was 25536.25 (sd=6379.52). This difference was shown to be significant using an unpaired t-test (t=2.065, df=14, p<0.05).

A Spearman's Rank correlation was used to examine whether there was a relationship between the two naming test measures. So, each subject's discrete-trial nonword median was correlated with each subject's nonword RAN latency. A significant correlation was observed (Rho=0.545, n=18, $P<0.05$). This suggests that the two measures are related in that if a subjects produces a long response on one measure he/she will also take longer on the other (and vice versa).

## Identity Matching

The two groups were measured on their time taken to match two physically identical nonwords. The median response latency was calculated for each subject over a total of eighteen trials. The mean for the Dyslexic group was 1152.87ms (sd=380.15ms); and the mean for the Control group was 790.81ms (sd=124.61ms).

An F-test demonstrated that there was a significant difference between the variance of the two groups (F=9.307, num. df=7, den.df=7, $p<0.01$). As a result the data were analysed using the non-parametric Robust Rank-Order Test which showed a significant difference between the two groups (U`=3.14, m,n=8 sign. at the 0.05 level).

**Training**

**Table 14: The mean number of errors per group made during: Phase 1, train AB and BC; Phase 2, AB and BC without reinforcement; and the mean number of repeat training sessions required by each group. Standard deviations are given in brackets.**

| | Phase 1 Number of Errors to Criterion | Phase 2 Unreinforced Errors | Repeat Number of Sessions |
|---|---|---|---|
| **Dyslexic Group** | 89.38 (68.85) | 2.5 (2.07) | 9.13 (5.41) |
| **Control Group** | 45.63 (62.28) | 1.5 (2.27) | 4.63 (4.53) |

Table 14 shows that the dyslexic group made approximately double the amount of errors to reach criterion. However, an unpaired t-test showed that this difference was non-significant (t=1.33, df=14, ns).

The dyslexic group made on average more errors on the unreinforced sessions but this difference was again non-significant according to an unpaired t-test (t=0.921, df=14, ns).

Finally, the dyslexic group required nearly double the amount of extra training sessions but this difference was again non-significant according to an unpaired t-test (t=1.803, df=14, ns).

Therefore, despite the fact that from looking at the means in

Table 14 it seems that the dyslexic group made double the amount of errors and required double the amount of training sessions but no significant group differences were found. Looking at the raw data the dyslexic means in both cases were inflated by 3 subjects and the control mean by 1 subject. Specifically, these 4 subjects made over 100 errors each which is many more than any of the other subjects. This made the standard deviations equal but in reality there is a trend for dyslexic subjects to make noticeably more errors and require more repeat sessions.

**Test Session**

The test trials were categorised as: baseline, symmetry, and transitivity; as in the previous experiments.

Errors made during the test session

The same pass/fail criterion was adopted as in Experiments 2 and 3.

**Table 15: Pass/fail rate for each group**

| | Pass | Fail |
|---|---|---|
| **Dyslexic Group** | 0 | 8 |
| **Control Group** | 1 | 7 |

Table 15 illustrates the number of subjects from each group who passed and failed according to the above criterion. No subjects from the dyslexic group passed and only one control subject (Control 29) managed to reach the experimental criterion.

The distribution of the errors made by the subjects who failed (n=8 dyslexic and 7 control subjects) is illustrated in Figure 21. Here the test session is divided into two halves to see whether the subjects' performances improve over time. It can be seen that this is clearly the case for: Dyslexic 30 who reaches criterion on the second half; and Control 33 who approaches criterion on the second half. Such an improvement can be observed but less markedly for: Dyslexic 29; and Dyslexic 34. Most subjects' performances remain more or less stable over time, for instance: Dyslexic 27; Dyslexic 31; Dyslexic 33; Control 27; Control 32; and Control 34. The remaining subjects' performances got slightly worse on the second half of the session. For example, Dyslexic 28; Dyslexic 32; Control 28; Control 30; and Control 31. So, overall for the majority of subjects little or no learning took place over the session. In other words, the majority of the subjects irrespective of group persistently failed.

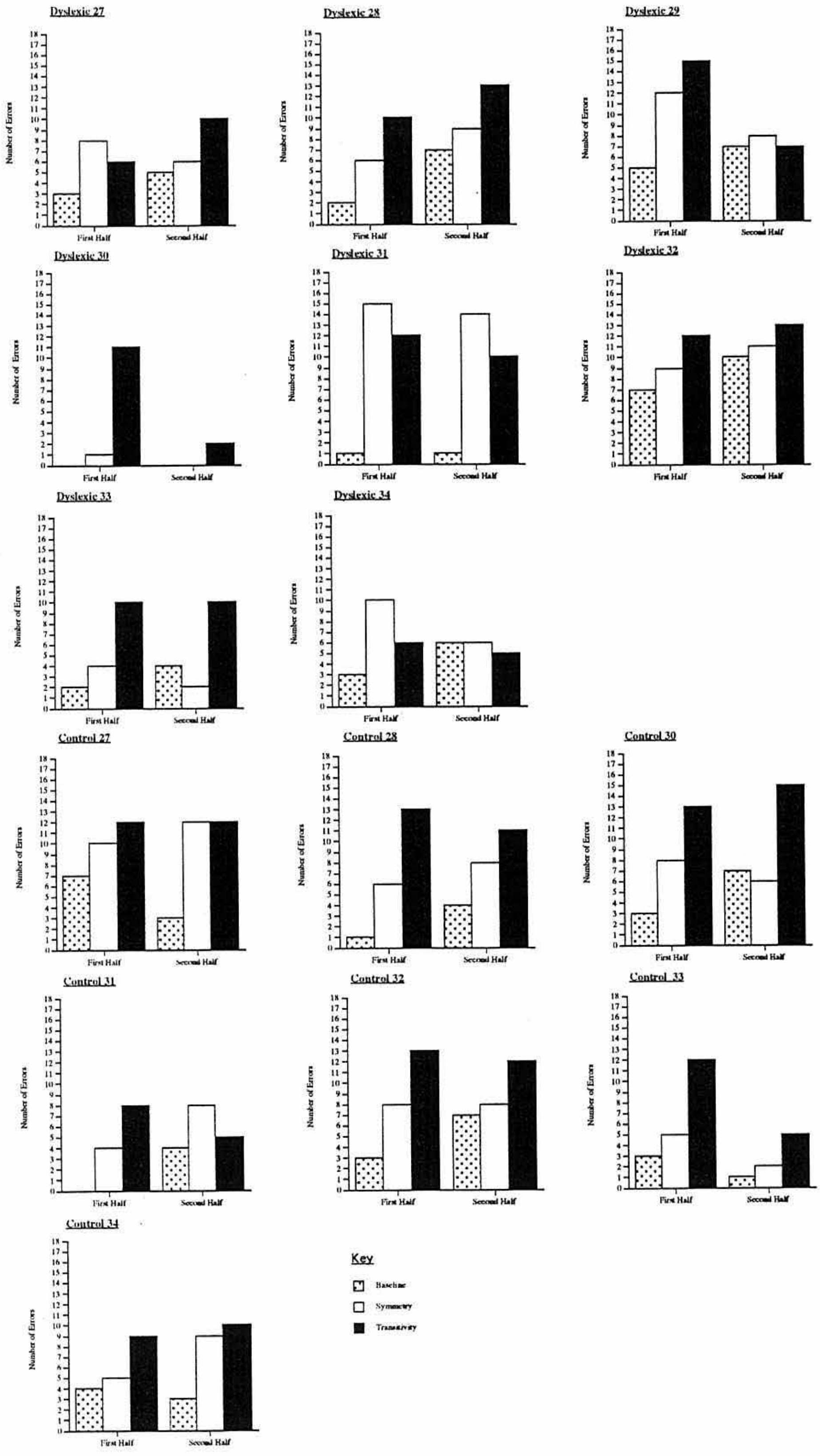
All subjects with the exception of Dyslexic 31; and Dyslexic 34 made the most errors on the transitivity trials.

Seven subjects (4 dyslexic and 3 control) showed evidence of baseline deterioration (that is made more than eight errors on the baseline category).

<u>Test trial response latencies</u>

As in the previous studies a median was calculated for each subject on each of the test session categories. However, in this experiment, such a median was calculated for the subjects who failed. In the three previous experiments

**Figure 21:** **Graphs illustrating the number of errors made by the subjects who failed, on each category, on each half of the test session in Experiment 4.**


Dyslexic 27
Dyslexic 28
Dyslexic 29
Dyslexic 30
Dyslexic 31
Dyslexic 32
Dyslexic 33
Dyslexic 34
Control 27
Control 28
Control 30
Control 31
Control 32
Control 33
Control 34
Number of Errors
First Half
Second Half
Key
Baseline
Symmetry
Transitivity

analyses were performed on the latencies of the subjects who showed evidence of equivalence, with the proviso that any latency differences observed may reflect underlying naming strategies which must be occurring (according to the Naming Hypothesis) in order for the subjects to pass.

Here the majority of the subjects have failed to demonstrate equivalence, therefore, the latencies cannot relate much concerning naming. However, it is of value to make the comparison between the pattern of responding in subjects who pass and those who fail. Given that they are responding differently (that is passing versus failing) are there any noticeable differences? Or were the patterns of responding predicted by Bentall et al (1993) also observed in the subjects who failed as in Experiments 1 and 3? This could be the case if subjects are naming but employing an incongruent strategy. To this end the same analyses, previously employed on the data of subjects who passed were applied to these subjects who failed.

**Table 16: The mean response latencies for each group on the three test session categories (standard deviations are given in brackets).**

| | Baseline | Symmetry | Transitivity | Total |
|---|---|---|---|---|
| **Dyslexic Group** | 3165.50 (786.75) | 5187.94 (1970.07) | 6242.00 (2748.09) | 4865.15 (2317.00) |
| **Control Group** | 3021.57 (876.63) | 6444.79 (1258.32) | 7413.79 (1412.84) | 5626.71 (2242.11) |
| **Total Mean** | 3098.33 (798.52) | 5774.47 (1743.68) | 6788.83 (2235.54) | |

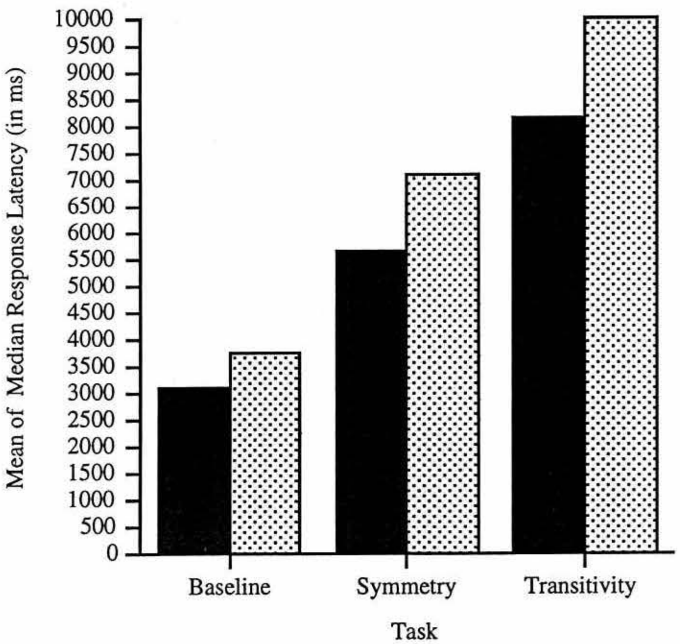
Looking at Table 16 it can be seen that the overall dyslexic group mean latency is shorter than that of the control group and this reflects the fact that the dyslexic group produces, on average, shorter latencies on the symmetry and transitivity trials. On the baseline tasks, the dyslexic group is slower. Both groups produced their shortest latencies on the baseline trials, followed by the symmetry trials, and then the transitivity trials.

A mixed design 2X3 repeated measures ANOVA demonstrated that there was no significant difference between the groups ($F=1.225$, $df=1,13$ ns), or a significant interaction effect ($F=1.445$, $df=2,26$ ns) however, a significant difference was observed between the test categories ($F=34.223$, $df=2,26$ $p<0.0001$). To determine where this difference lay Tukey's HSD tests were performed ($q\ (3, 26) = 1146.22$ at the 0.05 level of significance) and the following pattern was identified: Baseline < Symmetry and Transitivity.

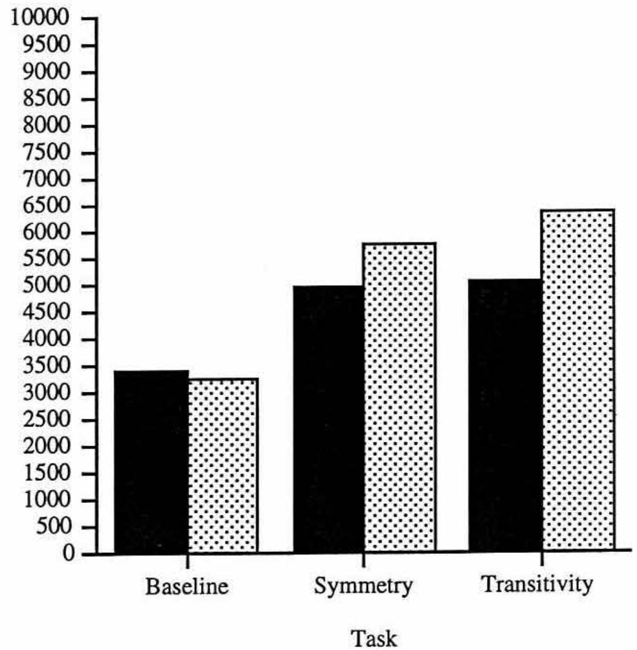
Graphs were plotted (see Figure 22) to determine whether the groups' latencies differed over the two halves of the session. Using visual inspection it can be seen that the difference between the groups remained consistent over both halves of the session. On all tasks (with the exception of the second half baseline task) the control group produced longer latency than the dyslexic group. Both groups' latencies dropped over time.

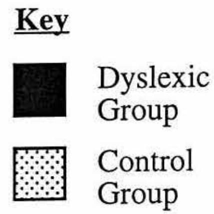
**Figure 22: Graphs comparing each group's mean of median latencies on the first and second half of the test session in Experiment 4.**

**Group Mean Response Latencies**
**First Half of the Test Session**




**Group Mean Response Latencies**
**Second Half of the Test Session**



**Discussion**

The first issue to be addressed was whether the groups significantly differed in naming speed. Using the discrete-trial format no significant group differences were found using the familiar stimuli (pictures, colours, numbers and letters) which replicates the findings of Experiment 2. The largest difference between the groups was seen on the nonword subtest but this was again non-significant. On this task the dyslexic group took 162.47ms longer that the control group. This may have been because the task required the subjects to read novel stimuli, a task which the dyslexic subjects would find difficult and one for which they could not have compensated.

A significant difference was, however, found between the two groups on the nonword RAN test with the dyslexic group taking 8745ms longer. So, at last a significant difference was found between the groups on a measure of naming. This was found using more complex stimuli than the ones employed in the naming tests in Experiments 2 and 3 (that is, words) and in a procedure which placed the subjects responding under the pressure of time.

Given that a significant difference was found using one methodology and not the other raises the question of whether this is because the two tasks measure different latent variables. In order to establish whether there was a relationship between the tests, a Spearman's Rank correlation was employed and a significant correlation was demonstrated. Therefore, a large latency on one test indicates a large latency on the other and vice versa. The correlation provides evidence that the two tasks are measuring the same thing. It seems that whether the difference is big enough to be detected depends on the measure employed. There is an indication that the two tests are both measuring a deficit in naming, but that the RAN test exaggerates any group differences because it is administered under the pressure of time.

Therefore, a significant naming difference was observed on a test which it has been argued represents (to a certain extent) the type of naming which should, according to the naming hypothesis, occur on the matching-to-sample trials if subjects are to successfully demonstrate equivalence.

A significant difference was found between the groups on the identity matching task using the nonword stimuli with the dyslexic group producing significantly longer reaction times. Nonwords were utilised as stimuli (in contrast to the abstract pictorial stimuli previously used in Experiments 2 and 3) in an attempt to attain a measure of nonverbal reaction time using stimuli similar to the ones to be included in the Experimental task. As a result the two measures could be compared; one verbal and one nonverbal. It was hoped that even though the stimuli were 'wordlike', because the task required the subjects to match the two identical stimuli, such matching would occur on a purely physical basis.

It would be especially pertinent if it could be said that this significant difference occurred because of the nature of the stimuli. In other words, the subject could not help but name the nonwords (even though they did not need to). This would result in another verbal measure on which the dyslexic group produced significantly slower latencies. However, this can not be said definitively as a significant difference was observed between the dyslexic and non-dyslexic group in Experiment 3 using abstract, pictorial stimuli. This suggests that the differences may not necessarily be stimuli related, but due to one group containing slower responders.

With regard to the group's performances in learning and maintaining the baseline relations, no significant group differences were observed. This was despite the fact that the dyslexic group had double the amount of training sessions. No significant differences were observed between the variances of the two groups but looking at the raw data the means of the dyslexic subjects on these two measures were inflated by three subjects and the control mean by just one subject. So, there is a slight trend for the dyslexic subjects to make more errors and require more exposure to the relations before their responding reaches criterion. As it can be said with a fair amount of confidence that the subjects were naming the stimuli this detrimental performance may reflect the dyslexic subjects' naming weakness.

Looking at the test session pass/fail rate (Table 15), only one subject from either group passed (Control 29). This is in marked contrast to the previous three experiments where the majority of subjects showed evidence of stimulus equivalence. Why should this be the case? The only difference between the methodologies was the stimuli used, so why should using nonwords as stimuli create so many problems?

In order to shed some light on this issue the subjects' post-experimental remarks were examined. The majority of subjects failed and even when they were taken through the training and test session trials they were not aware of what was being required of them. Most said that their memories had let them down and they could not remember all the relations.

Two potential strategies did emerge which arose from the unique, 'wordlike' qualities of the stimuli:
(1) The subject could treat each nonword as a syllable and therefore, compound the two components of the relation to make one, 'word'. For example, Control 34 stated this was exactly what she did so in effect she only had six words to remember instead of twelve. Similarly, Dyslexic 33 said that she immediately put A2-B2, MOXKEB, together as she is a former nurse and this word sounded like a drug. Dyslexic 27 and Dyslexic 28 both stated that B1-C1, WEFYIM, sounded like a word. Dyslexic 28 found it easy to remember in this form because it sounded like a Chinese surname.

If such a strategy is adopted this is incompatible with equivalence class formation because the relations are being processed as a unit which does not lend itself to being manipulated in the way required in tests of symmetry and transitivity. In other words, the stimuli are not being treated equivalently in the sense that they are interchangeable, if they are formed into a rigid unit.

(2) The second and most reported strategy involved the fact that the stimuli can be broken down into their component parts (ie letters). Many of the subjects reported memory problems, so they were searching for shortcuts. The most readily available one was to remember just one of the letters. It was impossible to use the middle vowel because it was repeated too often. So, some subjects used the initial or final consonant. For instance, Dyslexic 27, Dyslexic 28, Dyslexic 31, Control 27, and Control 28 reported doing this. Subjects reported that on A3-B3, YOF-ZEG, the Y and Z followed on from each other in the alphabet. Control 30 stated that he homed in on the last letters, meaning that:

| | | |
|---|---|---|
| B1-C1 | WEF - YIM | became FM (as in frequency modulated radio) |
| B2-C2 | KEB - ZIQ | became B & Q (as in a well-known DIY store) |
| B3-C3 | ZEG - MIB | became GB (as in Great Britain) |

Dyslexic 29 said that on the test session he matched together words with the same initial letters (for example, MOX-MIB, YOF-YIM).

Using just the initial letter to match the stimuli works well in the training phases. For example:

| | |
|---|---|
| A1-B1 = KW | B1-C1 = WY |
| A2-B2 = MK | B2-C2 = KZ |
| A3-B3 = YZ | B3-C3 = ZM |

However, when a subject using this strategy reaches the test session and is confronted with symmetry and transitivity trials, their learning works against them.

For instance, on the symmetry trial C3-B3 (MIB-ZEG), the

sample C3 (MIB) is presented followed by the comparisons, B1 (WEF), B2 (KEB), and B3 (ZEG). The subject will look at MIB and say M and from the relations learned in training say MK (A2-B2) and as a result respond MIB-KEB (C3-B2) which is incorrect. This could happen for a number of trials.
Put another way, attending to only one letter of the stimulus means that important information for discriminating between the stimuli is lost. This information may not be relevant during training but becomes so during the test session.
On consulting the raw data and especially the test session errors of subjects such as Dyslexic 27, Dyslexic 31 and Control 27 it can be seen that they are responding erroneously due to initial letter strategies acquired during the training phases.
Interestingly, the one subject who passed (Control 29), also reported that he had difficulty remembering the nonwords. However, he said that he manipulated them to give them meaning. For example, WEF became FEW and YIM became MY. So not only was he attending to the whole of the word but he was giving the stimuli meaning which could have facilitated his performance. In addition he verbalised a strategy whereby he said, "A goes with B" and so on, which as seen in the previous three studies is highly correlational with the demonstration of the properties of equivalence.

Despite the erroneous strategies employed here, once again this is an example of subjects who can be assumed to be producing the names of the stimuli intraverbally failing to demonstrate stimulus equivalence. Horne and Lowe's naming hypothesis stipulates that if such naming reliably occurs stimulus equivalence must arise but this is not the case, and was not the case for some of the subjects in Experiment 3. In this instance naming was detrimental to the subjects' performance, something not predicted by the naming hypothesis. Even though it can be said with some certainty that most subjects did employ verbal means here, their strategies could have been incongruent with the intended classes. Thus naming can facilitate or adversely affect equivalence class formation. It can not be said though, that naming is sufficient for the successful demonstration of equivalence.

Looking at Figure 21, the majority of subjects persistently failed, only two subjects' (1 dyslexic and 1 control) performances improved over the session, indicating that some learning did take place, but this was equally apparent in both groups. Seven subjects (4 dyslexics and 3 controls) showed evidence of baseline deterioration so even the trained relations were lost when the other test combinations were presented. This was the most deterioration observed on any of the studies but again it was not dependent on which group the subject was in. Rather it seemed to be a result of the nature of the stimuli used which made them difficult to maintain, possibly because they have no meaning.

Analyses were performed on the data of the subjects who failed to establish whether they performed any differently to the subjects who passed in previous experiments. No significant group differences were observed between the groups. Notably the dyslexic group produced shorter latencies on both the symmetry and transitivity trials. This suggests that the significant differences observed on the RAN test, where the dyslexic group were slower, were not reflected here on the test session latencies where it can be said with some certainty that the subjects were naming.

So, there is some discrepancy here. One possibility is that the dyslexic subjects are employing an alternative strategy which does not rely as heavily on verbal processing. Examining the subjects' post-experimental reports there is no evidence to suggest that there is any difference in the strategies employed by either group. The alternative is that the RAN test does not measure the same naming process which is exhibited on the matching-to-sample trials. Although both involve continuous naming (that is the repetition of the intraverbal strings, according to Horne and Lowe) the RAN is implemented under the pressure of time whereas the matching-to-sample trials are not.

The significant difference observed between the groups on the Identity Matching task was not reflected in the test trial latencies. Therefore, this difference did not appear to influence their performance on the equivalence task.

Finally, the data were considered as a whole in order to establish whether the same patterns of responding were observed in subjects who passed and subjects who failed. The same pattern of responding (baseline<symmetry/transitivity) was seen in this latency data as was observed in Experiment 3 with the subjects who successfully demonstrated equivalence.

So, what does this pattern indicate? It is a different pattern from the two outlined by Bentall et al (1993) and can not be accounted for by Field's, 'Associative Distance Effect'. As it can be said that subjects named in Experiments 3 and 4 this adds to the tentative conclusion that this pattern reflects a verbal strategy. Even though all but one of the subjects failed there was still a distinction between the test categories. This was despite the fact that the subjects claimed to have no knowledge of the task requirements.

The symmetry tasks took significantly longer than the baseline tasks.This, as in Experiment 3, suggested that this may be because the baseline relations are not obviously reversible because they evoke name-name relations, which in everyday experience tend to be unidirectional. The important point is that these patterns occur even when the subjects do not report being aware of having applied any strategy, so these differences in response latencies seem to have nothing to do with whether the subjects passed or failed.

In conclusion, dyslexic and non-dyslexic subjects were compared on a matching-to-sample task which employed nonwords as stimuli. This was in an attempt to accentuate any group differences whilst at the same time ensuring that the subjects employed the naming strategies, which according to the naming hypothesis, bring about equivalence. Only one subject successfully demonstrated equivalence, the remainder failed despite having all assigned names to the stimuli. This appears to challenge one of the major assumptions of the naming hypothesis.

However, this conclusion is disputable because many of the subjects employed erroneous strategies during training which were incongruent with the experimenter defined classes and worked against the emergence of equivalence. In order, to conclude anything definite about the role of naming in stimulus equivalence steps must be taken to eliminate such strategies.

A significant group difference was observed on the nonword RAN test which is the first naming difference demonstrated between these subjects so far. The questions it raises are; (i) was this a reliable finding?; (ii) does this difference apply solely to this type of stimulus or is the effect seen using other stimuli?

## Experiment 5

### Background

Two major issues arose from Experiment 4:

(1) Concerning the naming test methodologies,

(2) Concerning the nature of the nonword stimuli used.

(1) A significant group difference was observed using the RAN test methodology but not on the discrete-trial methodology. It was decided to pursue this issue further, to determine whether the difference is replicable. The same discrete-trial naming tests were administered but this time corresponding RAN tests were also presented (a Picture RAN; a Colour RAN; a Number RAN; a Letter RAN; and a Nonword RAN) using identical stimuli. This was to determine whether the same effect would be seen using familiar stimuli or whether it was merely limited to the nonwords.

(2) After examining the pattern of responding in the subjects who failed and transcribing their post-experimental interviews, it was seen that some of them had formed initial letter strategies which led them to base their test session matching on this and not on the intended equivalence classes. As a result, it was necessary to repeat the experiment but this time after an attempt was made to make such an erroneous strategy impossible.

This was firstly achieved by rearranging the stimuli in a manner that if a subject only uses the initial letters to remember the stimuli, then they should not reach the training criterion. The stimuli were as follows:

| | | |
|---|---|---|
| **A1** | **A2** | **A3** |
| **WEF** | **KEB** | **ZEG** |
| **B1** | **B2** | **B3** |
| **KOJ** | **MOX** | **YOF** |
| **C1** | **C2** | **C3** |
| **YIM** | **ZIQ** | **MIB** |

Therefore, AB training puts together KM (A2-B2, KEB-MOX). When the subject moves on to BC training, the subject is confronted with another K (B1 KOJ), so using the initial letter strategy the subject would put KOJ with MIB (C3) which is incorrect.

A second measure was also implemented this time to encourage the subject to attend to the whole word, from the beginning of training, and not to be tempted to break down the word. This was achieved by introducing a paired-associate learning phase. So, instead of the subject being presented with a sample stimulus followed by three comparison stimuli, the subject just sees the sample and is required to say its name aloud, then he/she must say the name of the corresponding comparison aloud before the single comparison is revealed. Therefore, straight away the subject has to read the whole word aloud and also produce the whole name of the comparison stimulus. Hopefully, this would ensure that the subject matches two whole words. All the AB and BC relations were trained this way, only then was the standard matching-to-sample format reverted to.

Importantly, the subjects were explicitly taught intraverbals which, according to the naming hypothesis, should lead to the successful demonstration of equivalence.

## Method

**Subjects** (See Appendix J for full subject details)

Eighteen subjects in total took part in this study (nine dyslexic subjects and nine control subjects). Each group was made up of eight females and one male subject. Ages in the dyslexic group ranged from 18-53 years (mean age = 29.33 years); and from 19-34 years (mean age = 23.78 years) for the control group.

All eighteen subjects were students at the University of Wales, Bangor. The nine control subjects were recruited via the School of Psychology's student subject pool, where participants signed up for research projects in return for course credits. The nine dyslexic subjects were recruited from the Dyslexic student subject pool which is a database of dyslexic students from all faculties who have volunteered to make themselves available (where possible) for experimental studies. None of the subjects were paid for their time and none were familiar with stimulus equivalence procedures.

Subject Selection

Identical pretests to those administered in Experiments 3 and 4 were presented.

On the Bangor Dyslexia test the dyslexic subjects scored between 1.5 and 8 dyslexia positives (mean 5.39); the control groups' scores ranged from 0 to 1.5 positives (mean 0.67). The dyslexic group's mean is unrepresentative in that only one subject (Dyslexic 35) scored 1.5 positives. If this score is excluded the range falls within the experimental criterion, it becomes 4 to 8 positives. Dyslexic 35's low score is coupled with a score of below 20 on the spelling test (13/30 correct); this added to the fact that all the dyslexic participants over all the studies had been previously diagnosed as being dyslexic at some previous date, justifies her inclusion in the dyslexic sample.

On the last thirty items of the Schonell spelling test the dyslexic group's mean was 10.44 correct; and the control group's mean was 23.67 correct. One dyslexic subject (Dyslexic 36) scored over twenty correct, but because of her previous diagnosis of dyslexia this can be viewed with some confidence as compensation. One control subject (Control 40) scored under twenty correct. This score was coupled with a score of just one positive on the Bangor Dyslexia Test, so it could be due to bad spelling. However, this subject (Control 40) was also an exceptional case as her first language was Danish.

On the Advanced Progressive Matrices task the dyslexic group produced a mean score of 20.22 out of 36 correct; and the control group a mean score of 21.89. Two dyslexic subjects (Dyslexic 39 and Dyslexic 41) scored below the standardised norm for university students (mean 21, standard deviation 4 - Raven 1962). However, some leeway was allowed in these cases as the norms given relate to subjects of 25 years and under and the above subjects were 39 and 32 years old respectively and produced scores of only 1 and 2 points outside the distribution.

In summary, all subjects complied with the experimental dyslexic/non-dyslexic criterion and all showed average and above average ability on the Matrices task.

All participants: signed a consent form; were reminded of their rights as subject; and were asked to fill out a short questionnaire concerning relevant subject details.

In addition to the above mentioned subjects, three dyslexic subjects also took part but were excluded from the final analysis. Two subjects were excluded because their scores on the Raven's Matrices test were too low and one because he failed to reach the training criteria in the experimental task. In order to balance the group sizes three control subjects were omitted at random from the sample.

## Apparatus and Materials

The experimental setup was identical to that illustrated in Figure 12.

The discrete trial naming tests were identical to those described in Experiment 4.

In this study five Rapid Automatised Naming tests were administered, using the same stimuli as were employed in the discrete-trial naming tests. All stimuli were reproduced to fit within a 3/4 inch square. The stimuli were as follows:

(1) Pictures - the five pictures employed in the discrete-trial naming test were produced in the standard RAN format. That is, each was presented randomly ten times in the form of a 10X5 grid on an A4 size sheet of white paper. The only stipulation was that no two of the same stimuli could appear side by side.

(2) Colours - the five colours (drawn as circle 3/4 inch in diameter) as above.

(3) Numbers - presented as above.

(4) Letters - presented as above.

(5) Nonwords - the nine nonwords were randomly presented five times in a 9X5 grid.

A microphone, tape recorder and stopwatch were also utilised for recording and timing purposes.

The stimuli and equipment used for the Identity Matching task were identical to those employed in Experiments 2 and 3 (that is they were abstract, pictorial stimuli- see Figure 13).

The experimental task consisted of the same stimuli used in Experiment 4, however, the classes were grouped differently as follows:

| | | |
|---|---|---|
| **A1** | **A2** | **A3** |
| **WEF** | **KEB** | **ZEG** |
| **B1** | **B2** | **B3** |
| **KOJ** | **MOX** | **YOF** |
| **C1** | **C2** | **C3** |
| **YIM** | **ZIQ** | **MIB** |

The equipment, computer applications and format employed were the same as previously utilised in Experiments 2, 3 and 4. A paired-associate learning phase was introduced. This involved the experimental stimuli and was written and presented using the SuperLab 1.5.9 Beta application.

The Post-experimental Interview followed the same structure as the ones in Experiments 3 and 4 with the stimuli altered appropriately.

**Experimental Design**

The experiment was run over two sessions.

Session one comprised of: the subject selection tasks, namely the Bangor Dyslexia Test, the Spelling test, and the Advanced Progressive Matrices; the Discrete-trial Naming tests; and the Rapid Automatised Naming tests. This lasted from one to one and a half hours.

Session two contained: the Identity Matching task; the Experimental task; and the Post-experimental task. This lasted between one and two hours.

All the experimental designs used were identical to those previously employed with the exception of the experimental task. A paired-associate learning phase was introduced to teach the initial AB and BC relations.

Following this the matching-to-sample format was reverted to for the AB/BC with reinforcement phase; the AB/BC without reinforcement phases; and the test session. These phases were identical to those presented in Experiments 2, 3 and 4.

**Procedure**

Discrete-trial Naming Tests

The procedure followed was identical to that described in Experiment 4. The subtests: pictures, colours, numbers and letters were administered together; followed by the nonword subtest.

Rapid Automatised Naming Tests (RANS)

Five of these were administered in the following order: pictures, colours, numbers, letters; and nonwords. The appropriate sheet was placed in front of the subject with the Experimenter holding the microphone in a constant position. The following instructions were given (as appropriate):

> You must say out loud as quickly as possible the names of the following pictures/colours/numbers/letters/nonwords. Please go from left to right and down the page as if they were lines in a book (Experimenter shows the subject). I will say go after which you must start. Are you ready? Go!

The whole session was tape recorded.

Identity Matching Task

The procedure here was identical to that employed in Experiments 2, 3 and 4. As the stimuli used were the same as those presented in Experiments 2 and 3 the corresponding instructions used there were repeated.

Experimental Task

(1) Paired-Associate Learning Phase

The initial training was in the form of a paired-associate learning task. The procedure was as follows (see also Figure 23):

(i) The sample appears on the screen.

(ii) The subject names the sample and attempts to name the corresponding comparison.

(iii) The Experimenter presses the space bar to reveal the correct comparison.

(iv) The Experimenter presses the space bar and the next trial

**Figure 23:** **An illustration of a typical trial from the Paired-associate learning training session in Experiment 5.**

| | |
|---|---|
| WEF | |

**The subject is presented with the sample stimuli (above) which must be named. The subject must then say aloud the name of the corresponding comparison stimulus before it is revealed (below).**

| | |
|---|---|
| WEF | KOJ |

begins.

The Experimenter controls the sequence to ensure that the subject vocalises the sample and comparison names correctly. Any feedback is given by the experimenter.

The trials utilising the above design were as follows:

| | | | |
|---|---|---|---|
| (i) Train | A1-B1 | 18 | trials |
| | A2-B2 | 18 | trials |
| | A3-B3 | 18 | trials |
| | Train AB | 36 | trials |
| (ii) Train | B1-C1 | 18 | trials |
| | B2-C2 | 18 | trials |
| | B3-C3 | 18 | trials |
| | Train BC | 36 | trials |

On the above sessions, if the subject made more than two errors per trial type then the session was repeated.

(iii) Train AB/BC 36 trials

The six AB and BC trials are presented together (in a psuedorandom order meaning that no three of the same trial type can appear consecutively). The subject must not make more than one error per trial type or the session was repeated. On the above sessions the following instructions were read out loud to each subject:

> In this task a nonsense word will appear on the screen, you must read this word aloud, and then guess which nonsense word will appear next. You must try and guess as many correctly as you can. At first you will not know which nonsense word is going to appear next, but once you have made your first guess the correct nonsense word will appear alongside on the screen. Do you understand?
> I will talk you through the first couple of examples.

(2) Matching-to-Sample Phase

(iv) Train AB/BC with reinforcement

Once the criterion had been met on the above session the format changed to that of the standard matching-to-sample format utilised in Experiments 2, 3 and 4. In other words, a sample and three comparisons were now presented to the subject and she/he was required to select the correct comparison using the keypad.

(v) Train AB/BC without reinforcement

Once the criterion on the above had been reached the reinforcement was dropped.

(vi) Test Session

This was identical to that presented in Experiment 4.

Post-Experimental Interview

This followed the same format as in Experiment 2, 3 and 4. The experimenter talked the subject through examples of the session participated in and asked the subjects to relate their thoughts, strategies used and so on.

This session was video-taped.

## Results

### Discrete-trial Naming Test

For each subject a vocalisation latency was measured for every stimulus and a median calculated for each subtest.

Table 17: Group means (presented in milliseconds) for each subtest of the discrete-trial naming tests (standard deviations are given in brackets).

| | Pictures | Colours | Numbers | Letters | Nonwords |
|---|---|---|---|---|---|
| **Dyslexic Group** | 1059.03 (117.55) | 947.73 (70.14) | 1069.19 (157.84) | 1098.56 (161.15) | 1299.74 (380.74) |
| **Control Group** | 1060.84 (86.85) | 1349.07 (1135.10) | 999.52 (112.65) | 978.23 (100.84) | 1033.38 (101.65) |
| **Total Mean** | 1059.94 (100.27) | 1148.40 (807.02) | 1034.36 (137.77) | 1038.39 (144.35) | 1164.06 (301.93) |
| *Difference* | *1.81* | *401.34* | *69.67* | *120.33* | *266.36* |

Table 17 presents the group means for each of the five subtests. The largest difference between the group means can be seen on the colours subtest, followed by the nonword subtest, the letters, the numbers, and lastly a minimal difference is observed on the pictures subtest.

As outlined in Experiment 4 the nonword subtest was analysed separately due to the fact that it was not administered in the same battery as the other subtests. As a result a 2X4 repeated measures ANOVA was used to analyse the four subtests (pictures, colours, numbers and letters). This revealed a non-significant difference (F=0.224, df=1,16 ns). There was also no within subjects effect (F=0.329, df=3,48 ns) nor any significant interaction effect (F=1.633, df=3,48 ns). Finally, the nonword subtest data was analysed using the non-parametric Robust Rank-Order Test (due to a significant difference found between the variance of the two groups using an F-test - F=14.031, num.df=8, den.df=8, p<0.001). Here a significant difference was observed between the groups on this measure (U`=2.42. m,n=9 sign. at the 0.05 level). This was the first time a significant group difference was found using a discrete-trial naming test.

Looking at Table 17 it can be seen that a large difference was observed between the groups on this measure but it was not the largest difference. This occurs on the colours subtest (401.34ms). I f this subtest is treated separately a non-significant result was observed using a Robust Rank-Order Test (U`=0.94, m,n=9 ns)) which accounts for the unequal variance between the two groups (which can be seen clearly by looking at the standard deviations). Returning to the raw data it can be seen that the control group's mean on the colours subtest was unduly inflated by just one subject's median (Control 35, who produced a median of 4371.29ms on this subtest). If this median is excluded from the group, the mean lowers to 971.30ms which is much closer to that of the dyslexic group.

**RAN analyses**

The time taken to perform each subtest was calculated in milliseconds for each subject. The group means are presented in Table 18.

Table 18: The mean performance of each group for each RAN subtest (standard deviations are presented in brackets)

| | Pictures | Colours | Letters | Numbers | Nonwords |
|---|---|---|---|---|---|
| **Dyslexic Group** | 40531.11 (4515.34) | 32536.67 (6335.86) | 22774.44 (3024.67) | 23008.89 (4430.92) | 43312.22 (12642.79) |
| **Control Group** | 35872.22 (5107.06) | 29773.33 (3709.24) | 19298.89 (4152.97) | 20667.78 (3067.79) | 27145.56 (4422.77) |
| **Total Mean** | 38201.67 (5254.89) | 31155.00 (5233.24) | 21036.67 (3952.09) | 21838.33 (3888.28) | 35228.89 (12393.87) |
| *Difference* | *4658.89* | *2763.34* | *3475.55* | *2341.11* | *16166.66* |

As can be seen in Table 18 the largest difference between the two groups is observed on the nonwords subtest. This is followed by the pictures, letters, colours, with the least difference being observed on the numbers subtest.

These data were analysed using a mixed design 2X5 repeated measures ANOVA. Firstly, a significant difference was found between the two groups on this measure ($F=12.341$, $df=1,16$ $p<0.005$). The total dyslexic group mean was 32432.67ms and the total control group mean was 26551.56ms. Therefore, overall the control group was significantly faster than the dyslexic group.

However, where did these differences specifically lie over the five subtests? In order to locate the group differences unpaired t-tests were performed for each subtest. No significant differences were found between the groups on: the pictures subtest ($t=2.05$, $df=16$, ns); on the colours subtest ($t=1.129$, $df=16$, ns); on the letters subtest ($t=2.029$, $df=16$, ns); or on the numbers subtest ($t=1.303$, $df=16$, ns). However, a significant difference was found between the groups on the nonwords subtest. Using a non-parametric Robust Rank-Order Test (due to a significant difference between the variance of the two groups as seen on an (F-test - $F=8.171$, num.df=8, den. df=8, $p<0.01$) $U`=10.35$, m,n=9 sign. at the 0.05 level.

Secondly, a significant within subjects effect was observed ($F=41.598$, $df=4,64$ $p<0.001$). Looking at the total mean for each subtest it can be seen that the pictures subtest took the longest, followed by the nonwords, colours, numbers then letters. These differences were compared using Tukey's HSD tests ($q$ (5, 64) = 4795.93 at the 0.05 significance level) and it was seen that: Pictures > Colours > Letters/Numbers but Pictures = Nonwords = Colours.

However, the picture is not as clear cut as this because a significant interaction effect was also found ($F=5.825$, $df=4,64$ $p<0.005$) suggesting that the order of difference differs for each group. Put another way the effect of the stimuli are different for the non-dyslexic and dyslexic subjects. In order to tease out where these effects differed two one-way repeated measures ANOVAs were performed: (1) for the dyslexic subjects only; and (2) for the control subjects only.

(1) The dyslexic ANOVA ($F=18.277$, $df=4,44$ $p<0.0001$).

The rank order of subtests was: Nonwords longest
Pictures
Colours
Numbers
Letters shortest

According to a Tukey HSD test (q (5, 32) = 9186.79 at the 0.05 significance level): Nonwords > Colours > Numbers/Letters but Pictures = Nonwords = Colours.

(2) The control ANOVA ($F=58.673$, $df=4,44$ $p<0.0001$).

The rank order of subtests was: Pictures longest
Colours
Nonwords
Numbers
Letters shortest

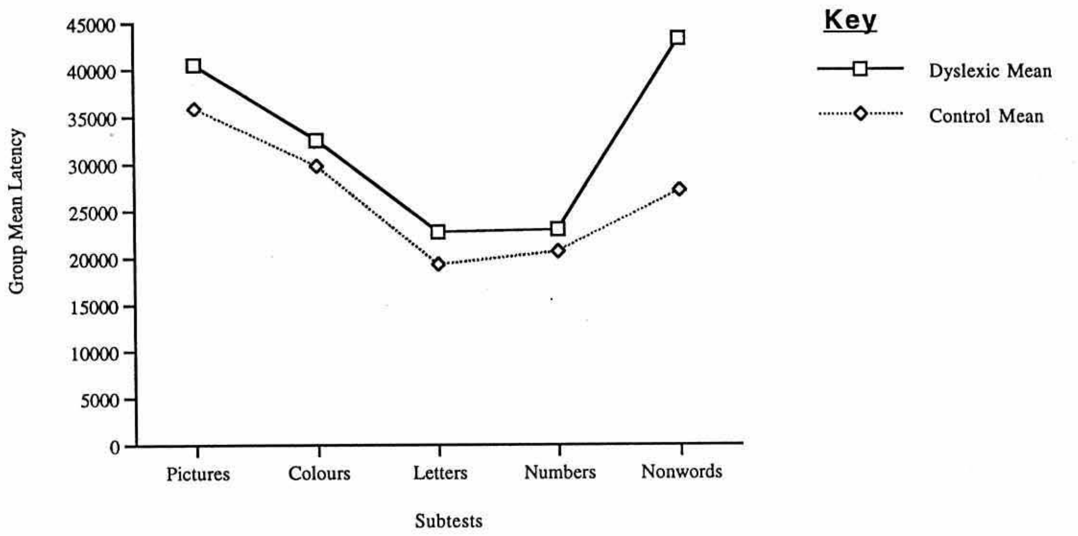
According to a Tukey HSD test (q (5, 32) = 3638.47 at the 0.05 significance level): Pictures > Colours/Nonwords > Letters/Numbers.

So, the difference between the groups is seen with regard to the effect of the nonword subtest. It has a greater effect on the dyslexic group. Figure 24 illustrates this graphically, showing the greatest disparity between the groups occurring on the nonwords subtest. This is supported by the unpaired t-test which demonstrated that this is the subtest where the significant group differences lie.

In an attempt to establish whether there is any relationship between a subject's performance on the discrete-trial naming test and on the RAN naming test, an overall mean was calculated for each subject on each measure. These were compiled from the subject's subtest median on the discrete trial naming test and from the five total time measures on the

**Figure 24: A graph comparing the group's mean latencies on the five RAN subtests.**

## The Groups Mean latencies on the Five RAN Subtests



RAN tests. These data were then compared by calculating a Spearman's rank order correlation coefficient and a significant correlation was found (Rho=0.612, n=18, $p<0.01$).

**Identity Matching**

Each subject's median latency was calculated on this measure. The dyslexic group's mean was 1145.67ms (sd=230.61ms) and the control group's mean was 973.89ms (sd=156.20ms). An unpaired t-test demonstrated that there was a non-significant difference between the groups (t=1.85, df=16, ns).

**Training**

As in the previous studies two measures were taken: (1) the number of errors to criterion and the number made on the unreinforced training trials only; and (2) the number of repeat training sessions each group required to reach criterion. An added factor in this study was that there were two training measures implemented (the paired-associate learning - PAL - followed by the matching-to-sample methodology - MTS). So, an extra question was addressed as to whether the groups performed differently on these two methodologies.

Table 19: A breakdown of the mean number of errors made on the various training sessions (standard deviations given in brackets).

| | Number of Errors to Criterion | PAL Errors | MTS Errors | Unreinforced Errors |
|---|---|---|---|---|
| **Dyslexic Group** | 43.89 (29.28) | 32.44 (20.35) | 9.67 (10.06) | 1.56 (0.88) |
| **Control Group** | 35.33 (21.07) | 29.44 (17.96) | 6.56 (6.52) | 0.67 (0.50) |

Table 19 shows that the dyslexic group made, on average, more errors overall, and on each phase of the training session. However, unpaired t-tests showed that there was no significant difference between the groups on the number of errors to criterion made (t=0.712, df=16, ns); the number of errors made on the PAL phase (t=0.332, df=16, ns); or on the number of errors made on the MTS methodology (t=0.778, df=16, ns). However, a significant difference was found between the groups on the number of errors made on the unreinforced matching-to-sample sessions alone (t=2.630, df=16, $p<0.05$) with the dyslexic group making more errors. This indicates that their performance was adversely affected by the removal of feedback.

The next question was whether there were any group differences on the number of repeat sessions needed during training.

Table 20: A breakdown of the mean number of repeat training session needed on the various training sessions (standard deviations are given in brackets)

| | Total Number of Sessions | PAL Sessions | MTS Sessions |
|---|---|---|---|
| **Dyslexic Group** | 3.33 (2.18) | 2.22 (1.56) | 1.11 (1.05) |
| **Control Group** | 3.33 (2.00) | 2.44 (1.33) | 0.89 (1.05) |

Table 20 shows that there was only a slight difference between the groups on any of the measures. There was no difference whatsoever between the groups looking at the total number of sessions needed. Unpaired t-tests showed that there was no significant difference between the groups on the number of PAL training sessions required (t=-0.324, df=16, ns); or on the number of extra MTS sessions needed (t=0.447, df=16, ns).

## Test Session

As in all the previous studies the test session trials were divided into the following categories: baseline, symmetry, and transitivity.

### Errors made during the test session

The same pass/fail criterion was adopted as in Experiments 2, 3 and 4.

Table 21: Pass/fail rate for each group

| | **Pass** | **Fail** |
|---|---|---|
| **Dyslexic Group** | 1 | 8 |
| **Control Group** | 1 | 8 |

Table 21 shows that only one subject from each group managed to reach the experimental criterion for passing the test session. So, there was no difference between the two groups on this pass/fail measure.

The distribution of the errors made by the sixteen subjects who failed (n=8 dyslexic subjects and 8 control subjects) is

illustrated in Figure 25. For subjects: Control 36, and Dyslexic 36, and Control 43; performances clearly improve in the second half of the session with Control 36 reaching criterion on the second half and Dyslexic 36 and Control 43 approaching it. This is less so for Dyslexic 37; and Dyslexic 38. Most subjects performances remained more or less stable over time. For instance: Control 38; Control 39; Control 41; Control 42; Dyslexic 35; Dyslexic 38; Dyslexic 39; Dyslexic 40; and Dyslexic 41. Two subjects' performances became slightly worse on the second half: Control 35; and Dyslexic 42.
So, the majority of subjects (irrespective of group) persistently failed and no learning took place over the course of the session. All subjects made the most errors on the transitivity trials. Five subjects showed evidence of baseline deterioration (1 control subject; Control 41 and 4 dyslexic subjects; Dyslexics 35, 36, 37 and 41). This suggests that there is more evidence of this occurring in the dyslexic group.

Test Trial Response Latencies

As in Experiment 4, the majority of subjects failed to demonstrate equivalence. Therefore, the following analyses concerning only the subjects who failed have been included in order to see whether the pattern of responding is the same for subjects who pass and those who fail.

Figure 25: Graphs illustrating the number of errors made by the subjects who failed, on each category, on each half of the test session in Experiment 5.

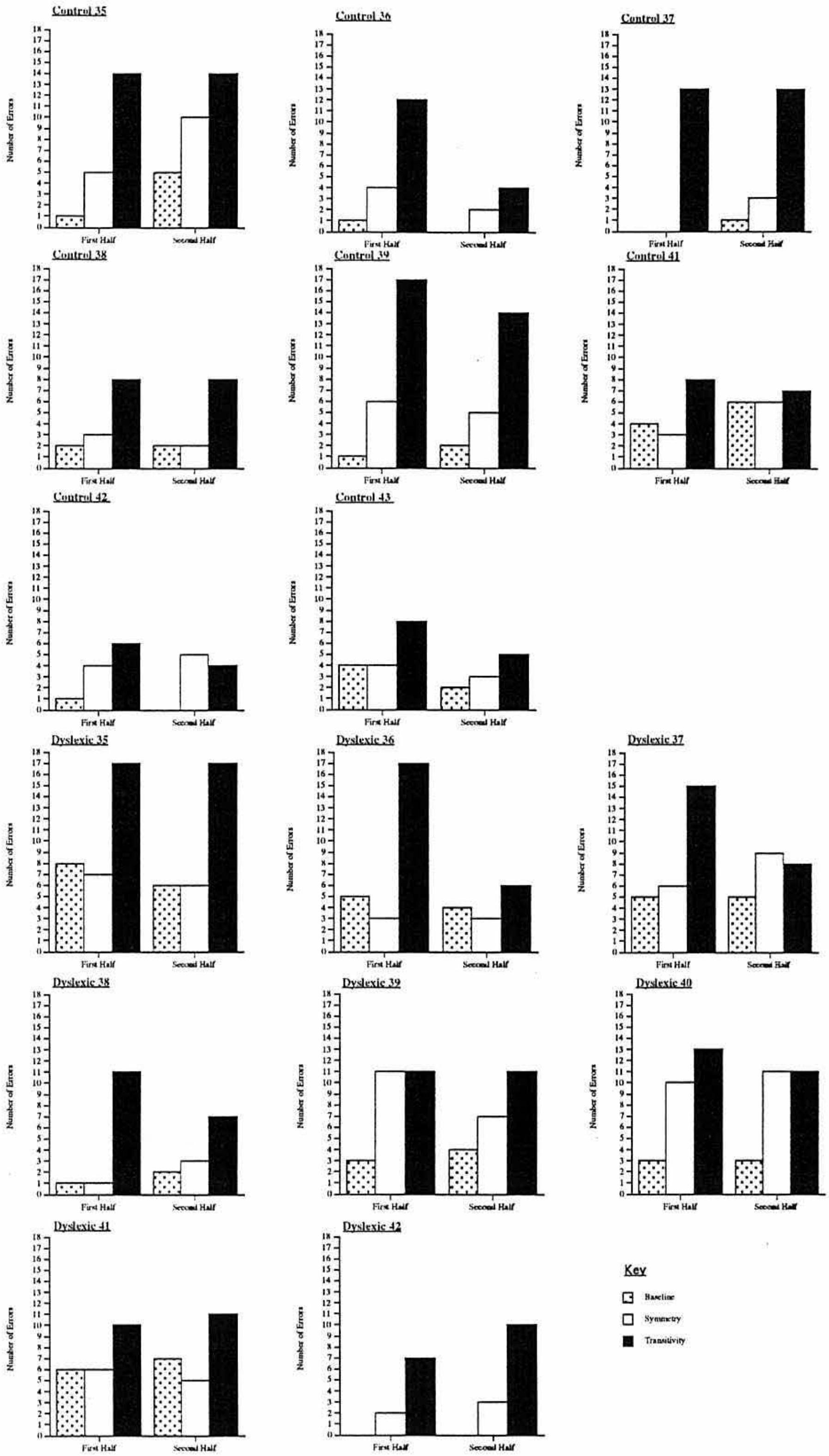

Control 35
Control 36
Control 37
Control 38
Control 39
Control 41
Control 42
Control 43
Dyslexic 35
Dyslexic 36
Dyslexic 37
Dyslexic 38
Dyslexic 39
Dyslexic 40
Dyslexic 41
Dyslexic 42
Number of Errors
First Half
Second Half
Key
Baseline
Symmetry
Transitivity

Table 22: The mean response latencies for each group on the test session categories (standard deviations are given in brackets)

| | Baseline | Symmetry | Transitivity | Total |
|---|---|---|---|---|
| **Dyslexic Group** | 4259.00 (1699.97) | 6564.69 (2908.14) | 8945.25 (3185.18) | 6589.65 (3218.83) |
| **Control Group** | 2810.12 (1116.69) | 5940.56 (1381.00) | 10388.19 (5331.13) | 6379.63 (4438.21) |
| **Total Mean** | 3534.56 (1578.08) | 6252.63 (2222.75) | 9666.72 (4307.30) | |

Looking at Table 22 it can be seen that the overall means for the two groups are very close. The dyslexic group produces longer latencies on the baseline and symmetry trials but has a shorter mean than the control group on the transitivity trials. Both groups produced their longest latencies on the transitivity trials, followed by the symmetry and lastly the baseline trials.

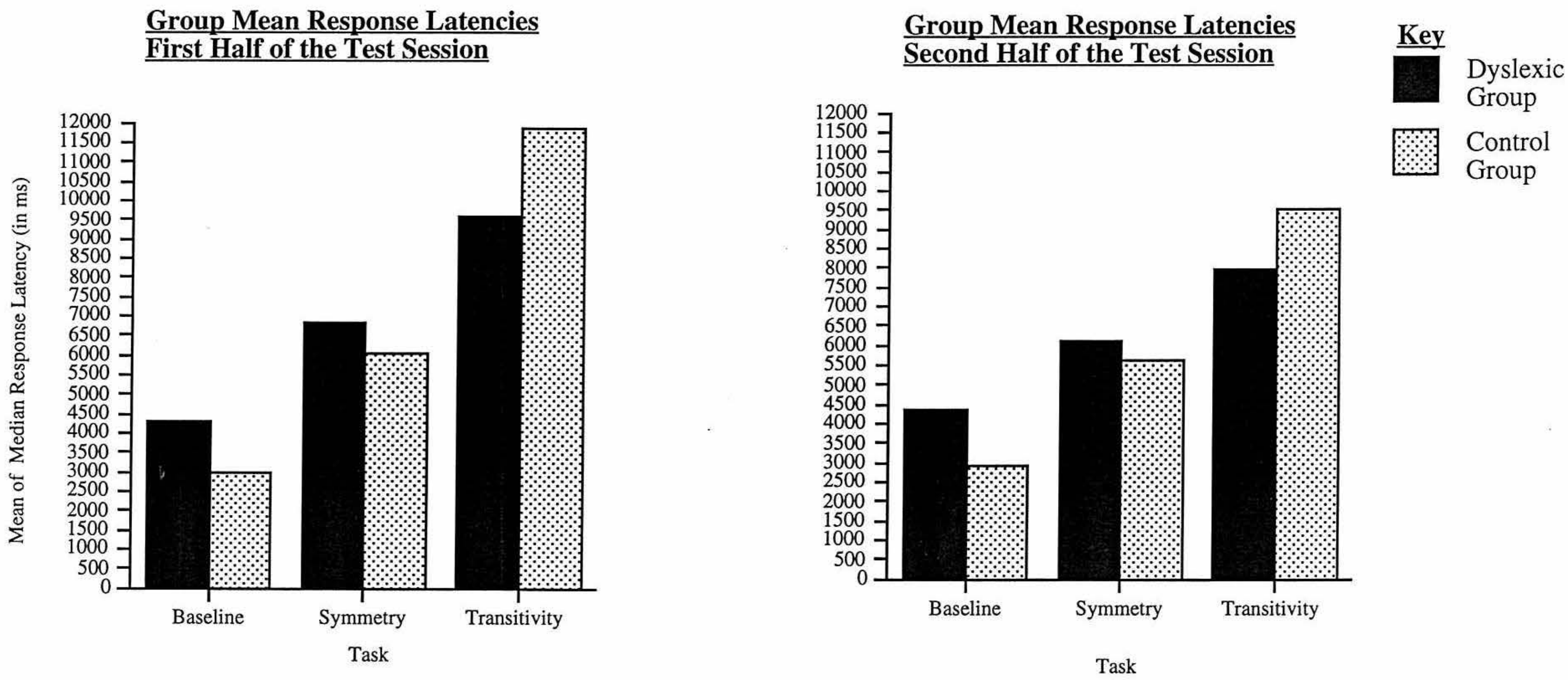
A mixed design 2X3 repeated measures ANOVA showed that there was no significant difference between the two groups ($F=0.039$, $df=1,14$ ns). There was, however, a significant difference between the categories ($F=23.037$, $df=2,28$ $p<0.0001$). To establish where this difference lay a pairwise comparisons of the category means were made using Tukey's HSD test ($q\ (3, 28) = 2243.20$ at the 0.05 level of significance) and it was found that:

Baseline < Symmetry < Transitivity.

No significant interaction effect was seen ($F=1.354$, $df=2,28$ ns).

Figure 26 shows the group's performances over the first and second halves of the test session. The latencies reduced slightly on the second half of the session but the differences between the groups remain equal across the session.

Figure 26: Graphs comparing each group's mean of median latencies on the first and second half of the test session in Experiment 5.


Group Mean Response Latencies
First Half of the Test Session
Mean of Median Response Latency (in ms)
12000
11500
11000
10500
10000
9500
9000
8500
8000
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Baseline
Symmetry
Transitivity
Task
Group Mean Response Latencies
Second Half of the Test Session
12000
11500
11000
10500
10000
9500
9000
8500
8000
7500
7000
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Baseline
Symmetry
Transitivity
Task
Key
Dyslexic Group
Control Group

## Discussion

Using a discrete-trial format the naming speeds of the dyslexic and non-dyslexic groups were compared. No significant differences were found between the groups using familiar stimuli (that is pictures, colours, letters and numbers) as was the case in Experiments 2 and 4. However, there was a significant difference between the groups when nonwords were employed. The dyslexic subjects were 226.36ms slower. In Experiment 4 the largest group difference was demonstrated on this subtest but it was not statistically significant. It appears that using more complex stimuli and, specifically, requiring the subjects to use their phonological processing skills to read the novel nonwords, had a detrimental effect on the dyslexic subjects' performances. This would be expected given that a phonological deficit is a chief characteristic of developmental dyslexic (Chapter 2). This confirms that the dyslexic naming impairment does persist into adulthood, but only for certain stimuli, namely words (this finding is also demonstrated by Watson and Brown, 1992 with adult subjects).

No significant overall difference was found between the subtests on this discrete-trial measure. Yet differences have been seen previously (Experiments 2 and 4) using this procedure. Why should this be the case? The dyslexic and control group were faster (and slower) on different subtests. Their rank orders were as follows:

| | | | |
|---|---|---|---|
| **Dyslexic Group** | Colours *fastest* | **Control Group** | Letters *fastest* |
| | Pictures | | Numbers |
| | Numbers | | Nonwords |
| | Letters | | Pictures |
| | Nonwords *slowest* | | Colours *slowest* |

and were almost the opposite of each other. This results in the total mean for each subtest landing somewhere in the middle for each subtest (see Table 17). Such a pattern would be evident looking at the interaction effect. This ($F=1.63$, $df=3,48$, $p=0.1941$) was not statistically significant but it was the nearest to significance out of the three values.

A significant group differences was also demonstrated on the RAN tests with the dyslexic group taking longer. This difference was also found on the nonword subtest.
So, significant group differences were found on the nonword subtest using two independent measures. To confirm that the two tasks were related a Spearman's Rank order correlation was performed on these data and a significant correlation was demonstrated. In this way performance on one test is related to performance on the other. This suggests that they are both reliable measures of naming.

In the RAN analysis a significant interaction was also observed indicating that the effect of the stimuli was different for each group. It was confirmed that the dyslexic group took longer on the nonwords subtest whereas the control group took longer on the pictures subtest. This explains why the greatest difference between the two groups was seen on the nonwords subtest and confirms that these stimuli accentuate the dyslexic naming weakness.
Here now is strong evidence that the dyslexic subjects take significantly longer in naming the experimental stimuli. The question is would this difference be reflected on the matching-to-sample test session latencies?

No significant difference was found between the groups on the nonverbal, Identity Matching task which implies that the groups did not differ on a simple, matching task based on physical features but do differ when a verbal element is introduced (for example, the naming tests).
The only measure on which the two groups differed during any of the training phases was on the number of errors made during the unreinforced sessions, with the dyslexic group making significantly more. This indicated (as in Experiment 1) that the dyslexic subjects were more adversely affected by the removal of feedback. One reason for this could be that the subjects had not received that much experience of the matching-to-sample trials as the basic relations were taught using the PAL methodology. So, this could have resulted in the dyslexic subjects being more unsure of their performances on these trials. The groups did not, however, differ in the number of sessions required. Therefore, even though the dyslexic subjects made more unreinforced errors, they had no extra exposure to the baseline relations, which is a hint that the quality of learning was not as high as for the control group.

On the test session there was no difference in the pass/fail rate of the two groups. Only one subject from each group managed to reach the criterion. This was despite the procedural amendments which ensured that the subjects attended to the whole of the stimulus. As it can be assumed that subjects are naming then this would suggest that naming has little to do with subjects passing or failing.

So then, why should the overwhelming majority of subjects fail when the stimuli are nonwords? For an insight into this question each subject's post-experimental remarks were examined.

The majority of subjects reported that they had paid attention to the whole of the word and had remembered the relations by saying the words together. However, this did not aid them on the test session. Six subjects (Control 35; Control 36; Control 42; Control 43; Dyslexic 36; and Dyslexic 38) reported, without prompting, that they were aware of what they were required to do and that the stimuli fell into three classes. They claimed their memory/attention span let them down. Other subjects (For instance: Control 37; Control 39; Control 41; and Dyslexic 40) said that they put together words which, 'sounded right together'. Or they put words together by elimination, that is they knew that stimuli which shared the same initial letter did not go together.

Some subjects reported that they did use individual letters as clues. Control 36 and Dyslexic 35 said that they remembered A3-B3, ZEG-YOF, because ZY are at the end of the alphabet. Dyslexic 42 said that she remembered A3-B3, ZEG-YOF, because Y was a letter less than Z and F was a letter less than G.

Other subjects reported compounding the stimuli. For example, Control 36, remembered A2-B2, KEB-MOX, as one word; Dyslexic 36 thought that A3-B3, ZEG-YOF, sounded like a Hebrew verb. Control 35 and Dyslexic 41, both stated that for B2-C2, MOX-ZIQ, they said, 'Mozambique'.

Further strategies included contracting the stimuli. For instance Control 36, said B1-C1, KOJ-YIM became KIM, somebody he knew. Similarly it became JIM for Dyslexic 36, her brother. For Dyslexic 42, A2-B2, KEB-MOX became BOX. Two dyslexic subjects, Dyslexic 43 and Dyslexic 37 reported that they remembered the stimuli visually.

The problems the subjects encountered fall into two categories:
(1) At least half the subjects reported that they knew what they had to do but failed because their memories let them down. This could be because the stimuli are meaningless or hard to discriminate between.
(2) For the other half the actual stimuli hampered the formation of equivalence classes. That is, they were picking words because they sounded right together and not because of any prior learning. Or they were using initial letters or compounding/contracting stimuli, resulting in information being lost or the relations becoming too rigid to manipulate. It has been argued that this is the most likely possibility because the subjects failed. However, from the subjects' reports it can not be determined when these strategies were implemented or what purpose they served. For example, they could have served as mnenomics but the word pairs were actually being named intraverbally, which is what Horne and Lowe's naming hypothesis state should bring about equivalence. As a result, as in Experiment 3, subjects were naming intraverbally yet still failing to demonstrate equivalence.

These are arguments implying that it was the unique characteristics of the stimuli themselves which prevented the demonstration of equivalence. However, in essence the subjects were taught to intraverbally name the stimulus pairs (in the form of the PAL training phases) and yet the majority of the subjects failed the test of equivalence. Once again (as in Experiments 3 and 4) it was demonstrated that to merely name the stimuli intraverbally is not enough to bring about equivalence as Horne and Lowe's naming hypothesis predicts.

The majority of subjects persistently failed over both halves of the session. Only three subjects (2 controls and 1 dyslexic) showed any improvement. So, there is no evidence that any of these subjects would be likely to demonstrate equivalence using this protocol. Five subjects (1 control and 4 dyslexics) showed evidence of baseline deterioration which indicates that there is a definite trend for dyslexic subjects to become disrupted on their baseline performances. This could suggest that their performance was not as concrete in the first place given that they made significantly more errors on the unreinforced training trials but did not receive any extra exposure to the relations.

Despite the significant naming differences no significant group differences were observed for the latencies on the test session trials, in fact, the group's means were very close. It must be remembered that this analysis was performed using subjects who failed, but they were subjects who, despite failing, should have been naming the stimuli (by virtue of the fact that they were 'wordlike') yet the significant naming difference was not observed on the test session trials.

So, there is a discrepancy. The naming hypothesis predicts that if subjects intraverbally name this will automatically lead to equivalence, so it could be argued that these subjects were not naming consistently and therefore, no latency differences were observed. Is it reasonable to assume that subjects when presented with wordlike stimuli would not name? These subjects were failing and so on being confronted with the test session trials, may have not been able to make any sense of it and therefore, were just responding randomly to get it over and done with and were not focusing at all on the stimuli names.

This is possible, but looking at the pattern of responding observed on the test session latencies it was found that: baseline < symmetry < transitivity. So, some distinction is being made between the trial types (the more complex tasks take longer) despite subjects claiming not to know anything about the task requirements.

This pattern is the same as seen in Experiment 4 in the subjects who failed in that once again a distinction is made between the baseline and symmetry tasks. It is also the same pattern as seen by subjects who successfully demonstrated equivalence in Experiment 3 and who were also assigning names to the stimuli. This strengthens the conclusion that this pattern reflects a verbal strategy but that this is a pattern which has little to do with the subject passing or failing.

## General Discussion

### Stimulus Equivalence

These two studies were devised as an attempt to examine whether the dyslexic and non-dyslexic samples differed on a matching-to-sample task using 'wordlike' nonword stimuli. If a naming difference exists between the groups and the stimuli used in the training sessions are nameable, then naming should occur during the tests of equivalence and as a result a group difference should, if naming is necessary for equivalence, be observed on the test-trial latencies.

Experiment 4 sets out to investigate this, but only one subject from either group managed to demonstrate equivalence. Subjects' post-experimental remarks suggested that some of the subjects failed on the test session because they had used only the initial letters of the stimuli to form the basic relations. The repetition of some of the initial letters meant that the same initial letter pairings could be implemented on some of the novel test session trials resulting in a high proportion of errors because key discriminating features of the stimuli had been lost.

It was decided to repeat the experiment, but to implement measures to ensure that the subjects attended to the whole of each nonword. Despite these safeguards only one subject from each group passed the equivalence test, which was in marked contrast to the previous experiments using the abstract, pictorial stimuli where the majority of subjects demonstrated equivalence.

So, why do so many subjects fail when the stimuli are nonwords? It seems that it is the particular characteristics of these stimuli which adversely affect the subject's behaviour. One characteristic of nonwords is that they could be viewed as syllables and the sample and comparison 'syllable' can be compounded to make a so-called 'word'. This occurrence was reported by some of the subjects (in both experiments) for particular reasons (for example, WEFYIM, the Chinese surname, or ZEGYOF, the Hebrew verb). Such a strategy is incompatible with the emergence of equivalence because the stimuli are treated as a unit and can not function independently as is stipulated in equivalence relations. Wulfert, Dougher and Greenway (1991) report this phenomenon occurring when subjects were required to use a, 'think-aloud' protocol during the matching-to-sample tasks. The subjects who passed the tests of equivalence described the relations between the stimuli, whereas the subjects who failed compounded the stimuli. In a follow-up study, subjects were pretrained to either compound stimuli or relationally respond to various stimuli. On a standard matching-to-sample task the subjects in the compounding group failed on the symmetry and equivalence trials where the relational group did not.

Subjects in Experiments 4 and 5 who reported compounding the nonwords did not treat the stimuli as relations, failed on the equivalence tests.
Similarly the contracting strategy reported by a few subjects (for instance, KOJ-YIM becoming KIM or JIM) would have the same effect of turning the trained relations into single units incompatible with equivalence class formation. Wulfert et al

(1991), however, report this occurrence in tasks utilising abstract, pictorial stimuli. So, compounding/contracting is not necessarily exclusive to nonword stimuli. These findings reported here may indicate that it is more likely to occur when nonwords stimuli are presented.

The fact that the stimuli are 'wordlike' may lead them to be related more inflexibly. The nonword relations may be treated by the subject in the same manner as many real word relations (for example, 'pink flower') in that they are related unidirectionally and are not likely to be reversed (yielding, for example, 'flower pink'). Although the two stimuli are not compounded the subject has learnt a two chain relation which is treated as a unit whose components are not independent of each other. Once again such processing is not conducive to the emergence of equivalence.

If the trained relations are inflexible in these ways when the subject is confronted with the untrained test trials he/she may not classify the stimuli into the intended equivalence classes. Some of the test session strategies which were reported bear this out, for example, putting together pairs of stimuli which shared the same initial consonant or matching nonwords, 'which sounded right together'.

This leads on to another characteristic of nonword stimuli, that is they have only one notable feature due to the fact that there is a one-to-one correspondence between the spoken and printed words. An adult human who is fully practised in reading, when presented with a series of unfamiliar words is extremely likely to use the phonology (the sound of the letters), and produce the name. In this way the subjects in these studies would be most likely to use the name alone to classify the stimuli (there were two dyslexic subjects who claimed to have visually discriminated between the stimuli due to their severe reading difficulties). In studies which employ abstract, pictorial stimuli a subject has at least two features of a stimulus to facilitate learning. Firstly, a name which he/she may choose to assign to it; and secondly its physical appearance which is entirely different to that name. This extra feature provides added discriminative power to the stimulus and may help to produce a more robust performance. Control 29, the only subject to pass in Experiment 4, reported that he had trouble remembering the nonword relations and so he attempted to make the stimuli more meaningful by transforming some of the stimuli into real words. Assigning meaning to the stimuli could be an important factor in treating them relationally (for example, A goes with B) as opposed to viewing them as just meaningless strings of letters.

The abstract, pictorial stimuli are meaningless in that they too have no real-world referent but they are given meaning in the process of assigning a name to them. If a subject assigns the name 'house' to a stimulus because a part of it looks like a house then the stimulus becomes a house and so on. Researchers in the field of memory (for example, Craik and Lockhart 1972) specify that stimuli are more effectively remembered and stored in long term memory if they undergo a deep level of processing by which meaning is attached to them (take mnemonics as an example). It is not as immediately obvious to give an individual nonword meaning by changing it into something recognisable. The strategies used by the subjects to give the stimuli meaning (compounding, contracting, initial letters) have not promoted equivalence. If no such strategy is adopted and the subject is just attempting to remember the two nonwords, then the stimuli remain meaningless and are vulnerable to short term memory failure or to outside interference when the novel test trials are introduced. This would result in the subjects making more errors and failing.

It can be concluded that there are certain characteristics of the nonword stimuli which hinder the emergence of stimulus equivalence. So, this manipulation to make the stimuli nameable is not compatible with the equivalence task, suggesting that a more suitable methodology is the, 'name-aloud' protocol adopted in Experiment 3, or using familiar pictorial stimuli.

However, it is important to remember that in these studies it can be reasonably argued that subjects were assigning names to the stimuli and as a result naming them intraverbally (this was explicitly taught in Experiment 5) but this did not lead to the automatic emergence of stimulus equivalence as implied by Horne and Lowe. This could have been (as already argued) because of a specific characteristic of these stimuli; but the same was found in Experiment 3 when familiar names were applied to the stimuli. So, merely naming the stimuli intraverbally is not sufficient for equivalence to emerge, something else must be operating on the subject's behaviour.

It was found that the subjects who failed in these two studies did display a different pattern of responding to the subjects who passed in Experiments 1 and 2 but the same pattern as the subjects who passed in the, 'name-aloud' study, Experiment 3. Specifically, a distinction was made between the baseline and symmetry tasks. It was during these three studies (3, 4 and 5) that it can be said with some certainty that subjects were assigning individual names (appropriately or not) to the stimuli. Throughout the five studies the research of Bentall et al (1993) has been cited, regarding the observed two patterns of responding in the latencies of subjects who successful demonstrated equivalence: (i) a flat pattern to which they attributed to common naming; (ii) a pattern whereby the trials which involved transitivity produced significantly more errors and took significantly longer, the authors said this could be accounted for by Field's, 'Associative Distance Effect'. It was further argued that in line with a naming hypothesis (Dugdale and Lowe 1990; Horne and Lowe, 1996) this pattern is not incongruous to the subjects assigning individual name to the stimuli. Yet in the three studies where it can be said almost unequivocally that this is what the subjects were doing, a different pattern was consistently observed.

So, a dichotomy is observed. One pattern is seen when subjects use a verbal strategy (regardless of them passing or failing) and another pattern is noted (in Experiments 1 and 2) where it is not known for certain what the subjects are doing, but there is the implication that the strategies are nonverbal. This means that adult subjects do not necessarily name the stimuli in order to demonstrate equivalence.

Naming Issues

Experiments 4 and 5 not only compared the groups' performances on the matching-to-sample task but also pursued further the dyslexic/non-dyslexic naming differences. Experiments 2 and 3 measured the subjects' naming latencies using a discrete-trial methodology with firstly familiar stimuli and then the experimental stimuli (which involved measuring the familiar names assigned by the subjects to the abstract, pictorial stimuli). No significant group differences were demonstrated using any of the stimuli.

Experiments 4 and 5 afforded another opportunity to measure potential group differences utilising a different type of stimuli - nonwords. Like on the previous discrete-trial naming test would no group differences be observed using the 'wordlike' stimuli? In addition, it was decided to introduce an alternative type of naming test on which reliable naming differences had been observed in adult subjects; that is the Rapid Automatised Naming test. Would this methodology bring out any group differences?

In Experiment 4, no significant group differences were observed between the groups on any of the discrete-trial subtests. However, the groups in Experiment 5 demonstrated a significant difference on the nonword subtest. Although the groups in Experiment 4 did not significantly differ on the nonword subtest it was on this subtest where the largest difference between the two groups was observed (162.47ms).

This all suggests that the dyslexic subjects produce longer latencies on the nonword subtest and it was using only these stimuli where significant differences were found using the discrete-trial methodology. Why should this be the case? The other stimuli so far measured have been familiar stimuli (pictures, colours, numbers and letters) which are well practised in adult subjects. Even if dyslexic children take significantly longer to produce these stimuli under discrete-trial conditions (Denckla and Rudel,1976b; Perfetti, Finger and Hogaboam, 1978; Stanovich, Freeman and Cunningham, 1983; Bouma and Legein, 1980; and Fawcett and Nicolson, 1994), by the time they are adults they will have had much experience in producing the names of such stimuli and may have compensated for their differences.

The discrete-trial naming test in Experiment 3 was slightly different in that the latencies of the names the subjects assigned to the experimental stimuli were measured. However, subjects were required to access familiar names which were chosen because the stimulus physically looked like the object in question. Therefore, it was almost as if the subject is naming familiar pictures, which was again a practised task. The stimuli are novel and have never been paired with this particular name but the naming test was administered after two hours of matching-to-sample sessions where the subject was required to name the stimuli aloud. So, even if it is not accepted that this task was a familiar one (as argued above), it was a practised one, be it only a couple of hours practice. The nonword naming task was a truly novel task. When the subjects were asked to name the nonword stimuli they had never seen them before (except once during the practice session). There was no opportunity for compensation and this could by why the group differences were observed using these stimuli.

A difference may also have been observed on nonword stimuli due to the fact that the task requirements are different in that the subject was required to read a whole word (not just a letter or number or to access a name based on a non-graphical representation as with the colours and pictures). This was reflected by Watson and Brown (1992) who found a significant difference between dyslexic and non-dyslexic adults when they were required to name real words aloud. So, it may be the reading component of the task which serves to emphasise the differences between the groups.

For all the subjects in Experiment 4 the nonwords subtest took significantly longer than the others and the colours subtest took significantly less time. In Experiment 5 no significant differences were observed between the subtests due to the fact that the two groups took the longest on different subtests (the dyslexic group on the nonwords; and the control group on the colours). It must be noted that the control mean on the colours subtest is skewed by one subject's data. If this subject's median is eliminated from the analysis, the colours mean is reduced to below that of the pictures and nonwords. A significant difference was observed between the groups in

Experiment 4 on the nonword RAN test. Similarly the groups in Experiment 5 significantly differed on the nonword RAN subtest when it was included as one of the five subtest administered. This demonstrates a consistent difference between dyslexic and non-dyslexic individuals on this measure. The analysis in Experiment 5 showed that the dyslexic subjects were more adversely affected by the nonwords, whereas the control group took longer on the pictures subtest.

So, what as a result can be said of the dyslexic/non-dyslexic naming differences? A naming difference is observed between the adult dyslexic and non-dyslexic populations but only using non-practised stimuli - nonwords. This mirrors the naming differences found in childhood using familiar stimuli which at that stage in development are not well practised either. By adulthood any naming discrepancy on familiar stimuli disappears due to compensation but the difference does persist and can be observed using novel stimuli such the nonwords. This difference can be seen using the discrete-trial methodology (Experiment 5) but is more reliably seen using the RAN test where a subject's responding is placed under the pressure of time and is in effect magnified. This supports Bower and Swanson (1991) who too demonstrated discrete-trial and continuous list differences in the same set of subjects.

In Chapter 2 it was argued that some authors (such as, Perfetti, 1985; Stanovich et al, 1983; and Wolf and Goodglass, 1986) have suggested that the RAN test involves more than just naming and involves other cognitive processes which may also differentiate good from poor readers. An attempt to address this issue was made by correlating a subject's performance on the discrete-trial methodology with that on the RAN test. In both samples (Experiments 4 and 5) a significant correlation was found between the two measures suggesting a relationship between a subject's performance on both measures. So it could be said that a deficiency on one test predicts a deficiency on the other indicating that they are both valuable measures of naming speed.

Crucially, even if it can be argued that the RAN procedure is limited as a measure of naming because it is contaminated by additional cognitive demands it does have strong similarities with Horne and Lowe's definition of the intraverbal naming which should occur during the test session if equivalence is to emerge, in that they both demand continuous naming and as a result should share some underlying sub-processes. As a result, attempting to establish a naming latency difference between the groups on this measure can be seen to be appropriate in this context despite it not being the most pertinent measure of pure naming.

Naming tests were introduced into the studies to establish that there was a significant difference in naming between the dyslexic and non-dyslexic groups. If this could be said, then it follows (according to the naming hypothesis) that the two groups will significantly differ on the matching-to-sample test trials latencies where naming is also occurring for equivalence between the relations to be successfully demonstrated.

The significant naming difference between the groups was observed using nonword stimuli, but on the matching-to-sample tasks involving these stimuli all but three subjects failed to demonstrate equivalence. Even though the subjects were thought to be naming intraverbally, little can be said about the role of naming in stimulus equivalence. If subjects had passed but still demonstrated no significant latency difference it could be hypothesised that naming was not taking place on these trials, but this was not the case. What can be said, however, is that simply requiring the subjects to name intraverbally does not always automatically bring about stimulus equivalence, as predicted by Horne and Lowe (1996). These studies show that something more is needed and it has been hypothesised that this is to treat the stimuli relationally either explicitly (as in a rule A goes with B or A is the same as B) as Rhys the subject who passed in Experiment 4 did, or implicitly. The subjects who passed in Experiment 5 reported to have formed the equivalence classes during training.

In addition, the subjects were almost definitely naming the stimuli (and failing) and were known to differ significantly in naming speed, but this was not reflected in their test session latencies. It could be argued that it was wrong to predict that this would be the case in the first place. However, it could also signify that the naming difference is being swamped by other processes coming into operation given that these were the latencies of subjects who failed. Put another way, responding to a test session trial could involve more than just naming the stimuli. So the dyslexic subjects may take significantly longer to produce the names of the stimuli but may not take significantly longer to respond especially with the subjects from these experiments who were failing and who did not realise what the task requirements were and as a consequence may have been deliberating over their choice. Just 'eyeballing' the data from the five studies it can be seen that for the first three studies the average overall mean latency for the subjects who passed was approximately 3000ms but for the subjects who fail in Experiments 4 and 5 it was roughly double that suggesting that something other than naming and selecting the appropriate response was taking place. It could be this extra processing which is drowning out the significant naming latency.

Finally, the more general question arises of why naming should in some cases facilitate equivalence (as in Experiments 1, 2 and 3) but in others (Experiments 4 and 5) actively prevent it from emerging? Is it simply, as was argued previously, that a characteristic of word-word relations is that they are not obviously reversible in everyday experience and therefore, this established behaviour could override the emergence of the new equivalence relation?

What must be concluded is that intraverbalising is not the magic ingredient which if present always leads to the demonstration of stimulus equivalence as Horne and Lowe assume.

# Chapter Six

# Conclusions

## Conclusions

The purpose of this thesis was to attempt to present evidence to evaluate the naming hypothesis (as first outlined by Dugdale and Lowe, 1990; but more recently described in detail by Horne and Lowe 1996). Specifically, the naming hypothesis states that naming (common naming or intraverbal naming) is necessary for stimulus equivalence to emerge (although Horne and Lowe concede that there is a second possible 'contingency-shaped' route through which nonhuman subjects could also successfully demonstrate equivalence, although they argue there is little evidence to support this).

Therefore, a population of subjects who were developmentally stable, of average intelligence, but characterised by a deficit in naming were compared to control subjects on the various matching-to-sample tasks, in order to examine the role played by naming during these procedures. The fundamental experimental hypothesis was that as a result of their naming weakness, more dyslexic subjects should fail the tests of equivalence and produce longer latencies on the test session trials, if naming is indeed necessary for equivalence.

Alternatively, the null hypothesis would be that no differences would be observed between the two groups on the test session. This would either imply that: (i) it is questionable whether adult dyslexic subjects persist to demonstrate a naming impairment; or (ii) the dyslexic subjects' performances are on a par with those of the control group because they are using a nonverbal strategy which is just as effective as a verbal one for passing the equivalence tests.

It is of value to note here that the adoption of the null hypothesis is only reliable if the design of the experiment also possesses a high level of power. As already discussed within the context of Experiment 3, if an experimental design has low power and no significant effect is found, it can not be concluded with any confidence that this is because no effect exists. There is always the possibility that there is a significant difference between the groups, however, the design is not powerful enough to detect it.

It has already been noted that the power of the group comparison on the test session latencies in Experiment 3 (where the largest mean difference between the groups was seen) was low in that, given the parameters of the data, there was only a 17% probability of correctly rejecting a null hypothesis. Cohen (1988, 1992) recommends aiming for a power level of 0.80 (80%). In order to attain such a level for Experiment 3, 41 subjects per sample would have been required to confidently adopt the null hypothesis and state that there is no significant effect present.

Several issues arise from this point. The power of a given experiment is most usefully calculated in the design stage of a study, so that the appropriate number of subjects can be employed. Therefore, when calculating the power of a statistical test, the expected effect size must be known or estimated. In Experiment 3 the effect size, using visual inspection, appeared large in terms of cognitive processing time (1783.18ms). Therefore, if the power level had been considered pre-experimentally, it would have been concluded that the effect size is large and as a result a small sample size would be sufficient. However, in the specific context of a matching-to-sample task is the mean difference a large one? It has been previously outlined that responding on a typical MTS trial involves more than just a simple reaciton time. Differences in such complex responses may need to be large in order to be notable. The point is that considerable care must be taken when estimating the effect size for a particular variable, what might be valid for one particular protocol might not be for another.

Similarly, an appropriate power level must be decided upon. The alpha level of a statistical test is conventionally set at 95% (that is there is 95% chance of correctly adopting the experimental hypothesis). If the corresponding beta/power level is set at 95% (that is there 95% chance of correctly rejecting the null hypothesis) it often results in requiring an impractically large sample size. For instance, in order for Experiment 3 to have high power, 41 subjects would be needed per group. Even if the effect size is considered to be large in Cohen's terms (see Cohen, 1992) 26 subjects are needed per group. Such a number of subjects may seem reasonable for a simple reaction time experiment but given the nature of learning experiments which are often intensive and time consuming such a sample size is unworkable. So, a trade-off must occur in such a situation whereby the Experimenter settles for lower power. However, a small sample size should still be effective if the procedure is experimentally rigorous.

So, what can be concluded about statistical power? It is an important and should be considered and there is no doubt that the power of the experiments reported is low due to the small sample sizes employed, so the adoption of the null hypothesis must be viewed with caution. In the case of Experiment 3 where the effect appears to be large it may not signify the absence of an effect. On the other hand, having confidence in the power level of the experiment means unequivocally knowing the effect size under investigation and using a methodology which favours using large sample sizes or indeed having an extensive population to draw the subjects from. In the case of learning schedules or employing a subset of the general population (that is dyslexic subjects) this may not always be practical and as a consequence high power may have to be sacrificed.

In this way it can be seen how the data can be interpreted on two levels. The first implication would provide an insight into the nature of developmental dyslexia in adults, and in particular the phonological processing deficit, as characterised by a naming weakness. The second questions the role of naming in the demonstration of stimulus equivalence. Is naming necessary to bring about equivalence? Is it possible that equivalence could occur even when the subjects are not naming the stimuli?

**Naming Issue**

If any valuable conclusions were to be made it had to be established that dyslexic and non-dyslexic subjects differed in naming speed.

In Experiment 1 no measure of naming latency was included in the test battery on the basis that naming differences had been reliably found in childhood on discrete-trial tests (Bouma and Legein, 1980; Perfetti et al 1978; and Stanovich et al 1983) and this deficit appeared to persist into adulthood (Fawcett and Nicolson, 1994; Watson and Brown, 1992).

No equivalence test-trial latency differences were observed in this study so it was decided to include a discrete-trial naming test in Experiment 2 to empirically measure the groups' naming speed. No significant differences were found on this measure. However, this task employed familiar stimuli (pictures, colours, letters and numbers). It had been predicted that if the groups significantly differed in naming using familiar stimuli then it follows that they would also significantly differ in the naming of the experimental stimuli. The non-significant difference with familiar stimuli does not rule out the possibility that a naming difference still existed for the experimental stimuli. Naming familiar stimuli would be highly practised for adult subjects meaning that the dyslexic subjects may have compensated for their difficulties. Compensation would be much less likely to occur with unfamiliar stimuli. So, perhaps the dyslexic subjects would be slower than controls at naming such stimuli. If that were the case and there were still no differences in subsequent equivalence test latencies it would suggest that the dyslexic subjects were employing a nonverbal strategy in passing the equivalence tests.

In Experiment 3 due to procedural changes it was possible to post-experimentally measure the subjects' naming speed on the actual experimental stimuli. Again, however, no significant difference was found between the groups on the discrete-trial naming test. This was a surprising finding given that this test required the naming of novel stimuli. From these data it was concluded that the dyslexic naming deficit can be seen to have been compensated for in adulthood when: familiar stimuli are used; familiar names are applied to abstract stimuli; and when the stimuli are presented individually. Compensation occurred to such an extent that even with abstract stimuli, the dyslexic group's performance was equal to that of the control group after only a few hours practice. Developmentally speaking, this would suggest that dyslexic individuals can produce a performance on a par with that of non-dyslexic individuals, but it just takes them longer to achieve parity.

It was argued that this finding, however, did not mean that no naming deficit whatsoever existed for the dyslexic subjects. Watson and Brown (1992) found significant naming latency differences between samples of dyslexic and non-dyslexic college students when the stimuli presented were printed words. In addition if it is hypothesised that the naming most likely to occur on visual-visual matching-to-sample tasks is intraverbal naming (in which the pairs of stimuli to be matched are rehearsed continuously), then it is possible that discrete-trial naming (in which rehearsal does not occur) may not be the most accurate reflection of the kind of verbal behaviour employed during equivalence tests. Therefore, a RAN test procedure (which involves continuous naming) was introduced on which dyslexic and non-dyslexic adult samples have also been found to differ significantly (Kinsbourne, 1990; Felton et al 1990; Wolff et al 1990).

In Experiment 4 a significant difference was found between the groups using nonword stimuli (which were the experimental stimuli) using the RAN methodology, with the dyslexic group producing longer latencies. This finding was replicated in Experiment 5. In addition, a significant difference was found between the groups on the discrete-trial procedure using nonword stimuli in Experiment 5.

It was argued that nonword stimuli would create more problems for the dyslexic subjects because they were totally novel and therefore, could not have been compensated for. They were also more complex stimuli which relied heavily on phonological processing as the subjects were required to read. So, it should be concluded that the dyslexic naming deficit does persist into adulthood but that it is only reliably observed under certain conditions. Those being: when the subjects are required to name continuously under the pressure of time; or when the stimuli involve non compensatable phonological processing.

It must be noted that these findings imply that caution should be taken when employing dyslexic adults as subjects. Specifically, it has been seen how findings reliably observed in childhood can not always be measured in an identical form in later years because of the dyslexic subjects' capacity to compensate for their deficiencies. This points to a problem of diagnosing dyslexia in adulthood. If a symptom is not observed this does not necessarily mean that it does not exist. It could signify that the individual has, with practise, compensated for the difficulty but that under certain conditions where the channels of processing are 'overloaded' (that is by the implementation of time restraints or more complex stimuli) the deficit becomes evident once again. From a practical perspective, if a 'purer' example of dyslexia is required then it seems wise to study dyslexic children who have not yet had the opportunity to compensate for their difficulties.

**Stimulus Equivalence**

The rationale behind the implementation of the equivalence paradigm was twofold. Firstly, if it could be used to discriminate between the two groups it may have a practical diagnostic purpose. Secondly, if it did discriminate between the two groups especially on time taken on the test session trials and they were also found to significantly differ in naming speed then this would imply that naming played a role in the successful demonstration of equivalence.

The two samples were measured on various aspects of the task and across the five studies under conditions which did or did not encourage the subjects to name.
The measures taken on each study were:
(i) the amount of training required to learn the initial baseline relations.
(ii) the maintenance of this learning.
(iii) pass/fail rate.
(iv) test trial latencies.
It was predicted that the two groups would differ on these measures and this could be related to the dyslexic subjects' naming impairment.

It should be noted that subjects do not need to name the stimuli in order to learn the baseline relations (because nonhumans can learn them). However, responding differently to sample stimuli (for instance by naming them) does facilitate the acquisition of these relations (Cohen et al, 1976; Sidman et al, 1982; Urcuioli, 1985).

In addition the question of maintenance may be indicative of a general dyslexic deficiency (that is not specific to naming) due to a lack of confidence in his/her performance resulting in the dyslexic subjects being more reliant on feedback.

Looking at the pass/fail rate on the matching-to-sample test session could reveal something about whether subjects were or were not assigning names to the stimuli but it could also reflect the use of alternative strategies which are just as effective as naming.

What strategies were being used by the subjects was difficult to establish. There were the post-experimental reports but these should be viewed with caution as it is not known whether they reliably reflect what was actually happening at the time of testing or whether they consisted of strategies which had been formulated as a result of subjects passing/failing. The only direct evidence concerning whether subjects were naming or not during the test session is derived from examining the test trial latencies, where any between-group naming differences should be reflected.

One way of determining subjects' strategies was to impose one on them. For example, in Experiment 3 subjects were required to name the stimuli out loud throughout the experimental sessions and in Experiments 4 and 5 stimuli were chosen which were wordlike and therefore highly likely to be named. This meant that the groups could be compared on the above measures in the knowledge that they were consistently naming. In contrast, Experiments 1 and 2 employed abstract, pictorial stimuli (which had no preestablished names) and subjects were left to their own devices when it came to learning the relations. As a result it could not be ascertained definitively whether the subjects who passed the equivalence test were or were not assigning names to the stimuli.

However, this does provide a valuable comparison in that there were three studies (3,4,5) where it was known that the subjects were naming and two where there was uncertainty over what the subjects were doing. If the groups' performances were the same across all five studies then the possibility remained that the subjects in Experiments 1 and 2 were indeed also naming. On the other hand, if their performances were different in any way from those in Experiments 3-5, it would raise the possibility that the former were not using a verbal strategy in successfully demonstrating equivalence in the case of the test session measures.

(i) Training - Errors to Criterion

The only significant difference between the two groups found on this measure were seen in Experiment 2 with the dyslexic subjects making more errors. Why should differences be found only in this experiment?

Experiment 1 utilised a simple design, there were only two comparisons per trial and only four relations to learn, and thus as the task was a relatively easy one compared to those that followed. This meant that perhaps dyslexic and non-dyslexic differences would not have been as apparent. Experiments 3, 4 and 5, unlike Experiment 2, required the subjects to name the stimuli. This indicates that naming helped both groups to learn the relations and that in these three experiments naming was consistent from the outset.

It is interesting that Experiment 2, which did not require stimulus naming was the experiment in which the dyslexic subjects made many more errors than the controls, and demonstrated a tendency to require more repeat sessions. This can be seen by looking at Table 23 which shows the errors to criterion made by the two groups on each experiment. This would suggest that these dyslexic subjects took longer to 'lock into' a strategy whereas in Experiments 3, 4 and 5 a strategy (to name the stimuli) was imposed on them from the very start. It could have been that in Experiment 2 the dyslexic subjects may have started by remembering the stimuli visually (aware of their naming weakness) but this was unsuccessful so then after a time they decided to use names. Alternatively, they could have been naming from the outset but inconsistently (in line with their weakness) and they therefore, took longer to produce an error-free performance. They may even have avoided naming completely.

**Table 23: Errors to criterion made by the two groups on Experiments 1- 5. Standard deviations are presented in brackets.**

| | Expt.1 | Expt.2 | Expt.3 | Expt.4 | Expt.5 |
|---|---|---|---|---|---|
| **Dyslexic Group** | 29.44<br>(17.33) | 93.63<br>(87.95) | 17.11<br>(18.39) | 89.38<br>(68.85) | 43.89<br>(29.28) |
| **Control Group** | 19.89<br>(18.16) | 21.63<br>(21.27) | 9.33<br>(7.68) | 45.63<br>(62.28) | 35.33<br>(21.07) |

If a comparison is made between Experiments 2 and 3 it can be seen that there was a huge improvement for the dyslexic subjects when they were required to name the stimuli. Fewer errors were made by the dyslexic subjects in Experiment 5 than in Experiment 4 which incorporated the PAL methodology. This suggests that the dyslexic subjects in Experiment 4 took longer to settle on a consistent strategy because their performance should have otherwise been similar.

The overall conclusions which can be made from these data is that there is strong evidence to support the notion that naming facilitates the learning of arbitrary stimulus relations via matching-to-sample for both groups and the dyslexic subjects in particular benefit. In addition, the procedures which encourage the subjects to name (Experiments 3 and 5) requires them to adopt a consistent strategy which may have otherwise taken the dyslexic subjects a long time to arrive at given their naming weakness.

(ii) Training - Unreinforced Errors

In Experiment 5 a significant difference was found between the groups on unreinforced training, with the dyslexic subjects making more errors. A similar trend was found in Experiment 1. Why should the removal of reinforcement affect the subjects in these two experiments and not the others? It could be due to the amount of exposure to the matching-to-sample task in that the dyslexic subjects may require 'overlearning' in order to perform on a par with non-dyslexic individuals. In Experiment 1 there was no such opportunity for dyslexic 'overlearning' as all subjects acquired the baselines quickly and this could have resulted in them making more unreinforced errors.

By contrast, in Experiment 2 the dyslexic subjects made a huge amount of errors and as a result required more repeat training sessions. So, in effect, because they made more errors they received more exposure which may have been crucial to them maintaining this learning.

In Experiments 3, 4 and 5 it has already been noted that the subjects' consistent naming facilitated the learning of the baseline relations and it seems also to have facilitated their maintenance in unreinforced conditions because in Experiments 3 and 4 no significant differences between the groups were observed. A significant difference was, however, demonstrated for Experiment 5 which differed from the previous naming studies because two methodologies were employed during training (PAL followed by the MTS sessions). Once the subjects reached the MTS unreinforced session they had only undertook one previous MTS session. This may have unduly affected the dyslexic subjects' performances in that they had little experience of this type of protocol and thus may have been unsure of their performance using this methodology. The notion that the dyslexic subjects may need to 'overlearn' the relations before they are maintained without feedback ties in with the idea that dyslexic individuals can achieve the same level of performance as the control subjects, but that it just takes them longer to get there (Miles and Done, 1978).

(iii) Pass/Fail Rate

**Table 24: Each group's pass rate for Experiments 1-5 expressed as percentages.**

| | **Expt.1** | **Expt.2** | **Expt.3** | **Expt.4** | **Expt.5** |
|---|---|---|---|---|---|
| **Dyslexic Group** | 44% | 62.5% | 44% | 0% | 11% |
| **Control Group** | 78%[1] | 75% | 67% | 14% | 11% |
| *Difference* | 34% | 12.5% | 23% | 14% | 0% |

Looking at Table 24 it can be seen that in Experiments 1 and 3 the biggest difference in pass rate between dyslexic and control subjects came in Experiments 1 and 3. Whereas in Experiment 2 there was a minimal difference between the groups. The dyslexic pass rate for Experiment 1 may be misleadingly low due to a procedural anomaly where subjects were only presented with one unreinforced session before proceeding on to the test session despite the number of errors made. This presents the possibility that the baseline relations may not established enough to be maintained without feedback. It already been noted that the dyslexic group made a large number of errors on this phase.

Could this reveal anything about the subjects' underlying strategies in Experiments 1 and 2 where it is not known what the subjects were doing? For instance, if the subjects in Experiment 2 were naming the stimuli then why did they do better than the subjects in Experiment 3 who were required to name? These data suggest more strongly the opposite, that is

[1] Calculated to include Control 7

in Experiment 3 requiring the subjects to name intraverbally disrupted their performance whereas subjects in Experiment 2 did not name intraverbally and were, as a result, more likely to pass. This is contrary to Horne and Lowe's predictions that if intraverbal naming occurs stimulus equivalence will emerge (that is naming is sufficient for equivalence).

Similarly, it was argued at the end of Experiment 1 that the stimuli were perhaps too easy to name, so making them more difficult to discriminate between might magnify any differences between the dyslexic and control groups. In fact the opposite occurred; the dyslexic performance improved for Experiment 2 to become almost on a par with the controls; it then deteriorated again in Experiment 3 when naming was required. So, perhaps naming facilitated the learning of the baseline relations for the dyslexic subjects but their lack of ability in manipulating verbal material (Vellutino, 1987) mitigated against them passing the tests (perhaps because, for them, the baseline relations were rigid).

Most importantly, the introduction of naming seemed to make subjects' performances deteriorate which is not predicted by the naming hypothesis. Subjects in Experiment 3 were definitely naming intraverbally and consistently which should, according to Horne and Lowe, lead to the demonstration of equivalence and yet eight subjects failed the test session. Horne and Lowe state that naming would not lead to stimulus equivalence if it is incongruous with experimenter-defined classes but do not give an example of what is meant by this with respect to intraverbal naming.

For Experiments 4 and 5 the majority of subjects from both groups failed. What does this reveal about the role of naming in stimulus equivalence? Again it seems that when the subjects were naming more subjects failed. However, Experiments 4 and 5 were different in that it appeared that the reason the subjects failed was because of the nature of the actual stimuli. The fact that they were nonwords prevented the emergence of equivalence in that: compounding made the relations too rigid; the stimuli could be broken down into their component parts (as seen in Experiment 4); subjects have a history of using words which does not broadly speaking include reversing word orders; and finally they were meaningless and therefore, harder to remember.

In conclusion, three findings emerge from these data: (a) a comparison of Experiments 2 and 3 suggests that subjects in Experiment 2 were not naming; (b) the evidence from Experiments 4 and 5 in particular, suggest that naming, rather than facilitating the emergence of equivalence, can in some cases be detrimental to the establishment of equivalence classes; (c) intraverbal naming (as seen in Experiment 3) is not the magic ingredient which automatically leads to the demonstration of equivalence. Either something else is needed or in this instance the fact that the subjects were naming intraverbally had nothing to do with them passing or failing.

(iv) Test Session Latencies

This measure was of importance because it offered a means of assessing whether subjects were or were not naming on the test session trials.

Looking at the subjects who passed on Experiments 1, 2 and 3 no significant naming latency differences were observed and no test trial latency differences were observed. For Experiments 1 and 2 it was not known whether subjects were or were not naming. It has been tentatively argued that they were not, therefore, no group differences in test session latencies would be necessarily expected. However, in Experiment 3 subjects were known to be intraverbally naming but despite this again no test session latency differences were observed. This seems to be evidence which strengthens the previous suggestion that the dyslexic and control groups did not differ in the naming of these experimental stimuli either under discrete-trial naming or matching-to-sample conditions.

In Experiments 4 and 5 significant naming differences were found using the experimental stimuli. However, unfortunately, the majority of subjects failed to demonstrate equivalence meaning that nothing could be concluded regarding the role of naming in the successful demonstration of equivalence.

In addition, across the five studies the test session latency data was also considered as a whole in order to investigate what the group's patterns of responding could reveal regarding the subjects' underlying strategies.

Bentall et al (1993) identified two patterns of responding in their subjects' test session response latencies; (a) a flat pattern whereby all the test categories take equal amounts of time which implies that subjects are assigning common names to the stimuli; and (b) a pattern where the trials involving transitivity take significantly longer than the other tasks.

For Experiments 1 and 2 the latter pattern: baseline and/or symmetry < transitivity was observed and it was argued that this pattern is consistent with subjects assigning individual names to the stimuli in that the more complex intraverbal strings (produced on the transitive relations) should take longer to produce. However, for Experiment 3 it was known that the subjects were intraverbally naming yet a different pattern was observed: baseline < symmetry/transitivity. The pattern (that is baseline < symmetry) was replicated again in Experiments 4 and 5 where it was known the subjects were assigning individual names to the stimuli but this time the subjects were failing.

This would indicate that the initial assumption was wrong, and that this latter pattern is the one which characterises the usage of a verbal strategy. This makes sense in that what differentiates this pattern from the first one is that the symmetry tasks take significantly longer than the baseline tasks. For subjects who have a history of applying words in a given order a task requiring them to reverse this word order would be an unfamiliar one and as a result it should be expected to take longer.

The data from Experiments 4 and 5 suggest that this pattern is not restricted to subjects who pass the tests but it reflects more the use of names rather than a successful strategy. It does, however, suggest that some amount of 'effortful' processing was occurring in Experiments 4 and 5 in that if subjects were merely responding at random all tasks would undergo more or less the same amount of processing.

This raises the question of whether this consistency in patterning across subjects who passed or failed was seen in Experiments 1 and 2? The data of the subjects who failed (n=7, minus the data of Control 7) in Experiment 1 were analysed (it was not possible to analyse the data from Experiment 2 because the numbers of subjects who failed was too small) using a one-way ANOVA ($F=30.37$, $df=2, 20$, $p<0.0001$). Tukey's HSD tests ($q\ (3,12) = 534.03$ at 0.05 level of significance) show that: baseline/symmetry < transitivity. This is a similar pattern to the one found with the subjects who passed in Experiment 1 and the same pattern as seen in Experiment 2. So, it seems that this pattern again is a consistent one and is not associated exclusively with passing the tests.

The final issue is that if it can be argued that the pattern demonstrated in Experiments 3, 4 and 5 reflects a verbal strategy, then given that the pattern seen in Experiments 1 and 2 is different, it may reflect a different type of processing, possibly nonverbal. This is in line with the tentative conclusions drawn on the earlier measures.

Although no significant differences were found between the dyslexic and non-dyslexic subjects on the matching-to-sample sessions this thesis should hopefully show that this line of enquiry is a fruitful one. The use of adult subjects may not be appropriate for the reasons already outlined (that is dyslexic adults are more likely to have compensated for their naming weaknesses). However, younger subjects who can demonstrate a naming deficit for a wide variety of stimuli may prove to be entirely suitable for such procedures.

Also, the significant difference found in the naming of nonword stimuli (Experiments 4 and 5) should be followed up. The majority of subjects failed when tested using a procedure whereby all the tests of equivalence were presented randomly, but this is not to say that they would fail using all test session protocols. As well the nonword stimuli could be arranged to minimise the possibility of subjects grouping them during the equivalence tests in a manner which is incongruent with the intended classes.

**The Naming Hypothesis**

Chapter 1 traced the history of stimulus equivalence back to the philosophers of the 18th and 19th centuries and demonstrated how learning theory had progressed down the ages. Whereas Hartley (1705-1757) and Mill (1773-1836) talked of associations between ideas and behaviour brought into effect by contiguity, modern behaviourists talk of relations and the effect of contingencies in the environment. From the training of certain stimulus-stimulus relations stimulus equivalence was demonstrated. Emergent relations were observed which could not be sufficiently explained by existing theories. There was as a result the need for a new hypothesis to accommodate it.

The focus turned to one such hypothesis in which the response term is vital. Horne and Lowe (1996) emphasise how it is the nature of the name relation which promotes stimulus equivalence. It is the fact that the name relations are bidirectional and this property emerges without reinforcement that leads to stimulus-stimulus symmetry (Dugdale and Lowe 1990).

It was specified that the naming most likely to occur in the visual-visual task is intraverbal naming which with self-repetition becomes bidirectional in that when either component is produced the other is emitted (without reinforcement) along with the appropriate listener behaviour. This is likely to occur when a subject assigns individual names to each stimulus and results in an intraverbal equivalence class.

Although there is no evidence presented in Experiments 1-5 which unequivocally disproves such a naming hypothesis, the data have raised several issues worthy of further consideration.

It has been argued that the subjects who passed the test of equivalence in Experiments 1 and 2 were <u>not</u> naming the stimuli. Horne and Lowe in their account of stimulus equivalence have presented a dual-route model via which equivalence could emerge. Firstly there is the route whereby naming is necessary for equivalence to emerge and secondly there is another contingency-based route by which nonverbal subjects could demonstrate equivalence. Here there is evidence to suggest that verbally competent adults when given the choice appear to have chosen a nonverbal strategy. This would suggest that naming is not always necessary for stimulus equivalence to emerge but neither is it always the route taken by verbal humans.

The evidence presented from Experiments 4 and 5 especially, suggests that naming the stimuli can be detrimental to the emergence of equivalence. It seems that because of the history names and especially intraverbals carry with them, that is, they are usually produced in a set order (for instance, black and white; sugar and spice) means that when produced in an experimental context, for example, they may become rigid units which are unlikely to be reversed according to the properties of equivalence.

Following on from this the data from Experiment 3 suggests that intraverbal naming does not always automatically lead to the demonstration of equivalence. In favour of the naming hypothesis it could be argued that a specific type of intraverbal naming is needed whereby the relationship between the stimuli is established. That is the subject must be aware that A is 'the same' as B or that B 'is interchangeable' with C and so on (this could be either implied or explicitly stated as in a rule). However, how could this type of intraverbal naming ever be quantified independently of an equivalence test? The alternative argument is that quite simply naming and specifically intraverbal naming has little to do with subjects passing or failing. In other words, if it occurs it does not mean that a subject is more likely to demonstrate equivalence.

This thesis began by arguing that the strength of the naming hypothesis over other theories (for instance, those put forward by Sidman and Hayes) was that it could be unequivocally falsified. Horne and Lowe's dual-route theory has somewhat muddied the waters and served to make their theory apparently untestable in that they acknowledge that nonverbal subjects could potentially demonstrate equivalence be it under a different source of control than usually observed.

In order to extract a usable theory it may be necessary to get back to basics and state that subjects must name in order to successfully demonstrate equivalence. This is what Lowe and Horne (1996) strive to do in their response to the commentaries made on their original paper, by stating that naming is both sufficient and necessary for stimulus equivalence to emerge (p. 328-334).

This thesis has presented evidence which raises questions regarding Horne and Lowe's predictions. They imply that naming is necessary for equivalence, yet there is some reason to believe that subjects in Experiments 1 and 2 may not have named the stimuli but still successfully demonstrated equivalence. There is yet stronger evidence from Experiments 3, 4 and 5 to suggest that intraverbal naming is not sufficient to bring about equivalence. It could be that to accommodate these data the naming hypothesis needs more qualification. What is evident though is that the naming issue is not as straightforward as first maintained.

Lowe and Horne themselves say (p.330):

> It must...be reckoned one of the odd ironies of work in this area of behaviour analysis that a behavioural variable (that is, verbal behaviour) that is known to bring about success on match-to-sample tests of equivalence, and that can be directly manipulated with major and immediate effects, has been almost completely ignored by researchers..

This thesis has hopefully presented a new means by which the role of naming can be directly assessed and generated some issues which merit further investigation.

# Appendices

## Appendix A

| Subject Number | Age | Occupation | BDT Score | Spelling (out of 30) |
|---|---|---|---|---|
| Dyslexic 1 | 25 | Student | 6+ves | 19 |
| Dyslexic 2 | 21 | Student | Diagnosed elsewhere | |
| Dyslexic 3 | 21 | Student | 4.5+ves | 15 |
| Dyslexic 4 | 24 | Student | 4+ves | 16 |
| Dyslexic 5 | 29 | Student | Diagnosed elsewhere | |
| Dyslexic 6 | 32 | Student | Diagnosed elsewhere | |
| Dyslexic 7 | 29 | Student | Diagnosed elsewhere | |
| Dyslexic 8 | 28 | Student | Diagnosed elsewhere | |
| Dyslexic 9 | 24 | Student | 4+ves | 13 |
| Control 1 | 18 | Student | 1+ve | 26 |
| Control 2 | 24 | Student | 1.5+ves | 27 |
| Control 3 | 24 | Student | 1.5+ves | 27 |
| Control 4 | 23 | Student | 1.5+ves | 28 |
| Control 5 | 22 | Student | 1+ve | 18 |
| Control 6 | 24 | Student | 2+ves | 24 |
| Control 7 | 20 | Student | 2+ves | 28 |
| Control 8 | 20 | Student | 2.5+ves | 16 |
| Control 9 | 22 | Student | 2.5+ves | 9 |

**Key:** **BDT = Bangor Dyslexia Test (Miles, 1982)** **Spelling = Schonell Spelling Test**

## Appendix B

### Post-Experimental Questionnaire Expt.1

(1) What do you consider the purpose of the experiment was?

(2) What did you have to do?

(3) Did you think the shapes were related to each other? If so, how were they related?

(4) Did you give any of the shapes names? If so, what names did you give to the shapes?

(5) Did you find any part of the experiment more difficult than the others? If so, in what way?

(6) Have you any further thoughts or comments concerning the experiments?

<u>**Appendix C**</u>

| Subject Number | Age | Occupation | BDT Score | Spelling (out of 30) |
|---|---|---|---|---|
| Dyslexic 10 | 28 | Student | Diagnosed elsewhere | |
| Dyslexic 11 | 20 | Student | 6+ves - Diagnosed elsewhere | |
| Dyslexic 12 | 21 | Student | 5.5+ves - Diagnosed elsewhere | |
| Dyslexic 13 | 19 | Student | Diagnosed elsewhere | |
| Dyslexic 14 | 19 | Student | Diagnosed elsewhere | |
| Dyslexic 15 | 26 | Student | Diagnosed elsewhere | |
| Dyslexic 16 | 21 | Civil Servant | Diagnosed elsewhere | |
| Dyslexic 17 | 23 | Volunteer | 7+ves | 5 |
| Control 10 | 32 | Student | 2.5+ves | 13 |
| Control 11 | 21 | Student | 2.5+ves | 20 |
| Control 12 | 22 | Student | 0+ves | 22 |
| Control 13 | 26 | Catering Worker | 2.5+ves | 25 |
| Control 14 | 23 | Student | 3.5+ves | 19 |
| Control 15 | 20 | Student | 3.5+ves | 24 |
| Control 16 | 21 | Student | 0+ves | 27 |
| Control 17 | 24 | Cook | 3.5+ves | 11 |

<u>**Key**</u>

**BDT = Bangor Dyslexia Test (Miles 1982)**

**Spelling = Schonell Spellign Test**

Appendix D

# Record Sheet

## Identity Matching

**Name:** ____________________

(1) A1 - A1 A2 A3
(2) C1 - C3 C2 C1
(3) C3 - C1 C2 C3
(4) B2 - B2 B3 B1
(5) A3 - A1 A3 A2
(6) B1 - B3 B2 B1
(7) C3 - C1 C3 C2
(8) A2 - A2 A3 A1
(9) C2 - C2 C3 C1
(10) B3 - B1 B2 B3
(11) B1 - B1 B2 B3
(12) A1 - A3 A2 A1
(13) A2 - A1 A2 A3
(14) C1 - C1 C2 C3
(15) B2 - B1 B3 B2
(16) B3 - B1 B3 B2
(17) A3 - A1 A2 A3
(18) C2 - C1 C2 C3

## Appendix E

## Post-Experimental Questionnaire Expt. 2

(1) It is often useful in experiments of this kind to discover what went through peoples minds when performing the task.
Can you tell me anything about what went through your mind on a typical trial?

(2) Were you aware of using any particular strategy when trying to work out which picture or symbol went with which?

(3) Did you have any names for any of the pictures or symbols? (Subjects were given a separate sheet with the stimuli on to remind them).

(4) Did you use mental images to help you solve the tasks?

(5) Did you use any kind of mental trick to help you to remember which pairs of pictures or symbols went together?

## Appendix F

## A transcript of the post-experimental interview administered in Expts. 2, 3, 4 and 5.

I am going to show you some examples of the stimulus/word pairs you were shown in the long test session at the end.

You were shown pairs which you'd seen before.

Show: A1 - B1 B2 B3
A2 - B1 B2 B3
A3 - B1 B2 B3

B1 - C1 C2 C3
B2 - C1 C2 C3
B3 - C1 C2 C3

How did you remember which stimuli/words went together? What went through your mind?

You were also shown the symmetrical/reverse versions of these pairs. What did you make of these? What went through your mind?

Show: B1 - A1 A2 A3
C1 -B1 B2 B3

Finally you were shown pairs of stimuli/words which you had never seen before.

Show: A1 - C1 C2 C3
C1 - A1 A2 A3

Did these mean anything to you?
Could you see any connection?
If yes, did you know that these words went together before this last session?

## Appendix G

| Subject Number | Age | Occupation | BDT Score | Spelling (out of 30) | Matrices (out of 36) |
|---|---|---|---|---|---|
| Control 18 | 26 | Student | 1.5+ves | 27 | 26 |
| Control 19 | 19 | Student | 3.5+ves | 22 | 17 |
| Control 20 | 33 | Student | 1.5+ves | 17 | 17 |
| Control 21 | 30 | Student | 0+ves | 22 | 25 |
| Control 22 | 19 | Student | 2.5+ves | 19 | 30 |
| Control 23 | 28 | Student | 0+ves | 19 | 32 |
| Control 24 | 20 | Student | 1+ve | 19 | 26 |
| Control 25 | 30 | Student | 0+ves | 30 | 29 |
| Control 26 | 45 | Student | 0.5+ves | 13 | 17 |
| Dyslexic 18 | 18 | Student | 6.5+ves | 10 | 22 |
| Dyslexic 19 | 44 | Student | 5.5+ves | 18 | 14 |
| Dyslexic 20 | 21 | Student | 4+ves | 19 | 24 |
| Dyslexic 21 | 22 | Student | 4.5+ves | 20 | 26 |
| Dyslexic 22 | 19 | Student | 5+ves | 4 | 29 |
| Dyslexic 23 | 55 | Student | 5+ves | 23 | 18 |
| Dyslexic 24 | 22 | Student | 4.5+ves | 16 | 24 |
| Dyslexic 25 | 18 | Student | 6+ves | 3 | 21 |
| Dyslexic 26 | 29 | Student | 3+ves | 9 | 23 |

**Key:** BDT=Bangor Dyslexia Test (Miles, 1982) Spelling=Schonell Spelling Test Matrices=Advance Progressive Matrices (Raven,1962).

## Appendix H

| Subject Number | Age | Occupation | BDT Score | Spelling (out of 30) | Matrices (out of 36) |
|---|---|---|---|---|---|
| Dyslexic 27 | 30 | Support Worker | 7+ves | 0 | 25 |
| Dyslexic 28 | 25 | Chef | 7+ves | 1 | 15 |
| Dyslexic 29 | 35 | Butcher | 4.5+ves | 0 | 14 |
| Dyslexic 30 | 27 | Student | 3.5+ves | 16 | 21 |
| Dyslexic 31 | 29 | Carpenter | 3.5+ves | 17 | 13 |
| Dyslexic 32 | 22 | Clerical Officer | 4.5+ves | 2 | 12 |
| Dyslexic 33 | 31 | Nurse | 8.5+ves | 13 | 24 |
| Dyslexic 34 | 18 | Care Assistant | 6+ves | 0 | 10 |
| Control 27 | 27 | Student | 0+ves | 15 | 25 |
| Control 28 | 22 | Unemployed | 2.5+ves | 16 | 13 |
| Control 29 | 27 | Town Planner | 0+ves | 25 | 27 |
| Control 30 | 26 | Student | 2+ves | 26 | 25 |
| Control 31 | 25 | Unemployed | 2+ves | 19 | 30 |
| Control 32 | 20 | Student | 2.5+ves | 30 | 23 |
| Control 33 | 30 | Unemployed | 1+ve | 27 | 30 |
| Control 34 | 18 | Student | 3.5+ves | 13 | 19 |

Key:
**BDT = Bangor Dyslexia Test (Miles 1982)**
**Spelling = Schonell Spelling Test**
**Matrices = Advanced Progressive Matrices (Raven 1962)**

## Appendix I

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| MIB | WEF | KEB | YOF | ZIQ | MOX | YIM | ZEG | KOJ |
| KEB | ZIQ | MOX | YOF | MIB | ZEG | MOX | KOJ | MIB |
| YIM | ZEG | WEF | KOJ | ZIQ | KEB | MIB | YOF | KEB |
| WEF | YIM | MIB | MOX | KEB | KOJ | YIM | ZEG | ZIQ |
| YOF | KOJ | YIM | ZEG | WEF | ZIQ | MOX | WEF | YOF |

## Appendix J

| Subject Number | Age | Occupation | BDT Score | Spelling (out of 30) | Ravens (out of 36) |
|---|---|---|---|---|---|
| Control 35 | 20 | Student | 0+ves | 24 | 19 |
| Control 36 | 24 | Student | 0+ves | 21 | 24 |
| Control 37 | 27 | Student | 1+ve | 25 | 24 |
| Control 38 | 29 | Student | 1+ve | 27 | 20 |
| Control 39 | 19 | Student | 1.5+ves | 28 | 19 |
| Control 40 | 22 | Student | 1+ve | 10 | 31 |
| Control 41 | 19 | Student | 0+ves | 27 | 25 |
| Control 42 | 20 | Student | 0+ves | 29 | 17 |
| Control 43 | 34 | Student | 1.5+ves | 22 | 18 |
| | | | | | |
| Dyslexic 35 | 19 | Student | 1.5+ves | 13 | 17 |
| Dyslexic 36 | 39 | Student | 4+ves | 25 | 22 |
| Dyslexic 37 | 53 | Student | 5+ves | 7 | 21 |
| Dyslexic 38 | 18 | Student | 8+ves | 1 | 25 |
| Dyslexic 39 | 39 | Student | 7.5+ves | 6 | 16 |
| Dyslexic 40 | 20 | Student | 5.5+ves | 10 | 18 |
| Dyslexic 41 | 32 | Student | 7.5+ves | 18 | 15 |
| Dysleixc 42 | 21 | Student | 5+ves | 11 | 27 |
| Dyslexic 43 | 23 | Student | 4.5+ves | 3 | 21 |

**Key BDT = Bangor Dyslexia Test (Miles, 1982) Spelling = Schonell Spelling Test Matrices = Advanced Progressive Matrices (Raven, 1962).**

# References


Ackerman, P.T. and Dykman, R.A. (1993). Phonological processes, confrontational naming, and immediate memory in dyslexia. *Journal of Learning Disabilities,* 26(19), 597-609.

Anderson, S.W., Podwall, F.N., and Jaffe, J. (1984). Timing analysis of coding and articulation processes in dyslexia. *Annals of the New York Academy of Science,* 433, 71-86.

Barnes, D., McCullagh, P.D., and Keenan, M. (1990). Equivalence class formation in non-hearing impaired children and hearing impaired children. *The Analysis of Verbal Behaviour,* 8, 19-30.

Bentall, R. P., and Dickins, D. W. (1994). Differences between animal and human learning: Implicit and explicit processes. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages.* Academic Press Ltd. London.

Bentall, R.P., Dickins, D.W. and Fox, S.R.A. (1993). Naming and equivalence: Response latencies for emergent relations. *The Quarterly Journal of Experimental Psychology,* 46B, 187-214.

Bentall, R.P., and Lowe, C.F. (1987). The role of verbal behaviour in human learning III: Instructional effects in children. *Journal of the Experimental Analysis of Behaviour,* 47, 177-190.

Bentall, R.P., Lowe, C.F., and Beasty, A. (1985). The role of verbal behaviour in human learning II; Developmental differences. *Journal of the Experimental Analysis of Behaviour,* 43, 165-181.

Berlin, R. (1887). *Eine besondere art der wordblinheit: Dyslexia [A special type of wordblindness: Dyslexia].* Wiesbaden: J. F. Bergmann.

Boelens, H. (1994). A traditional account of stimulus equivalence. *Psychological Record,* 44, 587-605.

Bouma, H., and Legein, C.P. (1980). Dyslexia: A specific recoding deficit? An analysis of response latencies for letters and words in dyslexic and in average readers. *Neuropsychologia,* 18, 285-298.

Bowers, P. G., and Swanson, L.B. (1991). Naming speed deficits in reading disability: Multiple measures of a singular process. *Journal of Experimental Child Psychology,* 51, 195-219.

Bradley, L., and Bryant, P.E. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. *Nature,* 271, 746-747.

Bradley, L., and Bryant, P.E. (1983). Categorising sounds and learning to read - a causal connection. *Nature,* 301, 419-421.

Brady, S., Shankweiler, D., and Mann, V. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology,* 35, 345-367.

Bruck, M. (1992). Persistence of dyslexic's phonological awareness deficits. *Developmental Psychology,* 28(5), 874-886.

Bush, K. M., Sidman, M., and DeRose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behaviour,* 51, 29-45.

Catania, A. C., and Harnard, S. (Eds.). (1988). *The selection of behaviour. The operant behaviourism of B. F. Skinner.* Cambridge: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences.* Second edition. Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin,* 112(1), 155-159.

Cohen, L. R., Looney, T. A., Brady, J. H., and Aucella, A. F. (1976). Differential sample response schedules in the acquisition of conditional discriminations by pigeons. *Journal of the Experimental Analysis of Behaviour,* 26, 301-314.

Craik, F. I. M. and Lockhart, R. S. (1973). Levels of Processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour,* 11, 671-684.

Critchley, M. (1981). Dyslexia: An overview. In G. Th. Pavlidis and T. R. Miles (Eds.), *Dyslexia Research and its applications to Education.* Chichester: Wiley.

Denckla, M. B. (1972). Color-naming deficits in dyslexic boys. *Cortex,* 8, 164-176.

Denckla, M. B., and Rudel, R. G. (1974). Rapid "automatised" naming of pictured objects, colors, letters, and numbers by normal children. *Cortex,* 10 , 186-202.

Denckla, M. B., and Rudel, R. G. (1976a). Rapid "automatised" naming (R.A.N.): Dyslexia differentiated from other learning disabilities. *Neuropsychologia,* 14, 471-479.

Denckla, M. B., and Rudel, R. G. (1976b). Naming of object-drawings by dyslexic and other learning disabled children. *Brain and Language,* 3, 1-15.

Devany, J. M., Hayes, S. C., and Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behaviour,* 46, 243-257.

Done, D. J., and Miles, T. R. (1978). Learning, memory and dyslexia. *Dyslexia Review,* 1, 13-14.

Dugdale, N. A. (1988). *The role of naming in stimulus equivalence. Differences between humans and animals.* Unpublished Ph.D. University of Wales.

Dugdale, N. A., and Lowe, C. F. (1990). Naming and stimulus equivalence. In D. E. Blackman and H. Lejeune (Eds.), *Behaviour analysis in theory and practice: Contributions and controversies* . Hove, England: Lawrence Erlbaum Associates.

Eikeseth, S., and Smith, T. (1992). The development of functional and equivalence classes in high-functioning autistic children: The role of naming. *Journal of the Experimental Analysis of Behaviour,* 58,123-133.

Elliott, C. D., Murray, D. J., and Pearson, L. S. (1983). *The British Ability Scales.* Windsor: NFER-Nelson.

Ellis, N. C. (1980). *Functional analysis of reading and short term memory in dyslexic children.* Unpublished Ph.D. University of Wales.

Ellis, N. C. (1981). Visual and name coding in dyslexic children. *Psychological Research,* 43, 201-218.

Ellis, N. C., and Miles, T. R. (1978). Visual information processing in dyslexic children. In M. M. Grunber, P. E. Morris, and R. N. Sykes (Eds.), *Practical aspects of memory.* London: Academic Press.

Ellis, N. C., and Miles, T. R. (1981). A Lexical Encoding Deficiency I and II. In G. Th. Pavlidis, and T. R. Miles (Eds), *Dyslexia research and its application to education.* Chichester: Wiley.

Fawcett, A., and Nicolson, R. (Eds.). (1994a). *Dyslexia in children: Multidisciplinary perspectives.* Hemel Hempstead: Harvester Wheatsheaf.

Fawcett, A. and Nicolson, R. (1994b). Naming speed in children with dyslexia. *Journal of Learning Disabilities,* 27(10), 641-646.

Felton, R. H., Naylor, C. E., and Wood, F. B. (1990). Neuropsychological profile of adult dyslexics. *Brain and Language,* 39, 485-497.

Fields, L., Adams, B. J., Verhave, T., and Newmand, S. (1990). The effects of nodality on the formation of equivalence classes. *Journal of Experimental Analysis of Behaviour,* 53, 345-358.

Fields, L., and Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behaviour,* 48, 317-332.

Fields, L., Verhave, T., and Fath. (1984). Stimulus equivalence and transitive associations: A methodological analysis. *Journal of the Experimental Analysis of Behaviour,* 42, 143-157.

Galaburda, A. M. (1994). Developmental dyslexia and animal studies: At the interface between cognition and neurology. *Cognition,* 50, 133-149.

Hartley, D. (1749). *Observations on man: His duty and his expectations.* Gainsville Fla.: Scholars' Facsimiles Reprints 1966.

Harzem, P., Lowe, C. F., and Bagshaw, M. (1978). Verbal control in human operant behaviour. *Psychological Record,* 28, 405-423.

Hayes, S.C. (1991). A relational control theory of stimulus equivalence. In L. J. Hayes, and P. N. Chase (Eds), *Dialogues on verbal behaviour: The first international institute on verbal relations.* Reno, Nevada: Context Press.

Hayes, S. C. (1992). Verbal relations, time, and suicide. In S. C. Hayes, and L. J. Hayes (Eds), *Understanding verbal relations. The second and third international institutes on verbal relations.* Reno, Nevada: Context Press.

Hayes, S. C., and Hayes, L. J. (1992). Verbal relations and the evolution of behaviour analysis. *American Psychologist,* 47(11), 1383-1395.

Head, H. (1926). *Aphasia and kindred disorders of speech.* London: Macmillan.

Hinshelwood, J. (1917). *Congenital Word-blindness.* London: H. K. Lewis.

Horne, P. J., and Lowe, C. F. (1996). On the origins of naming and other symbolic behaviour. *Journal of the Experimental Analysis of Behaviour,* 65, 185-241.

Howell, D. C. (1985). *Fundamental statistics for the behavioural sciences.* PWS - Kent Publishing Company: Boston, Massachusett.

Hume, D. C. (1911). *A Treatise of Human Nature.* London: J. M. Dent and Sons Ltd.

Jenkins, J. J. (1963). Mediated associations: Paradigms and situations. In C. N. Cofer, and B. S. Musgrave (Eds.), *Verbal behaviour and learning: Problems and processes.* New York: McGraw-Hill.

Jenkins, J. J., and Palermo, D. S. (1964). Mediated processes and the acquisition of linguistic structure. In U. Bellugi, and R. Brown (Eds.), *The acquisition of language. Monographs of the Society for Research on Child Development,* 29, (1, Serial No. 92), 141-169.

Kamhi, A. G. (1992). Response to historical perspective: A developmental language perspective. *Journal of learning disabilities,* 25(1), 48-52.

Katz, R. B. (1986). Phonological deficiencies in children with reading disabilities: Evidence from an object naming task. *Cognition,* 22, 225-257.

Katz, W. F., Curtiss, S., and Tallal, P. (1992). Rapid automatised naming and gesture by normal and language-impaired children. *Brain and Language,* 43, 623-641.

Kinsbourne, M. (1990). Cognitive deficits concomitant with adult dyslexia. In G. Th. Pavlidis (Ed.), *Perspectives on Dyslexia,* Vol. 2. Chichester: Wiley.

Kussmaul, A. (1978). Word-deafness and word-blindness. In H. von Ziemssen (Ed.), *Cyclopaedia of the practice of medicine: Vol. 14 (Diseases of the nervous system and disturbances of speech).* London, Sampston Row: Marston, Searle and Rivington.

LaBurda, M. C., and Defries, J. C. (1990). Genetic etiology of reading disability evidence from a twin study. In G. Th. Pavlidis (Ed.), *Perspectives on Dyslexia,* Vol. 1. Chichester: Wiley.

Lazar, R. M., Davis-Lang, D., and Sanchez, L. (1984). The formation of visual stimulus equivalence in children. *Journal of the Experimental Analysis of Behaviour,* 41, 251-266.

Leahey, T. H. (1992). *A history of psychology: Main currents in psychological thought.* New Jersey: Prentice Hall, Englewood Cliffs.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levine, D. N., Calvanio, R., and Popovics, A. (1982). Language in the absence of inner speech. *Psychologia,* 20(4), 391-409.

Livingstone, M. S., Rosen, G. D., Drislane, F. W., and Galaburda, A. M. (1991). Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia. *Neurobiology,* 88, 7943-7947.

Locke, J. (1877). *An essay concerning human understanding.* London: William Tegg and Co.

Lowe, C. F., and Beasty, A. (1987, April). *Language and the emergence of equivalence relations: a developmental study.* Paper presented at the British Psychological Society Annual Conference, Univ. of Sussex.

Lowe, C. F., Beasty, A., and Bentall, R. P. (1983). The role of verbal behaviour in human learning I: Infant performance on fixed-interval schedules. *Journal of the Experimental Analysis of Behaviour,* 39, 157-164.

Lowe, C. F., and Horne, P. J. (1996). Reflections on naming and other symbolic behaviour. *Journal of the Experimental Analysis of Behaviour,* 65, 315-340.

Manabe, K., Kawashima, T., and Staddon, J. E. R. (1995). Differential vocalisation in budgerigars: Towards an experimental analysis of naming. *Journal of the Experimental Analysis of Behaviour,* 63, 111-126.

Mandell, C., and Sheen, V. (1994). Equivalence class formation as a function of the pronounceability of the sample stimulus. *Behavioural Processes,* 32, 29-46.

Michael, J. (1985). Two kinds of verbal behaviour plus a possible third. *The Analysis of Verbal Behaviour,* 3, 1-4.

Miles, T. R. (1982). *The Bangor Dyslexia Test.* Cambridge: Learning Development Aids.

Miles, T. R. (1993). *Dyslexia: The pattern of difficulties.* Second Edition. London: Whurr Publishers.

Miles, T. R. (1994a). A proposed taxonomy and some consequences. In A. J. Fawcett, and R. I. Nicolson (Eds), *Dyslexia in children: multidisciplinary perspectives,* Hemel Hempstead: Harvester Wheatsheaf.

Miles, T. R. (1994b). Towards a rationale for diagnosis. In G. Hales (Ed.), *Dyslexia matters.* London: Whurr.

Miles, T. R. (1995). Dyslexia: The current status of the term: II. *Child Language Teaching and Therapy,* 11(1), 23-33.

Miles, T. R., and Miles, E. (1990). *Dyslexia: A hundred years on.* Milton Keynes: Open University Press.

Mill, J. (1969). *James Mill on Education.* London, Cambridge: University Press.

Miller, G. A. (1962). *Psychology. The science of mental life.* Middlesex, England: Penguin.

Morgan, W. P. (1896). A case of congenital word blindness. *The British Medical Journal,* 2, 1378.

Murphy, L. A., Pollatsek, A., and Well, A. D. (1988). Developmental dyslexia and word retrieval deficits. *Brain and Language,* 35, 1-23.

Nebes, R. D. (1975). The nature of internal speech in a patient with aphasia. *Brain and Language,* 2, 489-497.

Neel, A. (1971). *Theories of psychology.* University of London Press.

Nicolson, R. I., and Fawcett, A. J. (1990). Automaticity: A new framework for dyslexia research. *Cognition,* 35, 159-182.

Nicolson, R. I., and Fawcett, A. J. (1994). Comparison of deficit severity across skills: Towards a taxonomy for dyslexia. In A. J. Fawcett, and R. I. Nicolson (Eds.), *Dyslexia in children: multidisciplinary perspectives.* Hemel Hempstead: Harvester Wheatsheaf.

Nicolson, R. I., Fawcett, A., and Dean, P. (1995). Time estimation deficits in developmental dyslexia: evidence of cerebellar involvement. *Proceedings of the Royal Society,* London, B, 43-47.

Orton, S. T. (1937). *Reading, Writing and Speech Problems in Children.* New York: W. W. Norton.

Pavlidis, G. Th. (1990). Conceptualisation, symptomatology, and diagnostic criteria for dyslexia. In G. Th. Pavlidis (Ed.), *Perspectives on Dyslexia.* Vol. 2. Chichester: Wiley.

Pennington, B. (Ed.). (1991). *Reading Disabilities: Genetic and Neurological influences.* Dordrecht: Kluwer.

Perfetti, C. A. (1985). *Reading Ability.* New York: Oxford University Press.

Perfetti, C. A., Finger, E., and Hogaboam, T. (1978). Sources of vocalisation latency differences between skilled and less skilled young readers. *Journal of Educational Psychology,* 70, 730-739.

Rachlin, H. (1991). *Introduction to modern behaviourism.* New York: W. H. Freeman and Company.

Rack, J. P. (1994). Dyslexia: The phonological deficit hypothesis. In A. J. Fawcett, and R. I. Nicolson (Eds.), *Dyslexia in children: Multidisciplinary perspectives.* Hemel Hempstead: Harvester Wheatsheaf.

Raven, J. C. (1962). *Advanced Progressive Matrices.* London: H. K. Lewis and Co. Ltd.

Richardson, S. O. (1992). Historical perspectives on dyslexia. *Journal of Learning Disabilities,* 25(1), 40-47.

Saunders, R. R., and Green, G. (1992). The nonequivalence of behavioural and mathematical equivalence. *Journal of Experimental Analysis of Behaviour,* 57, 227-241.

Saunders, K. J., Saunders, R. R., Williams, D. C., and Spradlin, J. E. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record,* 43, 725-744.

Schonell, F. J., and Schonell, F. E. (1952). *Diagnostic and Attainment Testing.* Edinburgh: Oliver and Boyd.

Schusterman, R. J., and Kastak, D. (1993). A california sea-lion (zalophus californiaus) is capable of forming equivalence relations. *The Psychological Record,* 43, 823-839.

Schwartz, B. (1989). *Psychology of learning and behaviour.* Third Edition. New York: W. W. Norton and Company.

Siegel, S., and Castellan, N. J. Jr. (1988). *Nonparametric statistics for the behavioural sciences.* Second Edition. New York: McGraw Hill.

Skinner, B. F. (1957). *Verbal Behaviour.* New York: Appleton-Century-Crofts.

Sidman, M. (1971). Reading and auditory-visual equivalence. *Journal of Speech and Hearing Research,* 14, 5-13.

Sidman, M. (1987). Two choices are not enough. *Behaviour Analysis,* 22, 11-18.

Sidman, M. (1994). *Equivalence Relations and behaviour: A research story.* Boston, MA: Authors Cooperative Inc.

Sidman, M., and Cresson, O. (1973). Reading and crossmodal transfer of stimulus equivalence in severe retardation. *American Journal of Mental Deficiency,* 77, 515-523.

Sidman, M., Cresson, O., and Willson-Morris, M. (1974). Acquisition of matching-to-sample via mediated transfer. *Journal of Experimental Analysis of Behaviour,* 22, 261-273.

Sidman, M., Kirk, B., and Willson-Morris, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behaviour,* 43, 21-42.

Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., and Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkey, baboons and children. *Journal of the Experimental Analysis of Behaviour,* 37, 23-44.

Sidman, M., and Tailby, W. (1982). Conditional discrimination vs. matching-to-sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behaviour,* 37, 5-22.

Sidman, M., Willson-Morris, M., and Kirk, B. (1986). Matching-to-sample procedures and the developmental of equivalence relations: The role of naming. *Analysis and Intervention in Developmental Disabilities,* 6, 1-19.

Snowling, M., Wagendonk, B., and Stafford, C. (1988). Object-naming deficits in developmental dyslexia. *Journal of Research in Reading,* 11, 67-85.

Spring, C., and Capps, C. (1974). Encoding speed, rehearsal and probed recall of dyslexic boys. *Journal of Educational Psychology,* 66(5), 780-786.

Stanovich, K. E., Freeman, D. J., and Cunningham, A. E. (1983). The development of the relation between letter-naming speed and reading ability. *Bulletin of the Psychonomic Society,* 21(3), 199-202.

Steele, D., and Hayes, s. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behaviour,* 56, 519-555.

Tallal, P., Miller, S., and Fitch, R. H. (1993). Neurobiological basis of speech: A case for the preeminence of temporal processing. *Annals of the New York Academy of Sciences,* 1993.

Thomson, M. (1990). *Developmental Dyslexia.* London: Whurr Publisher Ltd.

Underwood, B. J., and Schulz, R. W. (1960). *Meaningfulness and verbal learning.* Chicago: Lippincott.

Urcuioli, P. J. (1985). On the role of differential sample behaviours in matching-to-sample. *Journal of Experimental Psychology: Animal Behaviour,* 26, 301-314.

Urcuioli, P. J., Zentall, T. R., and DeMarse, T. (1995). Transfer to derived sample-comparison relations by pigeons following many-to-one versus one-to-many matching with identical training relations. *The Quarterly Journal of Experimental Psychology,* 48B(2), 158-178.

Vellutino, F. (1987). Dyslexia. *Scientific American,* 256(3), 20-27.

Wagner, R. K., and Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin,* 101(2), 192-212.

Watson, F. L., and Brown, G. D. A. (1992). Single-word reading in college dyslexics. *Applied Cognitive Psychology,* 6, 263-272.

Wechsler, D. (1976). *Wechsler intelligence scale for children (WISC-R).* New York: Psychological Corporation.

Wolf, M. (1984). Naming, reading and the dyslexias: A longitudinal overview. *Annals of Dyslexia,* 34, 87-115.

Wolf, M. (1986). Rapid alternating stimulus naming in the developmental dyslexias. *Brain and Language,* 27, 360-379.

Wolf, M., Bally, H., and Morris, R. (1986). Automaticity, retrieval processes, and reading: A longitudinal study in average and impaired readers. *Child Development,* 57, 988-1000.

Wolf, M., and Goodglass, H. (1986). Dyslexia, dysnomia and lexical retrieval: A longitudinal investigation. *Brain and Language,* 28,154-168.

Wolf, M., and Obregon, M. (1992). Early naming deficits, developmental dyslexia and a specific deficit hypothesis. *Brain and Language,* 42, 219-247.

Wolff, P. H., Michel, G. F., and Ovrut, M. (1990). Rate variables and automatized naming in developmental dyslexia. *Brain and Language,* 39, 556-575.

Wulfert, E., Dougher, M. J., and Greenway, D. E. (1991). Protocol analysis of the correspondence of verbal behaviour and equivalence class formation. *Journal of the Experimental Analysis of Behaviour,* 56, 489-504.

Wulfert, E., and Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behaviour,* 50,125-144.

Yamamoto, J., and Asano, T. (1995). Stimulus equivalence in a chimpanzee (Pan Troglodytes). *The Psychological Record,* 45, 3-21.

Zentall, T. R., and Urcuioli, P. J. (1993). Emergent relations in the formation of stimulus classes by pigeons. *The Psychological Record,* 43, 795-810.