DOCTOR OF PHILOSOPHY

The role of naming in stimulus equivalence : differences between humans and animals.

Dugdale, Neil A.

Award date:
1988

Awarding institution:
Bangor University

UNIVERSITY OF WALES

THE ROLE OF NAMING IN STIMULUS EQUIVALENCE: DIFFERENCES BETWEEN HUMANS AND ANIMALS

NEIL A. DUGDALE

PH.D.

1988

ACKNOWLEDGEMENTS

It seems to me rather ironical that this page should come immediately after the 'declaration of independence' which I am required to sign for strictly administrative reasons. The production you see before you, to coin a theatrical analogy, has a very large supporting cast, without whom it could never have 'gone public'. Professor C.F. Lowe served ably as thesis Director, by providing constant and friendly support and guidance, and by showing an uncanny knack for inspiring me at just the right time. Dr. P.J. Horne, Assistant Director, provided a constant supply of extraordinarily incisive comments which helped to trim away many of the superfluous constructions to which I am prone. Professors Duane Rumbaugh and Sue Savage-Rumbaugh were my Directors 'on location'. I am indebted to them for all they did to make Experiment 1 a reality. The parents and staff of the Our Lady Roman Catholic School, especially Sister Alphonsus and Sister Margaret, are thanked for their roles in supplying the 'cast' for Experiments 3 - 6. Their children are a credit to them. Of course, no production could go ahead without its Technical Staff and none come better than Gaerwyn, David and Alan. Thanks are also due to the rest of the staff and postgraduates at the Psychology Department for sitting through many a 'rehearsal'.

But, above all, there is my wife, Sally. She not only typed this thesis with lightning speed, but she also maintained my sanity throughout; I am more than aware of the sacrifices she made throughout this production. Without her support and encouragement I very much doubt whether this thesis would have taken to the stage at all.

SUMMARY

When subjects learn to match a sample stimulus to a non-identical comparison stimulus, the stimuli may become equivalent, or substitutable for each other. Matching-to-sample procedures have generated stimulus equivalence with humans aged 3 years and upwards. Animals, however, have thus far failed tests of symmetry, one of the defining properties of equivalence. This human-animal difference suggests that language may be related to equivalence formation. In developmental studies by Beasty (1987), young children who failed equivalence tests later passed when taught to name the sample-comparison pairings during baseline matching trials. Naming, then, appears to be necessary for stimulus equivalence. Experiments in the present thesis further investigated equivalence formation in children and animals.

The first two experiments yielded further evidence against equivalence in animals. Experiment 1 found no evidence of equivalence in the arbitrary matching performances of two chimpanzees involved in an ape-language training programme. In Experiment 2, pigeons failed symmetry tests despite receiving extensive symmetry exemplar training.

The final series of studies examined naming and equivalence in 30 normal 4-5 year old children. In Experiment 3, children often gave the same name spontaneously to non-identical stimuli before matching them in equivalence tests. Experiments 4(a) - 6 systematically investigated common naming and showed it to be an extremely simple but effective way for naming to mediate equivalence. As well as suggesting a functional definition of naming, the results indicated that the subjects' pre-existing stimulus names may selectively interfere with equivalence formation by affecting the common naming relations introduced during the experiment.

These results support the view that language is a major determinant of human behaviour (Lowe, 1979; 1983) and they also emphasise the need for a functional analysis of language development.

CONTENTS

CHAPTER 1

INTRODUCTION

1. Background.
2. The stimulus equivalence paradigm.
3. The importance of stimulus equivalence to a behavioural analysis.
4. The role of naming in the formation of equivalence.
5. Where does equivalence come from?

BACKGROUND

Fifty years have elapsed since Skinner published the 'Behavior of Organisms' (Skinner, 1938), in which he urged psychology to embrace a science of behaviour. Fifty years seems an awfully long time to the present author and, given the pervasiveness of Skinner's arguments, one might have predicted that the 'experimental analysis of behaviour' would have grown by now into a dominant approach in psychology. Indeed, the approach got off to a flying start, thanks mainly to the experimental methods of operant conditioning. Basic research with animals yielded powerful techniques for altering behaviour and led to the derivation of presumably 'fundamental' conditioning principles. Behaviourism had something which rival approaches, particularly cognitive psychology, seemed to lack. Sidman puts the case against cognitivism most vociferously:

> "There is no body of systematized principles, no unique set of data, no characteristic measurement techniques, and no typical investigative procedures to which a cognitivist can point and say, 'That is my Science'....The basic units of cognition - - representations, intentions, plans, rules, programs, and mental structures - - are linked to actual behavior only if that becomes necessary. When such necessity does arise - - for example, in carrying out experiments - - the logic of the linkage need not be compelling. For the cognitivist, behavior is important only as a product of mental processes, but criteria do not exist for determining whether different instances of behavior represent the same mental processes. Given an interest in some particular process, each observer is privileged to decide which behavior will provide the appropriate window into the mind" (Sidman, 1986, p.214).

Perhaps this was why conditioning principles derived from animal research were eagerly and uncritically applied to problems of human behaviour. Initially, behaviour modification procedures attracted widespread attention both in clinical and educational psychology. The future for behaviourism looked bright.

Nowadays, however, behaviourism can hardly be called a dominant force in psychology. Paraphrasing Branch and Malagodi (1980), the spark of commitment to behaviourism, which previously glowed so brightly, is barely visible these days. Lowe (1983) goes further:

> "The power to predict and control complex human behavior, which behaviorism promised, proved to be decidedly elusive. Recently, behaviorists themselves have soul-searched (cf. Brigham, 1980; Cullen, 1981; Michael, 1980; Branch and Malagodi, 1980; Repucci and Saunders, 1974), outsiders have been eager to announce behaviorism's demise (Mackenzie, 1977), and out of the disillusionment the hydra-headed monster of mentalism .... has resurfaced in the form of contemporary cognitivism" (Lowe, 1983, p.73).

The behaviouristic soul-searching is particularly prevalent in the 1980 and 1981 volumes of the 'Behavior Analyst' journal, which include numerous contributions concerned with the 'flight from behaviour analysis' (e.g. Cullen, 1981; Michael, 1980). As the title of one paper asks, 'Where have all the behaviorists gone?' (Branch and Malagodi, 1980).

Lowe (1983) maintains that behaviourism has lost its way precisely because of its almost exclusive reliance on an animal model of human behaviour. Operant research with human subjects has been virtually neglected; the vast majority of basic operant research uses animals as subjects (Buskist and Miller, 1982), often on the assumption that animal behaviour and human behaviour have similar determinants and are governed by the same general principles. But this continuity assumption has not been substantiated by basic research (Lowe, 1983), particularly the recent finding that human operant behaviour can differ qualitatively from that of animals (Bentall, Lowe and Beasty, 1985; Lowe, 1979). Given the evidence for qualitative human-animal differences (summarised later in this chapter), it is hardly surprising that behaviourism has dwindled in applied settings. As Lowe puts it: 'If the animal model does not hold good for human operant behaviour under controlled experimental conditions why should it do so in the hospital, school or stock exchange?' (p.73). Furthermore, an almost slavish preoccupation with animal behaviour may explain why there seems to have been relatively little outside interest in basic behaviour-analytical research. Sidman (1986) sums up the problem:

> "An easy criticism has been that Behavior Analysis deals well with uninteresting behavior but ignores everything that makes human beings superior to all other creatures. The concepts of stimulus and response have seemed impoverished, unable to capture the rich complexity of the human intellect" (Sidman, 1986, p.215).

However, Sidman is well aware that not all behavior analysis can be so easily criticised; he himself has pioneered an area of operant research which has generated enormous interest in recent years. The study of stimulus equivalence is of obvious relevance to human activity because it opens the door to an experimental analysis of symbolic behaviour, and, in so doing, it may help to give behaviourism a new lease of life.

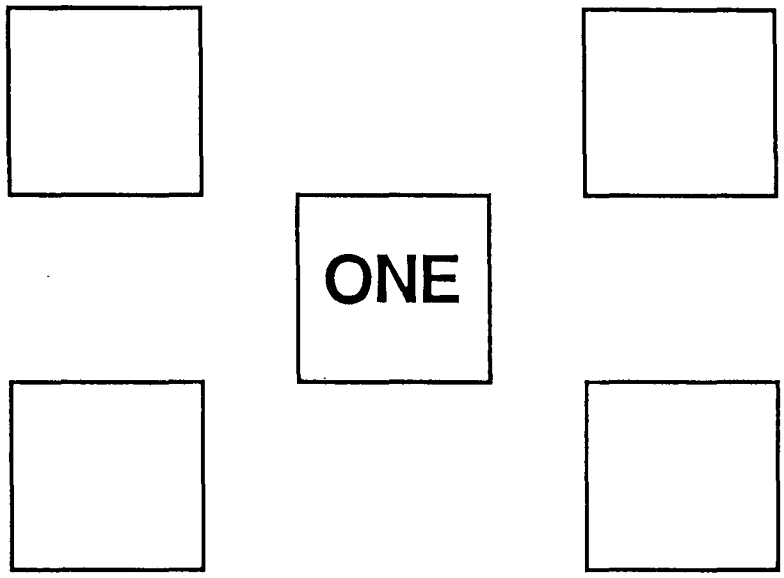
THE STIMULUS EQUIVALENCE PARADIGM

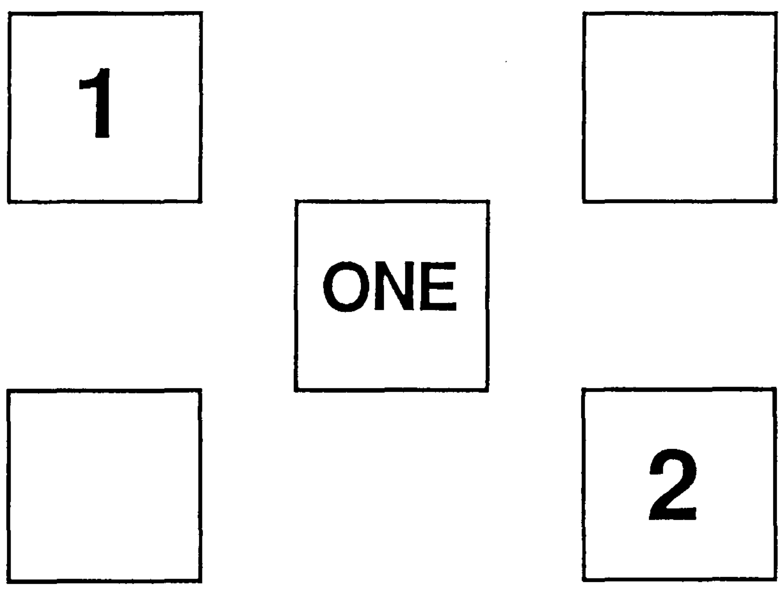
Matching-to-sample procedures are commonly employed in the study of stimulus equivalence. Figure 1.1 depicts one example of a matching-to-sample (MTS) task. A subject sits in front of a five-key response panel. The trial begins with the presentation of a stimulus on the centre key; this stimulus is the sample. The sample in this case is the printed word **ONE**. The subject then touches the sample, and additional stimuli appear on two of the four outer windows. These stimuli are the printed digits **1** and **2**, and these are the comparisons. The subject has to touch the comparison digit which corresponds to the printed number word sample. If the subject chooses the correct digit, a reinforcer is delivered. If the incorrect digit is chosen then no reinforcer is delivered. In either event the display goes blank, and a few seconds later another sample appears.

When the subject has learned to match each of a set of digits to the appropriate printed number word we may perhaps suspect that his behaviour is symbolic, that he is reading the words with comprehension. But we cannot be sure. Pigeons have learned MTS tasks that are just as arbitrary, but we might be less inclined to call their behaviour symbolic. How can we tell if the subject's performance on this or any other arbitrary matching task is symbolic, or whether it represents a simple stimulus-response chain, or conditional relation, which has no symbolic relevance?

If the stimuli in Figure 1.1 were acting as symbols then one would expect each to stand for the other. They would, in a sense, be equivalent; equivalent stimuli are by definition substitutable or interchangeable for each other (Sidman and Tailby, 1982). Stimulus equivalence appears to be a pre-requisite of symbolic activity (cf Catania, 1984; Devany, Hayes and Nelson, 1986; Sidman, 1977).

Figure 1.1 Schematic representation of an arbitrary matching-to-sample task. At the start of the trial a sample appears in the centre window of the five key response panel (see upper section). Touching the sample brings on the comparisons on any two of the four outer keys (see lower section). Reinforcers are delivered for selecting the comparison that corresponds to the sample.


ONE

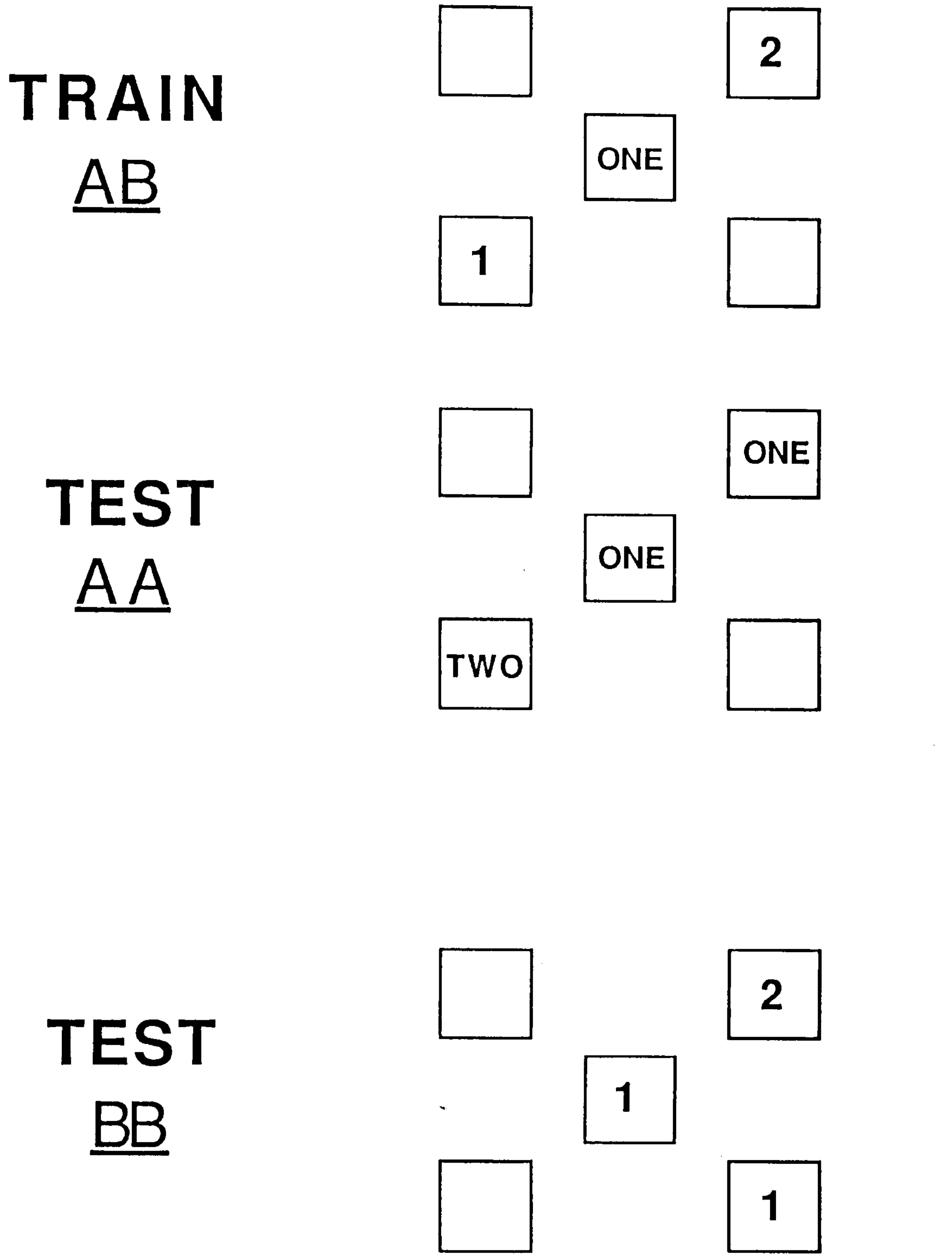

1
ONE
2

Equivalence is itself defined by the three properties of reflexivity, symmetry and transitivity (Sidman and Tailby, 1982). Each property may be tested independently. Figure 1.2 depicts a reflexivity test. The subject was initially trained to select Set-B digits conditional upon Set-A printed word samples. If the Set-A and Set-B stimuli are equivalent, then the subject should be able to match each stimulus to itself without additional training. In other words, generalised identity matching is the behavioural proof of reflexivity; if A=B, then A=A and B=B.

Reflexivity is not as trivial as it might appear. The subject may have learned AB arbitrary matching but this does not automatically guarantee AA and BB identity matching. During AB training, the Set-A stimuli always appear on the centre key as samples, and the Set-B stimuli always appear on the outer keys as comparisons. These invariant locations may become defining characteristics of the samples and comparisons (see Iversen, Sidman and Carrigan, 1986). If so, then, for example, the sample **ONE** and the comparison **ONE** would be as different from each other as, say, the sample **ONE** and the comparison **TWO**. Success on a reflexivity test shows that, as far as the subject is concerned, the stimuli remain identical when they change location from samples to comparisons or from comparisons to samples. Testing for reflexivity provides an empirical basis for the concept of identity, which is itself a pre-requisite for equivalence (Sidman 1986).

Symmetry, the second defining property of equivalence, is tested by interchanging the former samples and comparisons (see Figure 1.3). After training AB, the subject is tested on BA; digits now appear as samples and printed number words as comparisons. If the subject's AB matching exemplified equivalence then he should be capable of

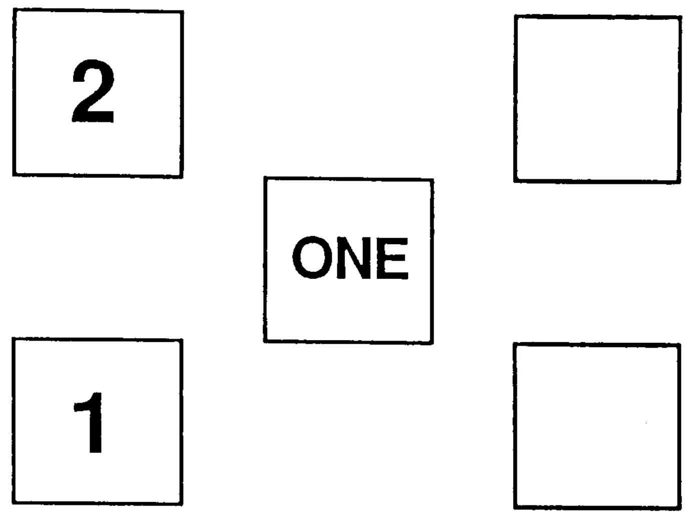
Figure 1.2 Schematic representation of a reflexivity **test** (see text).



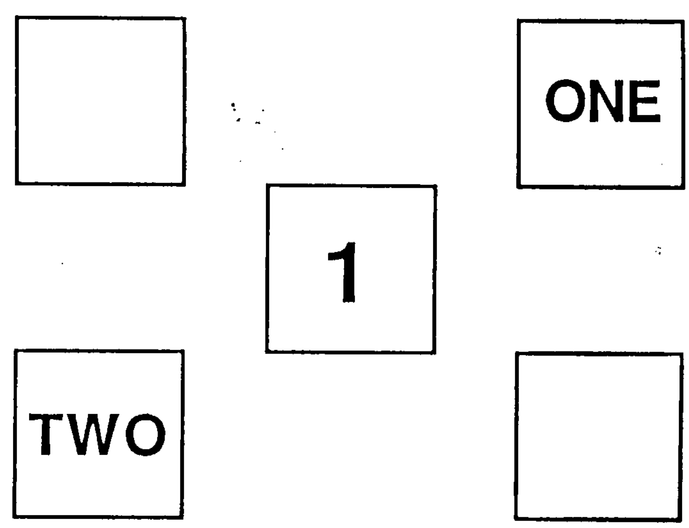
REFLEXIVITY TRAIN AB TEST AA & BB

Figure 1.3 Schematic representation of a symmetry test (see text).

TRAIN
<u>AB</u>



TEST
<u>BA</u>



	TRAIN	TEST
SYMMETRY	AB	BA

responding appropriately to the novel BA combinations without additional training (i.e. select the comparison ONE when the sample was 1, select TWO when the sample was 2, etc.). Symmetry translates behaviourally into the reversibility of sample and comparison roles; if A=B, then B=A.

The final requirement, transitivity, can be tested only after the subject has learned a second arbitrary relation, BC. An example is given in Figure 1.4. The subject has learned to match printed number words to digits (AB) and digits to numerical quantities (BC). Transitivity is demonstrated if the subject can then match printed words to quantities (AC) without additional training. If the training establishes equivalence between corresponding A, B and C stimuli then the subject should be capable of passing the transitivity test; if A=B and B=C, then A=C.

A subject's failure on any one of these tests would suggest that the stimuli had not become equivalent. Instead, the subject may have learned mere conditional or 'if-then' relations (e.g. if A then B, if B then C) which are fixed in sequence and of no symbolic relevance.

The three defining properties of equivalence can be evaluated simultaneously. Figure 1.5 depicts one such combined test. The subject is taught AB and AC and is then tested on BC and CB. In order to respond correctly on BC without additional training, symmetry of the trained AB relation is initially required, so that AB produces BA via symmetry. Then, given the derived relation BA and the trained relation AC, transitivity may yield BC (BA and AC, therefore BC). The CB relation may also emerge in similar fashion (CA and AB, therefore CB).

What is perhaps less obvious is that reflexivity is also required for BC and CB to emerge. Neither relation will emerge unless the subject views each B and C stimulus as identical across training and test conditions. During training, the Set-B stimuli only appear as

Figure 1.4 Schematic representation of a transitivity test (see text).

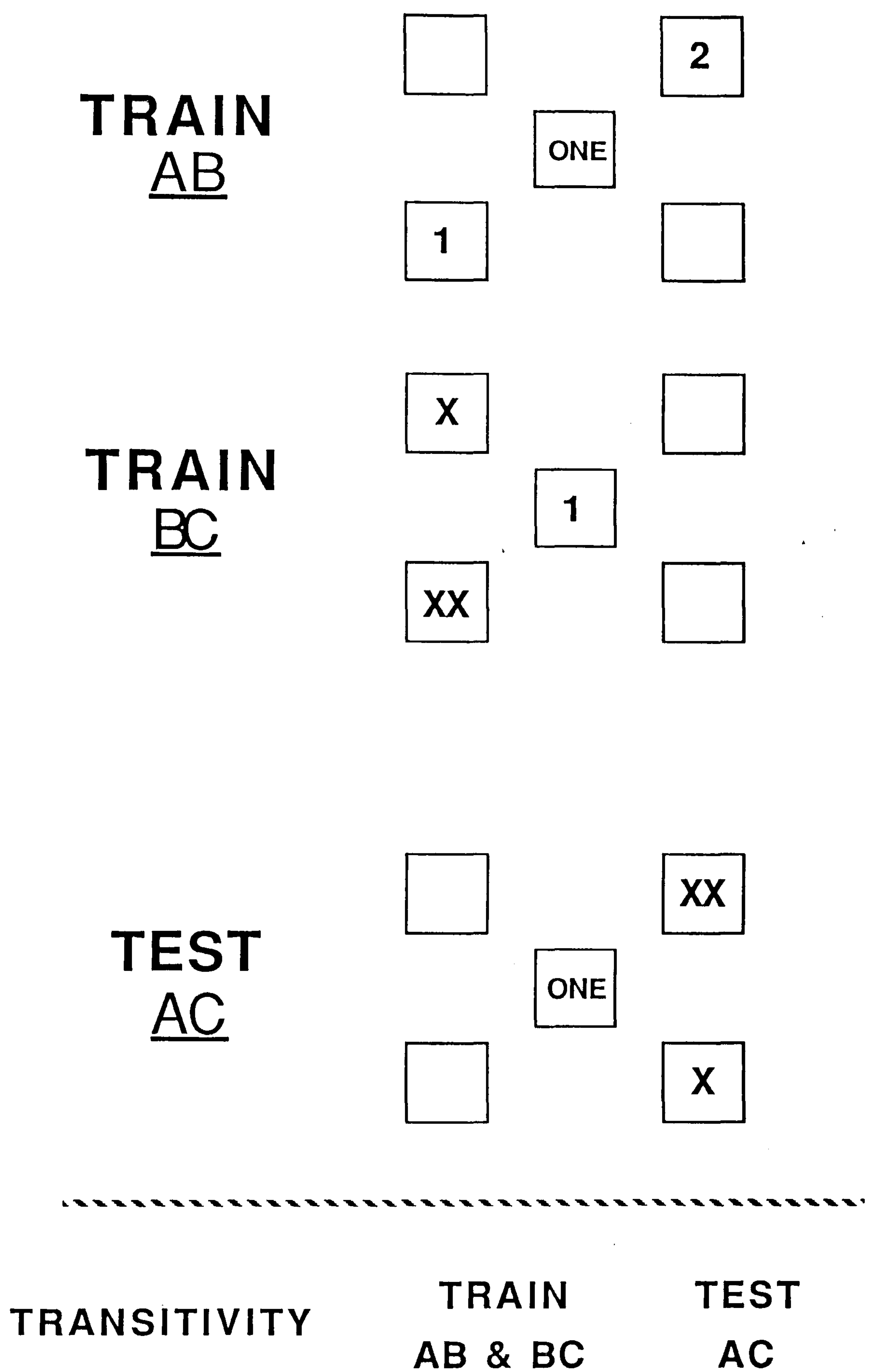
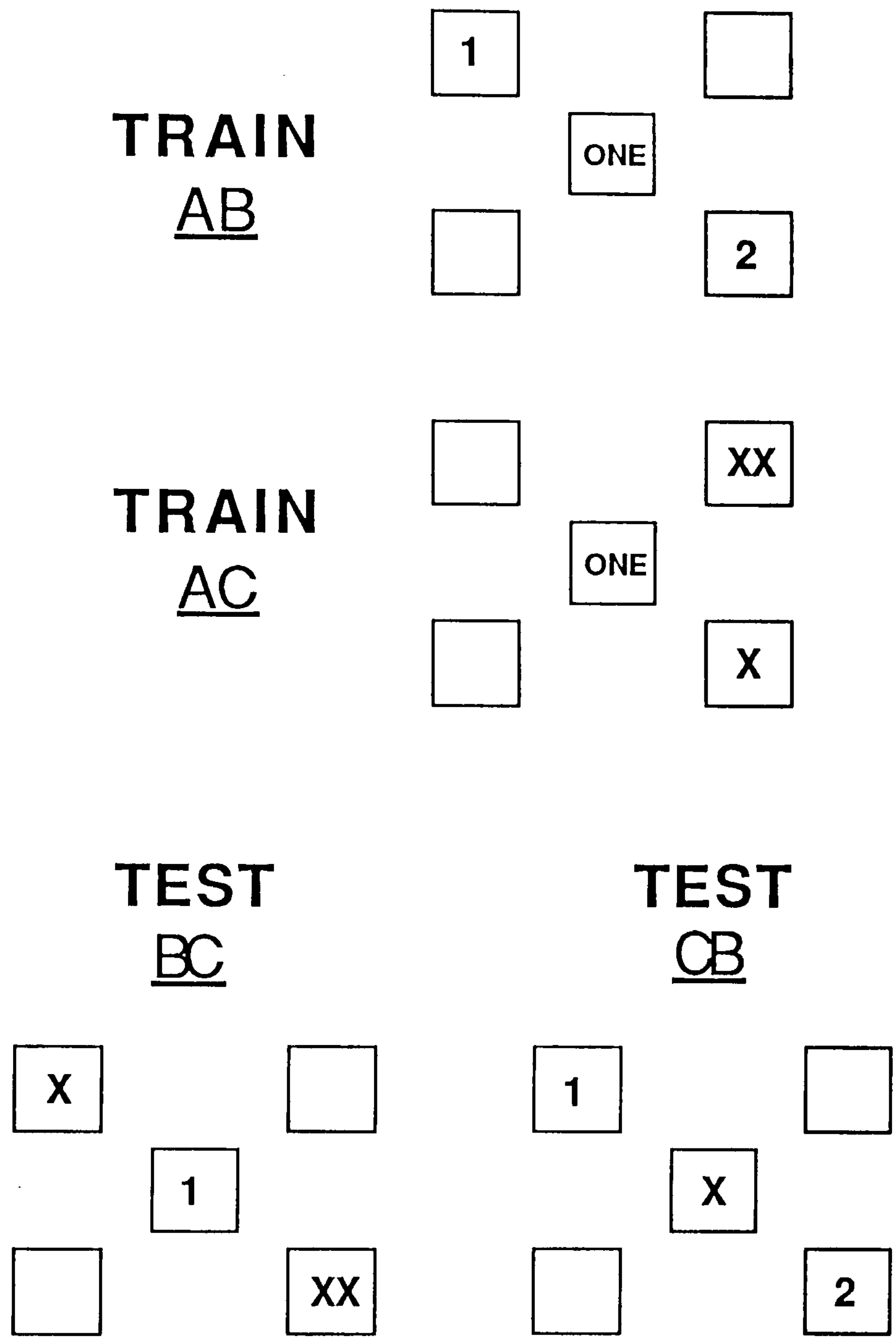

TRAIN
AB
2
ONE
1
X
TRAIN
BC
1
XX
XX
TEST
AC
ONE
X
TRANSITIVITY
TRAIN
AB & BC
TEST
AC

Figure 1.5 Schematic representation of a combined test for equivalence.



EQUIVALENCE	TRAIN A-B & A-C	TEST B-C & C-B

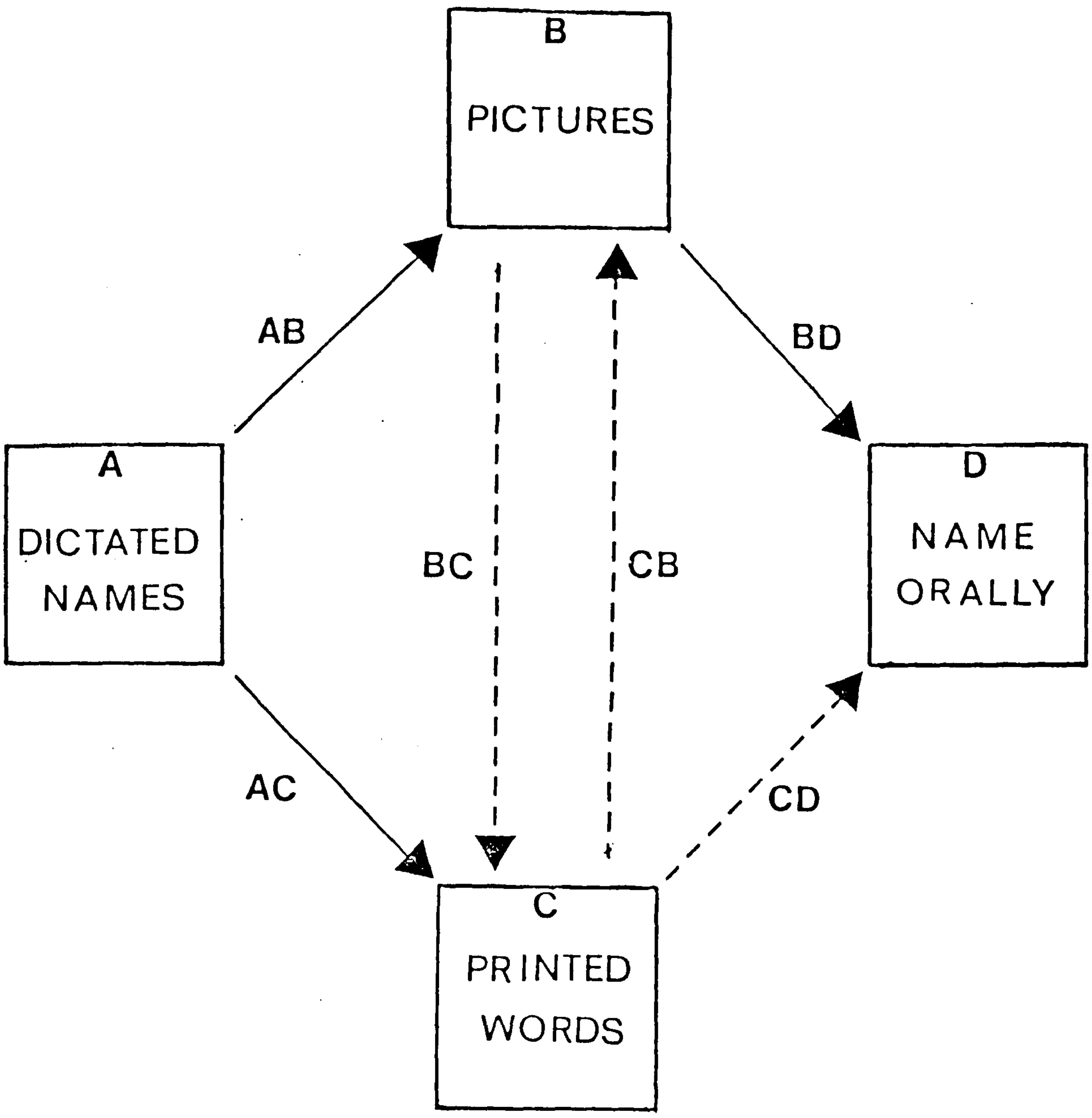
comparisons but during BC tests they appear, for the very first time, as samples. BC could not possibly emerge unless each B stimulus remains identical across this transposition. The same rationale applies to the Set-C stimuli upon changing from comparisons during AC training to samples during CB testing. If BC and CB emerge then the Set-B and the Set-C stimuli are reflexive, as well as symmetrical and transitive. The paradigm in Figure 1.5 therefore represents a simple, convenient and economical test of stimulus equivalence.

It is worth emphasising that equivalence is defined by the emergence of untrained relations. If equivalences form in the example given above, then BC and CB may emerge without explicit training; there would be no need to reinforce correct responses on the BC and CB test trials. Equivalence, then, is defined by functional, and not formal, properties. It would be possible to directly teach BC and CB via differential reinforcement, but the final performance need not represent equivalence. Responding that can be described in terms of equivalence need not be based upon equivalence per se.

THE IMPORTANCE OF STIMULUS EQUIVALENCE TO A BEHAVIOURAL ANALYSIS

The very process of analysing arbitrary matching performances to determine whether they involve equivalence can yield enormous practical benefits. The tests themselves are a vehicle for teaching (Sidman, 1977). In early experiments, Sidman and colleagues (Sidman 1971; Sidman and Cresson, 1973) assessed the performance of retarded youths who had been given up by others 'as hopeless prospects for any type of pre-academic training' (Sidman, 1977). Figure 1.6 shows the paradigm adopted. The subjects learned (or demonstrated they were already able) to select picture comparisons conditionally upon any of 20

Figure 1.6 Equivalence paradigm from the studies by Sidman (1971) and Sidman and Cresson (1973). Arrows point from samples to comparisons. Solid arrows represent relations present prior to testing and broken arrows indicate relations assessed during testing.


B
PICTURES
AB
BD
A
DICTATED
NAMES
BC
CB
D
NAME
ORALLY
AC
CD
C
PRINTED
WORDS

corresponding sample words, dictated by the experimenter. This performance is represented as the AB relation in Figure 1.6. Examples of the twenty dictated words and corresponding pictures were AXE, BED, BEE, BOX etc. After AB was established, the subjects learned AC - to match dictated word samples to corresponding printed word comparisons. Finally, BC and CB equivalence tests were given (see the broken arrows in Figure 1.6). The subjects all proved able to relate pictures to printed words (BC) and printed words to pictures (CB) even though they were not explicitly trained to do so. The direct teaching of 40 conditional relations (20 AB and 20 AC) resulted in the emergence of 40 more (20 BC and 20 CB). In addition, the subjects were able to name each stimulus aloud. As Sidman has stated on numerous occasions, stimulus equivalence permits an impressive economy and efficiency in teaching and learning; you train some and you get many more for free. The retarded youths emerged with a reading vocabulary of 20 words, 'a substantial starting point for a teacher who would otherwise be at a loss as to how even to begin to teach such students to read' (Sidman, 1977, p.357). The same techniques have established equivalences not just between words and pictures but also between words and numbers (Friedman, 1974; Gast, Van Biervliet and Spradlin, 1979), words and colours (Mackay and Sidman, 1984), coins of differing value (McDonagh, McIlvane and Stoddard, 1984) and words, manual signs and pictures / objects (Clarke, Remington and Light, 1986; Van Biervliet, 1977).

But the practical benefits of stimulus equivalence transcend the restricted teaching environment, by expanding the scope of what we learn ordinarily and incidentally, without conscious teaching efforts. Hayes and colleagues (in Hayes, Devany, Kohlenberg, Brownstein and Shelby, in press) draw attention to the maladaptive products of equivalence. They argue that through equivalence:

> "Instances of generalization may occur that have a degree of scope and precision that could not readily occur otherwise. For example, a phobic person may see a graphic description of a plane crash on television and may then avoid riding in the family car because the pictures of the plane are in a class with the word 'plane' and the word 'plane' is (under some contextual conditions) in a class called 'transportation vehicles' and this class contains the word 'car' " (Hayes et al, in press).

On a more positive note, equivalence also allows a person to behave adaptively to situations that are not directly experienced. For example, let's assume a child has learned to avoid the flames of a real fire, and that the word 'hot' is also evoked when he / she sees the flames. If the child then learns that radiators, which look quite unlike 'real' fires, are also 'hot' then, given equivalence between the elements (no pun intended!), he / she will subsequently avoid the radiator. The emergent behaviour is extremely adaptive - the child does not have to experience a burn from the radiator in order to know of its potential dangers. It is interesting to speculate how many other disasters may have been avoided by the safety net of equivalence.

Practical benefits apart, the study of equivalence has some startling theoretical implications for behavioural analysis. In some respects equivalence may be seen as problematical to behaviourists because it is (to them) an unexpected phenomenon. The laws of learning derived from animal research are unable to account for it. More specifically, equivalence is not encompassed by the three-term contingency (Catania, 1984), the fundamental unit of stimulus control (Sidman, 1986). In a three-term contingency, if a particular discriminative stimulus is present and if the subject produces an appropriate response to it then a reinforcer will follow. The three terms are related via conditionality; the contingency involves unidirectional relationships which do not reverse as required by equivalence tests. Hayes (in press) gives the example of a primate hiding in a thicket to avoid a nearby predator. Approaching a thicket given a lion does not imply the reverse i.e. approaching a lion given a thicket:

> "In the natural environment, the contingencies supporting conditional discriminations rarely seem to be symmetrically arranged in this sense... Most commonly, if the functions were reversed the consequence seemingly would either be extinction, or, as in the example of the lion, notable punishment" (Hayes, in press).

Evidence reviewed later in this chapter strongly suggests that the three term contingency's failure to predict equivalence is a direct consequence of its derivation from research with animals. It is apparently no coincidence that the equivalence paradigm evolved from research with humans. Indeed, the very fact that equivalence has appeared only recently on the behaviourist's agenda is a graphic illustration of the dangers, previously expressed by Lowe (1983), of relying exclusively on animal behaviour to the detriment of analysing the behaviour of humans.

Behaviourists have embraced equivalence, despite the problems it may pose them, because it promises to fill a space traditionally occupied by cognitivists:

> "The emergence of equivalence from conditionality permits Behavior Analysis to account for the establishment at least of simple semantic correspondences without having to postulate a direct reinforcement history for every instance. Instead of appealing to cognitions, representations, and stored

> correspondences to explain the initial occurrence of appropriate new behavior, one can find a complete explanation in the (equivalences) that are the prerequisites for the emergent behavior" (Sidman, 1986, p.236).

Of course one must still explain the equivalences themselves, but the general point of Sidman's comment remains.

Before equivalence research began, the experimental analysis of conceptual behaviour was restricted to stimulus classes which can be formed and transferred on the basis of physical similarities shared between each class member. Such concepts (often denoted as 'concrete' (Goldstein and Scheerer, 1941) or 'non-arbitrary' (Hayes and Brownstein, unpublished)) are readily learned even by pigeons. Typically, the birds first learn to discriminate between photographs according to whether the pictures do or do not show instances of the concept, and then the birds transfer discriminative responding when novel instances are shown. These skills have been demonstrated with a variety of concepts including people, trees, fish, bodies of water, pigeons, a specific person, leaves, man-made objects and letters of the alphabet (see Herrnstein, 1979). However, in all these cases, transfer may be based on nothing more remarkable than the fundamental process of stimulus generalisation, acting on a single feature or set of features common to all members of the concept.

But not all concepts are like this. Concepts such as 'colour', 'number' or 'noun' seem to be governed by processes other than stimulus generalisation, because their individual exemplars cannot be logically associated on the basis of physical similarity. Equivalence is important partly because it now permits, for the very first time, a behaviour analysis of these 'abstract' or 'arbitrary' concepts. Williams (1984) chose to emphasise this point:

> "The research by Sidman and his collaborators is virtually unique in its investigation of how arbitrary conceptual categories may be created. Given that the origin of stimulus equivalence is perhaps the most venerable issue in the study of cognition, the applicability of behavior analyses to that issue, with both human and nonhuman subjects, will have major implications for the future of research on complex stimulus control" (Williams, 1984, p.481).

But perhaps the area in which equivalence has the greatest impact is the behavioural analysis of <u>language</u>:

> "In the thirty years since the publication of 'Verbal Behavior' (Skinner, 1957), empirical progress in the behavior-analytic understanding of language has been disappointing...the study of stimulus equivalence provides another, possibly more fruitful, avenue for the study of language phenomena" (Devany et al, 1986, p.256).

Cognitive psychologists have rejected the traditional behavioural approach to language, as defined in Skinner's book, 'Verbal Behavior' (Skinner, 1957). They have focussed on two main areas to support their claim that language learning 'lies beyond the conceptual limits of behaviourist psychological theory' (Chomsky, 1972, p.72).

The first area involves the symbolic nature of words and semantic meaning. From a Skinnerian viewpoint, symbols are nothing other than discriminative stimuli which derive their 'meaning' from participating in a three-term contingency. For example, the word 'food' means FOOD inasmuch as, for instance, food-producing behaviour is occasioned by the word. Similarly, an utterance 'refers' to something to the extent that a stimulus (the 'referent') exerts conditional control over the utterance.

However, most psychologists would argue that symbols are something other than mere discriminative stimuli and that there must be something more to meaning than basic stimulus control. But what are the extra 'somethings'? According to Devany et al (1986) answers to this question are notably lacking:

> "In traditional views of language, much is made of the symbolic nature of words, but relatively little work has been done to show why or how words came to function as symbols. Instead, the literature has asserted that words do act as symbols and has traced their use. Verbal humans are said to be able to 'manipulate symbols' (Clark & Clark, 1977), to 'map words onto internal concepts' (Nelson, 1974), or to use words to 'refer' to objects, events, or relations (Premack, 1976). Exactly what constitutes a symbol and what gives rise to symbolic relations in verbal humans is rarely addressed. For instance, the textbook quoted above by Clark and Clark repeatedly refers to the symbolic nature of language, but fails even to include the word 'symbol' in its index. It is as if the origin or nature of symbolic activity per se need not be explained" (Devany et al, 1986, p.243).

Devany et al see equivalence as the behavioural key to freeing symbolic activity from its illusory cognitive web:

> "In the context of stimulus equivalence, a 'symbol' and its 'referents' form a class of functionally substitutable elements. The relation between a symbol and its referent is not a unidirectional conditional relation (although the members of the class are conditionally related to each other); the relation is functionally reversible. The relations among the members of an equivalence class appear to approximate what psycholinguists and others mean when they say that a word represents or 'stands for' its referent in a way that a conditionally related response does not" (Devany et al, 1986, p.244).

They add that the relation between a symbol and a referent seems necessarily bi-directional. A word 'stands for' another event only if the event 'is called' the word.

Sidman views equivalence as a pre-requisite for simple semantic correspondences:

> "The equivalence paradigm provides exactly the test that is needed to determine whether or not a particular conditional discrimination involves semantic relations" (Sidman and Tailby, 1982, p.20).

Sidman suggests elsewhere that when, for example, numbers and printed number words become equivalent then we may say that they have the same meaning or that each is the meaning of the other (Sidman, 1986).

The enormous complexity of language, and the sheer speed with which it is acquired during childhood, have led many to reject the notion that it is governed by operant laws of learning. Chomsky led the revolt:

> "It is simply not true that children can learn language...... through careful differential reinforcement. ....It is also not easy to find any basis to the claim that reinforcement contingencies are the single factor responsible for maintaining the strength of verbal behavior. The sources of the 'strength' of this behavior are almost a total mystery at present" (Chomsky, 1959, pp.42-43).

When Chomsky wrote this, he claimed (not without justification) that behaviourists had no real means of explaining the appearance of novel grammatical utterances during language acquisition. Behaviourists at the time attempted to account for novel behaviour by appealing to the principle of stimulus generalisation. But, as Chomsky pointed out, this was not sufficient:

> "Every time an adult reads a newspaper, he undoubtedly comes upon countless new sentences which are not at all similar, in a simple, physical sense, to any that he has heard before..... Talk of 'stimulus generalization' in such a case simply perpetuates the mystery under a new title. These abilities indicate that there must be fundamental processes at work quite independently of 'feedback' from the environment" (Chomsky, 1959, p.42).

Now, however, behaviourists no longer need to overburden stimulus generalisation with the task of accounting for novel verbal behaviour. Perhaps stimulus equivalence is the 'fundamental process' which Chomsky unknowingly referred to in the quote above. But then it would be quite wrong to assume that the process works independently of environmental feedback, because the equivalence paradigm exposes a source of reinforcement for novel verbal behaviour:

> "By definition, the existence of a class of equivalent stimuli permits any variable that affects one member of the class to affect all members. Even when stimuli bear no physical resemblance to each other, their inclusion within a class provides a route for extending the influence of reinforcement and other variables" (Sidman and Tailby, 1982, p.20).

The transfer of function from one member of an equivalence class to others has already been applied to the related problems of generative grammar and syntax. In a recent study, Lazar and Kotlarchyk (1985) first established two separate classes of five equivalent stimuli with 5-6 year old children. Then, <u>contextual control</u> was established over sequential responding. The subjects were presented with one member from each class, red from Class A and green from Class

B. In the presence of Tone 1 the subjects were taught to touch red first, and then green, whereas in the presence of Tone 2, the reverse was required - touch green first, and then red. Next, in subsequent test phases, the subjects transferred contextually controlled sequential responding to all four remaining members of each equivalence class.

This result is far from trivial. Firstly, it represents the initial step toward a functional analysis of novel syntactical relations. Lazar (1977) gives the simple example of a young child taught to say 'red ball' in the presence of that object. Given that the adjective 'green' is in the same equivalence class as 'red', and the noun 'hat' is in the same class as 'ball', then the child is also likely to produce the grammatically correct utterance, 'green hat' the first time he sees one. Moreover, in demonstrating <u>contextual</u> <u>control</u> over sequential responding, Lazar and Kotlarchyk (1985) have expanded the analysis further. The significance of this data is not lost on Wulfert and Hayes (in press), who replicated and extended Lazar and Kotlarchyk's findings:

> "Consider the example of an English speaker who in the presence of a red traffic light might utter 'red light', whereas a Spanish speaker in the same context would say 'luz roja' (literally 'light red'). In a different context, an English speaker's utterance controlled by the color of a garment might be 'light red', while a Spanish speaker under the same stimulus conditions would emit 'rojo claro' (literally 'red light'). Whether a bilingual speaker will order a response sequence in terms of 'property first, object second' (English) or the other way around (Spanish), will depend on the control exerted by a particular audience. A similar argument could be made for active vs. passive voice and other language phenomena which require an inversion of word order, but conserve the meaning of an utterance" (Wulfert and Hayes, in press).

WHERE DOES EQUIVALENCE COME FROM?

The previous section suggests that there is much to gain from discovering the origins of stimulus equivalence. The issue of how physically different stimuli acquire similar controlling properties has been a concern for many years. Beasty (1987) has conducted a thorough review of several paradigms which have previously adopted the term 'stimulus equivalence'. However, not all of these map directly onto present day equivalence tests. There are, nevertheless, two paradigms which appear to 'fit the bill', and both have postulated similar determinants for equivalence.

In their theory of the acquired equivalence of stimuli, Miller and Dollard (1941) have emphasised the human ability to react equivalently to stimuli with widely discrepant perceptual features. They considered acquired equivalence as critical to 'higher mental processes' such as reasoning and foresight and they postulated a plausible explanation of how equivalence is obtained and how it may function in complex human behaviour. Miller and Dollard proposed that stimuli become equivalent by controlling a common mediating response, usually a common verbal label. Counting is a simple example. Although five 10 pence coins and one 50 pence coin present completely dissimilar perceptual cues, after counting the 10 pence pieces we accept them as equivalent to the 50 pence piece provided our count produces the common label, 'fifty'.

Miller and Dollard's hypothesis has received considerable empirical support (see Beasty, 1987 and Reese and Lipsitt, 1970 for reviews). An early study by Birge (1941) illustrates the phenomenon. The subjects (third, fourth and fifth grade school children) each participated in three distinct experimental phases. In Phase 1, each child was shown four boxes, identical except for four nonsense shapes drawn on their covers. The subjects were required to call one pair of boxes 'towk' and the other pair 'meef'. After the subjects had learned labels for each pair of boxes, Phase 2 began; only one of the 'towk' and 'meef' boxes were presented, and the subjects learned that candy could always be found under the 'towk' box regardless of its spatial position relative to 'meef'. Finally, in Phase 3, the other pair of boxes were presented, to test for transfer of the choice responses. The children were split into four groups according to whether 'towk' and 'meef' verbalisations were required during Phase 2 and/or 3 (obviously, all subjects were required to name during Phase 1). Group One were required to name during Phases 2 <u>and</u> 3, Group Two during Phase 2 only, Group Three during Phase 3 only, and Group Four during neither phase. The results indicated that it was not sufficient to simply learn common labels for the stimuli - neither Group Three nor Group Four gave any evidence of transfer during Phase 3. Transfer only occurred for subjects in Groups One and Two, all of whom produced common names whilst learning the choice response in Phase 2. Furthermore, the best performance was recorded by the Group One subjects who continued to use common labels <u>throughout</u> the experiment.

The study of verbal learning via <u>paired-associates</u> also appears to correspond, at least procedurally, with modern day studies of equivalence. In a typical paired-associate experiment, pairs of nonsense syllables are presented and the subject has to learn to associate the first stimulus of a pair to the second. The subject is shown the first stimulus and is required to say what the second stimulus is, before it appears. After learning the AB relation (where A is the first term of the pair and B is the second) the subject may be tested for <u>backward association</u>; the B term appears, and the subject has to give the A term as a verbal response (BA). Alternatively, after learning AB and BC, a chaining test may be given, which tests the subject's ability to label each A term with the appropriate C term (AC). Response equivalence involves teaching AB and AC and assessing BC and CB. Finally, stimulus equivalence refers to establishing AB and CB before testing AC and CA. The backward association paradigm resembles the modern day procedure for testing symmetry, chaining resembles transitivity, and response equivalence and stimulus equivalence both resemble procedures now adopted in combined tests of equivalence. And, just like the acquired equivalence studies, the paired-associate experiments led to widespread acceptance of response mediation as the mechanism for transfer during test phases (see Jenkins, 1963, 1965; Jenkins and Palermo, 1964).

Modern day studies of stimulus equivalence have seemed less concerned than their historical predecessors with determining the necessary or sufficient conditions for equivalence formation. Most studies now appear content with evaluating quantitative parameters (such as the number of stimuli that may be incorporated into a class, or the number of classes that may emerge at any one time), and with adding to the complexity of the phenomenon by evaluating higher-order contextual control of equivalence classes. Such studies are not without significance but, thus far, they have failed to promote an understanding of the origins of equivalence. Stimulus networks of staggering complexity have been established (e.g. Matthijs, 1988) with apparently little concern for what might be producing the simplest reducible component. Furthermore, few studies have bothered to record or report the subjects' verbal behaviour (in particular, stimulus naming) during equivalence tests. This is an extraordinary omission given the historical link between verbal mediation and equivalence. If the rest of this review seems somewhat brief this only reflects the paucity of studies addressing the question 'where does equivalence come from?' The reader may find a more exhaustive review of other equivalence studies elsewhere (see Beasty, 1987).

We begin by focussing on a series of equivalence studies conducted by Sidman and colleagues (Sidman, Cresson and Willson-Morris, 1974; Sidman and Tailby, 1982; Sidman, Kirk and Willson-Morris, 1985; Sidman, Willson-Morris and Kirk, 1986). The results from these studies have led Sidman et al to conclude that naming is neither necessary nor sufficient for equivalence formation. They accept that differential responses may mediate (and may possibly facilitate) the emergence of new stimulus relations, but they also claim that equivalence may form in the absence of mediational naming. This view has been reinforced by others (Lazar, Davis-Lang and Sanchez, 1984), and its acceptance may explain why there has been relatively little interest in the verbal behaviour of subjects during equivalence experiments.

The following review aims to show that (i) evidence against the critical role of naming in equivalence formation is not particularly convincing when examined in detail, and that (ii) there is compelling evidence to support the contrary view that naming is necessary for stimulus equivalence.

THE ROLE OF NAMING IN STIMULUS EQUIVALENCE

An early study by Sidman, Cresson and Willson-Morris (1974) provides what seems to be the best evidence for the independence of equivalence and naming. The subjects were two severely retarded Down's syndrome adolescents (JC and PA). Extensive pre-tests confirmed that both subjects were unable to name or perform arbitrary matching with any of the experimental stimuli (both subjects, however, were considerably experienced at naming and matching other stimuli). The subjects were taught AB and BC matching, and then they were given tests for AC (transitivity), CB (symmetry), and oral naming of the Set-B and Set-C stimuli. The Set-A stimuli were dictated words corresponding with 20 pictures (Set-B) and their printed word equivalents (Set-C) for subject JC, and 9 upper-case printed letters (Set-B) and their lower-case equivalents (Set-C) for subject PA. Both subjects passed the AC and CB matching tests. However, neither subject was able to consistently name the Set-B and Set-C stimuli, and naming of the Set-C stimuli was particularly poor. The authors therefore concluded that equivalence had not been mediated by stimulus names produced by the subjects.

Now this conclusion appears reasonable, but it stands (and falls) on the assumption that the subjects' naming test scores were <u>meaningful</u>. But this assumption may be incorrect. Both subjects, for example, scored about 50% correct on the Set-C naming trials. This, however, does <u>not</u> mean the subjects were unable to name the stimuli consistently; it merely represents the fact that on half the trials the subjects did not produce the name <u>required by the experimenter</u>. Precise details were not presented for all of the subjects' naming responses, but the few examples given were particularly telling. Subject JC said 'hammer' to both the picture of an axe and the printed word AXE, and he also said 'cow' to both the picture of the pig and the printed word PIG. Subject PA also produced the common name 'Seh' to both upper and lower case G. All these responses counted as 'errors' and contributed to the poor naming scores. Furthermore, Sidman et al admitted that 'it is difficult to attribute such naming errors to any process other than expressive mediation' (Sidman et al, 1974, p.272).

But there are other, more fundamental, problems with naming tests. In a recent study, Hird and Lowe (1985) set out to examine equivalence in mentally handicapped adults. A conventional MTS procedure was used, but the experiment differed from all others in one important respect - all training and test sessions were recorded on audio and videotape. These tapes were analysed, and notes were made of any spontaneous verbal behaviour produced by the subjects while performing on the matching tasks. In addition, all subjects were given a post-experimental naming test of the kind presented to Sidman et al's subjects. In the naming test, the subjects saw each stimulus one at a time and were asked, 'What is it?' and 'Do you have a name for it?'

Four of the five subjects passed the equivalence tests, and they all named the stimuli at some stage during training. During the post-experimental naming test, however, three of the subjects gave the stimuli different names than the ones they had employed spontaneously while performing arbitrary matching (see Table 1.1). During the naming test, the subjects appeared to interpret the experimenter's questions and prompts in a complex fashion, often giving complex analytical geometric descriptions of the stimuli. For example, John's naming test responses indicated that he was searching for the description which he thought the experimenter was looking for. During training, John spontaneously labelled the green hue as 'yellow', but during the naming test he said, 'I just call it a square...square with lines down' (the computer drew this colour as a square block made up of a succession of closely spaced vertical lines). John's elaborate description of the cross also bore no semblance to his spontaneous label for the same shape. In David's naming test, it appeared that he did not have distinctive names for red and green because he called them both 'squares'; during training, however, he freely used the conventional names for the hues. Ian produced similar differences in hue naming, and when prompted to name the vertical line he said, 'looks like an 'I' to me', despite previously naming it spontaneously as a 'line'.

Table 1.1 Spontaneous and elicited naming by the subjects in the Hird and Lowe (1985) study. (semi-colons separate a subject's first and second naming test responses; V=vertical line, G=green, R=red).

SUBJECT	STIMULUS	NAMING RESPONSE: SPONTANEOUS (During Training)	NAMING RESPONSE: PROMPTED (During Naming Test)
IAN	V	Line	looks like an 'I' to me; 'I'
	O	Circle	circle; circle
	G	Green	square; square green
	R	red	its a square red; square red
	Δ	triangle	triangle; triangle
	X	cross	cross; cross
JOHN	V	straight one	from the top to the bottom its the same; oblong
	O	round one	circle... its round; circle
	G	yellow	I just call it a square; square with lines down
	R	red one	its red but, its square; square with no lines
	Δ	triangle	triangle; triangle
	X	other one	its two lines... its cut into triangles... a triangle but no bottom... there are little squares; 'triangle...them two are the same'
DAVID	V	line	line; line
	O	'O'	circle; circle
	G	green	square?; square
	R	red	Another square; square again, huh
	Δ	triangle	triangle; triangle
	X	cross	cross; an 'X'

The Hird and Lowe (1985) study simply confirms what logic demands - a subject's verbal responses elicited in the contrived context of a naming test need not necessarily correspond with those emitted spontaneously, and within the distinctly different context of matching-to-sample. The subject's naming test responses may depend upon how the experimenter's prompt is interpreted. Questions put by others normally have a purpose; they are often meant to correct errors e.g. asking 'what did you say?' upon hearing someone speak inappropriately. When the experimenter asks what appears to be a simple question during the naming test, the subject may produce a different name than usual, perhaps because the question appears to indicate to the subject that his former spontaneous utterances were incorrect. In naming tests, the subjects may simply try to work out what is required, before responding in the way they deem best. Alternatively, if the subjects fail to interpret the situation, or if they are anxious about compounding any apparent naming 'errors', then they may opt to say nothing at all.

The Sidman et al (1974) study produced other data that may be understood better by appreciating the complexity of naming tests. Firstly, why were the subjects generally better at naming the Set-B than the Set-C stimuli? The answer may lie in the 'dynamics' of the naming test. The Set-B stimuli had already been named by the experimenter (on AB training trials) whereas the Set-C stimuli had not. So during naming tests, the subjects may have had less confidence in naming the Set-C stimuli, even if they had named them correctly before. However, as the naming test progressed the subjects may have grown more confident until they felt prepared to offer their names for the Set-C stimuli. Both subjects did, in fact, produce a 'aha' reaction during naming trials, followed by increased production of appropriate Set-C names.

In commenting on these events, Sidman et al revealed what appears to be a behavioural 'blind-spot' in their own terminological repetoires:

> "His [JC's] 'aha' reaction during the oral reading [naming] test suggested that although he had failed to read the words aloud up to that point, he had actually been capable of doing so, and exercised his new capability only in the course of this test. Nevertheless, even though he may have been capable of naming the printed words, he clearly had not been doing so" (Sidman et al, 1974, p.271).

And commenting on subject PA's 'aha' reaction, Sidman et al argued that:

> "Subject PA was apparently capable of naming more lower-case letters than he actually did upon initial testing, but he had not been using those letter names to mediate the crossmodal matching of lower-case to dictated letters" (Sidman et al, 1974, p.271).

But these comments only make sense if one mistakenly equates naming with <u>overt naming</u>. It seems as if Sidman et al had failed to recognise behaviour which the reader of this text is probably engaging in right now, and which few would wish to deny, namely <u>covert verbal behaviour</u>. Bentall, Lowe and Beasty (1985) have discussed the seemingly inexplicable way in which many operant researchers readily reject, or avoid acknowledging, human covert behaviour:

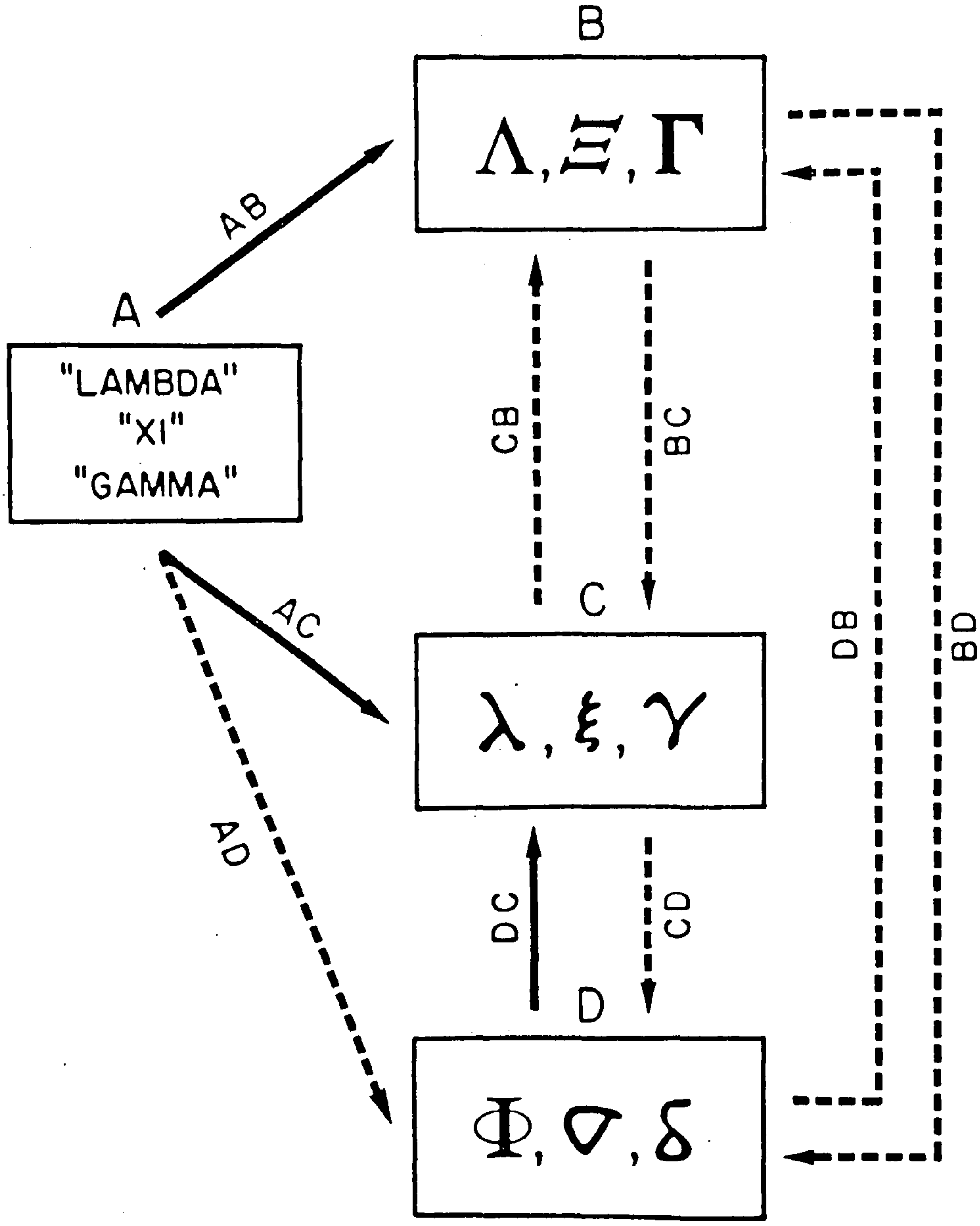
> "That reference to covert behaviour should be considered suspect, in principle, is indeed a strange irony inasmuch as Skinner established the identity of radical, as opposed to methodological, behaviorism largely on the basis of its recognition of the importance of covert events in human behavior (Skinner, 1945, 1957, 1963,

> 1966, 1974). For example, in defining rule-governed behavior, a key concept in contemporary behavior analysis, Skinner (1966) described how an individual constructs his own rules, and may do so overtly or covertly: 'Any actual formulation of the relation between a response and its consequences (perhaps simply the observation 'whenever I respond in this way such and such an event follows') may, of course, function as a prior controlling stimulus' (p.243). Similarly, Bijou, who has contributed much to the study of child behavior, has shown how the analysis of covert events is both consistent with behaviorist theory and is a practical necessity in dealing with problem-solving behavior in children (Bijou, 1976, pp.70-74; Bijou and Baer, 1967). Of course, each researcher is free to choose his / her own research strategy, which may or may not embrace an analysis of the role of covert behavior, but it should be clearly recognised that the radical behaviorist thesis, as articulated by Skinner, does not eschew consideration of such events, but, rather, maintains that it is folly for science to ignore them" (Bentall, Lowe and Beasty, 1985, p.179).

The problems associated with the Sidman et al (1974) study appear to have been compounded in later studies. Figure 1.7 shows the paradigm used by Sidman and Tailby (1982) in their study of equivalence with normal children of five years and older. Four sets of unfamiliar Greek stimuli were used to avoid lengthy pretests of the training and test performances. Once again, Set-A were letter names dictated by the experimenter. The subjects were taught the relationships depicted by the solid arrows in Figure 1.7, and the broken arrows depict relations assessed during unreinforced test trials.

All six children who passed the tests gave consistent and appropriate common names to the stimuli during a 90-trial naming test. Within each set of letters shown in Figure 1.7 (B, C and D) the six children consistently called the one at the left, 'lambda', the centre

Figure 1.7 Equivalence paradigm employed by Sidman and Tailby (1982). Arrows point from samples to comparisons. The stimuli are arranged, for expository purposes, so that auditory "lamba" is matched to the letter on the left in each box, "XI" to the centre letter and "gamma" to the letter on the right; in all other instances letters are matched to each other according to their relative positions in the boxes (see text).


B
Λ, Ξ, Γ
A
"LAMBDA"
"XI"
"GAMMA"
AB
AC
AD
CB
BC
C
λ, ξ, γ
DB
BD
DC
CD
D
Φ, σ, δ

one, 'XI', and the letter at the right, 'gamma'. Furthermore, two of the six children were given naming tests prior to completing the equivalence tests, thus proving capable of common naming even before passing the BC, CB, AD and (in one case) CD tests.

These results raised the possibility that naming may have mediated the emergent matching performances. However, Sidman and Tailby rejected this possibility on the grounds that, during a naming test, one of the six subjects (EW) hesitated before giving the correct names for each Set-D stimulus:

> "Subject E.W. gave all the stimuli names that were consistent with their class membership, but his hesitations and expressions of doubt indicated strongly that although he was capable of naming the Set-D letters, he had never done so until the naming test. The new conditional discriminations involving the D-stimuli emerged before he had ever applied names to those letters.
>
> Subject E.W. was the only one of the eight who yielded such a finding, but his demonstration that the stimulus class could form in the absence of naming cannot be dismissed" (Sidman and Tailby, 1982, p.21).

But if the authors had demonstrated anything here it is nothing but their complete disregard of covert naming. Subject E.W's doubts and hesitations over naming the Set-D stimuli do not demonstrate a prior absence of naming, but (and consistent with the earlier analysis), they may reflect the fact that those stimuli were the only ones not to be named by the experimenter during training (cf. Sidman et al, 1974).

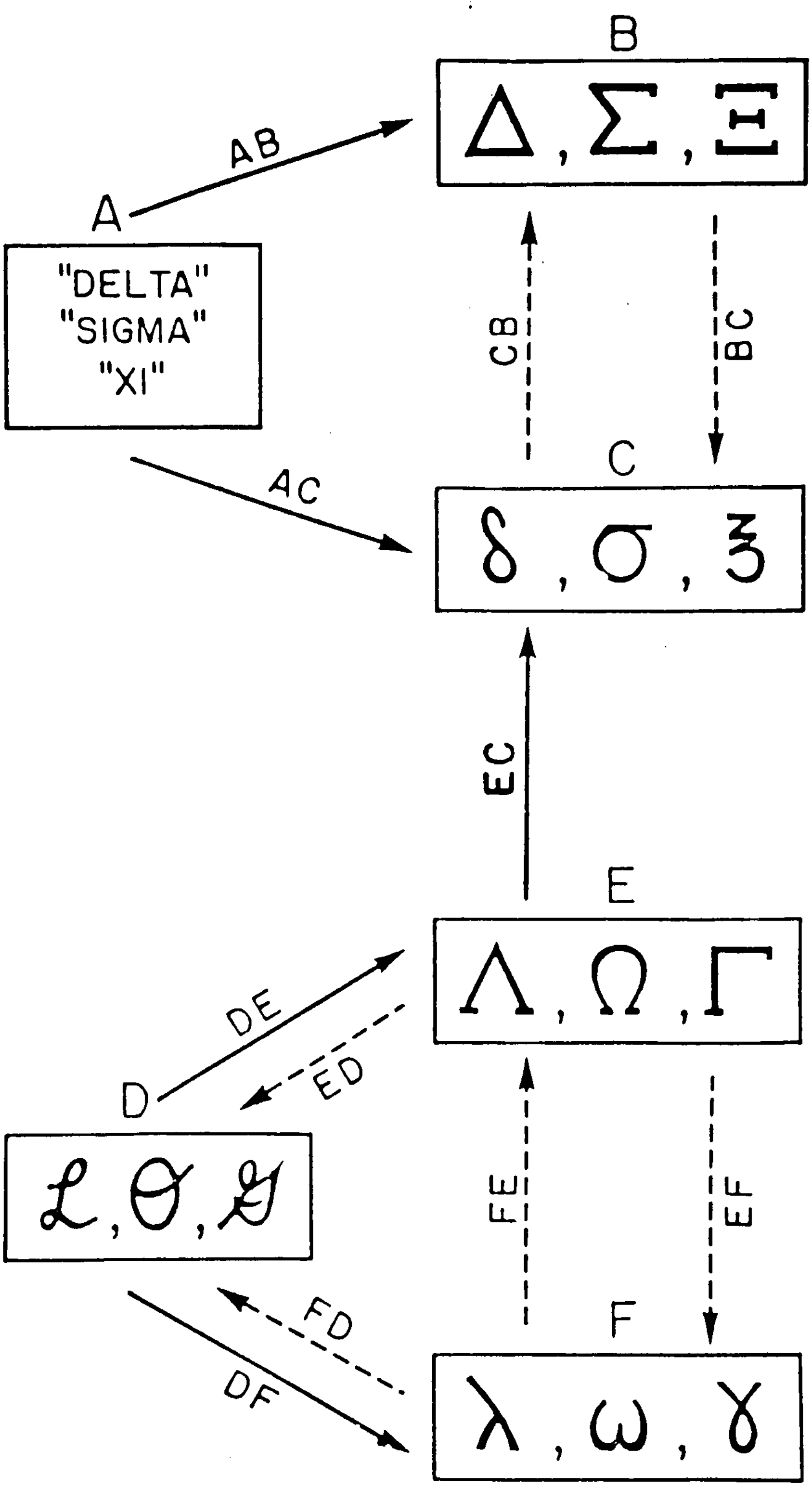
Two of Sidman and Tailby's subjects failed the equivalence tests, but one of these subjects (JO) was able to name the stimuli consistently and appropriately during the naming test which followed. After a repeat equivalence test showed no change in JO's matching performance, Sidman and Tailby concluded that naming was not sufficient for stimulus equivalence.

But even this conclusion requires qualification. Sidman and Tailby have no evidence that the subject continued to name the stimuli during the critical equivalence tests. Perhaps JO failed equivalence because he failed to name the stimuli spontaneously at the time of testing. The results from many experiments suggest that mediation is effective only if the subject continues to produce the mediating response while performing the task in question (see Birge, 1941; Kail,1979; Kendler and Kendler, 1975). The term 'production deficiency' has been applied to instances where the subjects fail to produce a mediating name during testing, despite being able to in other contexts. Subject JO may have named the stimuli during the naming test because he was asked to do so by the experimenter, but then failed to name during the critical equivalence test because the necessary prompts were absent. In this sense, it is not altogether clear from the evidence produced by Sidman and Tailby (1982) that naming is not sufficient for equivalence.

In a later study, Sidman, Kirk and Willson-Morris (1985) expanded the paradigm to include six sets of stimuli (see Figure 1.8). Eight of the eleven subjects eventually passed all of the tests depicted by the dotted lines, while the other three failed some, but not all, of the tests. The three unsuccessful subjects (two normal 5-6 year old children and one Down's syndrome adult with a mental age of 4) left the study prematurely with no assessment of stimulus naming. The eight successful subjects (seven normal children aged 5-10 years, and one 22 year old normal adult) were each given up to two 90-trial post-experimental naming tests.

There were two main points worth noting in connection with the successful subjects. Firstly, in most cases the emergence of

Figure 1.8 Equivalence paradigm from the study by Sidman, Kirk and Willson-Morris (1985). Arrows point from samples to comparisons. The stimuli are arranged, for expository purposes, so that auditory "delta" is matched to the letter on the left in each box, "sigma" to the centre letter and "XI" to the letter on the right; in all other instances letters are matched according to their relative positions in the boxes (see text).


A
"DELTA"
"SIGMA"
"XI"
B
Δ, Σ, Ξ
C
δ, σ, ξ
D
E
Λ, Ω, Γ
F
λ, ω, γ
AB
AC
CB
BC
EC
DE
ED
FE
EF
FD
DF

equivalence was far from automatic; repeated testing and / or the temporary removal of some baseline tasks was necessary for all but one subject, and there was some evidence that a particular baseline task (EC) was somehow interfering with equivalence formation. At no stage, however, did the experimenters consider whether these inconsistencies might be linked to the way in which the subjects named the stimuli. Secondly, after passing the tests, all the successful subjects were able to give the appropriate Set-A Greek names to each of the visual stimuli. In other words, all the subjects had common names for the stimuli. Despite this, Sidman et al still concluded that common naming is not necessary for equivalence. Their 'evidence' came from two subjects who were able to name either all or some of the stimuli not just with their corresponding Greek names but also with English names derived from the Set-D stimuli, L, O and G. (see Figure 1.8).

> "Subject PH applied the Set-A names appropriately to the Class-1, -2, and -3 stimuli in the upper triangle and to the Class-2 stimuli in the lower triangle; sometimes, however, he gave the Set-A names and at other times he gave the English names of the Set-D letters to Class-1 and -3 stimuli in the lower triangle. For two of the classes, therefore, he did not give the same name to each member. ....Subject F.M. gave the English names to the Set-D letters, but applied the Set-A names to all others in both the upper and lower triangles. Although in subsequent tests she proved capable of giving either the Set-A or the English-letter names to all stimuli, her first naming test indicated that she had not originally given the same name to all members of any class" (Sidman et al, 1985, p.41).

These comments seem to indicate that in addition to failing to distinguish between names produced after and during equivalence tests, Sidman et al were now confusing common naming with <u>consistent</u> naming.

But consistent naming is irrelevant; all that a mediational account requires is that the subject gives the same name to each prospective member of an equivalence class. It should hardly matter if the subject can do this with both English and Greek names for all equivalence classes or with English names for some equivalence classes and Greek names for others.

The stimuli in Figure 1.8 also appeared in a later study by Sidman, Willson-Morris and Kirk (1986), who investigated equivalence with two normal 5 year old children and four mentally retarded adolescents. The subjects were taught AB and AC, and DE and DF, and then tests were given for auditory-visual ABC and visual-visual DEF equivalence classes (the two classes remained separate because the subjects were not taught the EC relation in Figure 1.8). Equivalences were formed by all 6 subjects, although three required repeated tests. In post-experimental naming tests, the subjects appeared to give common names with greater consistency to stimuli in the auditory-visual than the visual-visual classes. During the naming tests, the subjects often responded with 'I don't know' or some such similar response.

From these results alone, Sidman et al concluded that the emergence of equivalence did not require mediation by naming. However, no such thing had been demonstrated. The results, at best, only suggested that equivalence may emerge without common naming, and not without naming per se. It is perhaps unreasonable to assume that common naming is the only way in which linguistic processes may mediate equivalences. Furthermore, the potentially unreliable nature of naming tests casts doubt upon whether this study even satisfactorily demonstrates a case against common naming. What is somewhat ironical is that this point has been recognised by two of Sidman et al's closest associates. In commenting on the study above, Stoddard and McIlvane

(1986) asked the following question:

> "Do these data lay to rest the question of response mediation as the critical basis for stimulus equivalence? Probably not.... Some examples may serve to illustrate the difficulty of this research question. Suppose a given subject characterizes all the stimuli in the entire visual classes with a common descriptive adjectival term, like 'rounded', 'pointed' or 'pointing that way', perhaps derived from primary stimulus generalisation....Alternatively, suppose a common descriptive term, such as 'Set 1' vs. 'Set 2' was applied, as we do in talking about stimuli within classes. When asked the question, 'What is it?', in relation to a given stimulus, perhaps the subject's verbal conditioning history had not prepared him or her to use descriptive terms as labels, leading to 'I don't know' (its name) responses on the naming tests. Would other methods of testing have evoked descriptions?
>
> What would it mean if additional 'questioning' did reveal some common response, emitted in the presence of each member of a class? On the one hand it might appear that one had isolated a potential form of response mediation. On the other hand, one might argue that the additional questioning had merely set the occasion for further discriminative behavior, capable of verbal description by the subject. Such responses need not have been functional in the original formation of equivalence relations among the stimuli. Additional research would be required to separate these accounts - research that would likely be extremely difficult to accomplish" (Stoddard and McIlvane, 1986, p.157).

Difficult but (as Stoddard and McIlvane imply) not necessarily impossible. Perhaps the first step toward progress would be to record spontaneous naming during the matching tasks. Then perhaps we may begin to discover ways in which the same names may be elicited from the subject via prompting, either during naming tests or during the matching task itself.

Sidman et al's dismissal of naming has been backed up by one other study conducted by Lazar, Davis-Lang and Sanchez (1984), but this suffers from the same problems as the studies reviewed above. The subjects were four normal 5-7 year old children. This time the stimuli were all visual, comprising of Greek and Hebrew letters. All subjects eventually passed the equivalence tests (but, once again, some children required repeated testing and baseline manipulations before doing so). Lazar et al did not record, or report of, any spontaneous naming by the subjects. Instead, stimulus names were elicited from the subject in two separate contexts - a post-experimental naming test (the subjects were shown each stimulus in turn and were asked 'Tell me what this is' or 'What is it?') and during nine trials of an equivalence test (here, the instructions were 'Don't touch; just point to them and tell me what it is'). In both conditions, subjects produced distinctive names for each of the stimuli.

Once again, and despite only having evidence against common naming, the conclusion formed was that equivalences may emerge in the absence of mediating names, and that naming is not a pre-requisite for equivalence formation. But, as argued earlier, even the evidence against common naming is equivocal when based upon elicited names. Perhaps the most significant finding of all served to highlight the problems of eliciting verbal responses. In approximately 40% of the naming trials, different names were given to the same stimuli across the two prompting conditions.

In summary, thus far there has been no convincing evidence against the critical role of naming in the emergence of equivalence. Now we must ask: is there any evidence that naming is necessary for equivalence?

If equivalence requires naming then animals should be unable to form equivalences of the sort readily found in language-able humans. Traditionally, attention has focussed on reflexivity and transitivity in animals, but, for a variety of reasons, this data will not be examined here. Animal studies of reflexivity have been abundant in the past, and no doubt will continue to be in the future, given the current level of debate on the topic (see Beasty, 1987). There has been some evidence of transitivity in non-humans (cf. D'Amato, Salmon, Loukas and Tomie, 1985; McGonigle and Chalmers, 1977) but it has been noted that transitivity may be more amenable to direct conditioning explanations (Devany at al, 1986). In addition, neither reflexivity nor transitivity are sufficient for proving equivalence. Symmetry, too, is required. And, at present, the evidence against symmetry in animals appears incontrovertible.

An early study by Gray (1966) investigated symmetry in pigeons. Gray concluded that symmetry was present, but the data does not support this conclusion. After reaching criterion on AB trials (matching red to green, and blue to yellow), the three pigeons' scores on a 56-trial BA symmetry test were 64%, 64%, and 57% correct. These scores were taken as evidence for symmetry because, according to statistical tests, they were significantly above chance level (50% correct). However, this is clearly an inappropriate comparison; a true test of symmetry compares BA performance with a fixed criterion of accuracy, usually 85% correct or better. The above chance symmetry test scores probably reflected initial learning due to differential reinforcement; although correct responses on test trials did not produce food reinforcers, incorrect responses produced a ten-second blackout in the test chamber.

In another study, Rodewald (1974) trained three pigeons to match a red sample to three vertical lines and a green sample to three horizontal lines, before presenting a 90 trial symmetry test in which all correct responses were reinforced. The results were very similar to Gray's; the birds' symmetry test scores were 63%, 63% and 73% correct. Rodewald came to the following conclusion:

> "There is little, if any, evidence that the animals learned the invariant symbolic relations between the color and figure stimuli. Rather, they seemed to have learned how to respond in the presence of each sample stimulus. When the sample was changed, further learning was necessary" (Rodewald, 1974, p.990).

Holmes (1979) came to the same conclusion. His pigeons failed a symmetry test in spite of being trained on a large number of identity matching problems, some of which included the symmetry test stimuli.

Hogan and Zentall (1977) assessed symmetry in pigeons by comparing two groups according to their rates of learning a novel arbitrary matching task. Group 1 pigeons were trained on a new task which was the symmetrical counterpart of a task already learned, whereas Group 2 pigeons were given new tasks and old tasks which were not symmetrical counterparts. Both groups learned the new tasks at the same rate, thus indicating an absence of symmetry. The results were replicated in a second experiment, using different stimuli, and in a third, using a zero-delay matching procedure. Using essentially the same design, D'Amato, Salmon, Loukas and Tomie (1985) also found no evidence of symmetry in the matching performances of pigeons and cebus monkeys.

Sidman, Rauzin, Lazar, Cunningham, Tailby and Carrigan (1982) conducted symmetry tests in three experiments with rhesus monkeys, and one experiment with baboons. Both species failed the tests, even when reinforcement was available on symmetry test trials. In contrast, six normal 4-5 year old children passed exactly the same tests without differential reinforcement.

Finally, two studies have presented pigeons with combined tests of equivalence. Both experiments were similar, but somewhat unorthodox.

Kendall (1983) presented pigeons with one of two equivalence tests, either (i) Train AB and CB; Test AC or (ii) Train AB and AC; Test BC. However, unlike conventional studies, all three stimulus sets were available during training and testing. Set-A were two 'signal' lights, Set-B were two keys on the front wall of the test chamber and Set-C were two keys on the side wall. This arrangement was designed to encourage the birds to engage in overt mediating behaviour during the equivalence tests. All of the birds failed the equivalence tests, even though reinforcers were available for correct responses. The results were later confirmed by Lipkens, Kop and Matthijs (1988), who used essentially the same procedures.

There are, to date, only two studies that appear to have anything like positive evidence of equivalence in animals. One, by McIntire, Cleary and Thompson (1987), used monkeys as subjects, whereas the other, by Edwards, Jagielo, Zentall and Hogan (1982), used pigeons. In both studies, however, the results are readily explained in terms of simple conditioning principles rather than the emergence of untrained relations. The features distinguishing these studies from true tests of equivalence are real but somewhat difficult to appreciate at first, though they become far clearer in the context of new data from experiments forming part of this thesis. Consequently, both studies are examined in detail in Chapter 5 of this thesis.

We may conclude that, to date, and despite considerable efforts, there has been no success in unequivocally demonstrating stimulus equivalence in any non-human species, including higher primates. These data, of course, are entirely consistent with the view that naming is necessary for equivalence, and so it is all the more remarkable that there have been hardly any attempts to confirm the 'naming hypothesis' with humans. Generally speaking, few have seen the need for such studies (but see Dixon and Spradlin, 1976; Lazar, 1977). The potential relationship between naming and equivalence has eluded recognition even in the most obvious circumstances. For example, Stromer and Osborne (1982) noted that out of twelve retarded adolescents, only one, M.P., failed a standard equivalence test. What was so different about M.P? Stromer and Osborne noted that:

> "Except for M.P., all of the present subjects were relatively proficient in expressive language. They frequently engaged in spontaneous conversation with the experimenter, and related detailed accounts of past and future activities. M.P., however, displayed neither spontaneous expression nor vocal imitation" (Stromer and Osborne, 1982, p.347).

As far as the present author is aware, this was the first published test of equivalence with a human subject completely lacking in functional expressive language. This result does not <u>prove</u> the claim that naming is necessary for equivalence, but it certainly supports such a claim; nevertheless few have considered its potential significance (but see Beasty, 1987).

There has in fact only been one published study attempting to systematically investigate the role of language in equivalence. Devany, Hayes and Nelson (1986) compared equivalence formation in three groups of subjects: normal 2 year old children, 2-4 year old retarded children with functional spontaneous speech and signing, and 2-4 year old retarded children with no functional verbal skills. The subjects were trained on AB and AC matching tasks with visual stimuli, before being tested on BC and CB equivalence. The results completely vindicated the 'naming hypothesis'. Both the normal and retarded / language groups passed the equivalence tests, whereas the retarded /

no-language group failed. Devany et al concluded that language and stimulus equivalence are closely related in some, hitherto unknown, way.

The above study, however, is not without problems. Firstly, the experimental procedures left much to be desired. For instance, the experimenter and subject sat together, either at a table (retarded children) or on a rug (normal children). The experimenter presented the stimuli by hand and she also administered a variety of reinforcers, including 'social praise', directly to the subject. Situations like these may allow the experimenter to unwittingly cue correct responses from the subject. Furthermore, there appeared to be other opportunities for adventitious reinforcement. Reinforcers were delivered on every third or fourth test trial. Although such schedules are procedurally non-contingent they do not necessarily guarantee an independence between reinforcer deliveries and correct responses.

The above analysis suggests that the normal and retarded / language subjects may have passed equivalence tests simply by learning to respond correctly through adventitious reinforcement. None of these subjects passed the tests straight away; their performances gradually improved during testing, as if they were learning what to do. But if reinforcement was responsible for their performances, why did the retarded / no language group <u>fail</u> the tests? The answer may lie in the training data. The retarded / no language subjects took the longest to learn the original AB and AC baseline tasks. It would not be surprising, then, if they also needed longer to learn the BC and CB relations during 'testing'. So, all the groups may have been learning BC and CB but the 'language' subjects may have given the <u>impression</u> of passing the tests by learning faster than, and reaching criterion before, their 'non-language' counterparts. Furthermore, any differences in learning may have been enhanced by a number of procedural biases. At the beginning of each trial the experimenter pointed to the sample and said, 'Touch the one that goes with this one'. There could be little doubt that the 'language' children understood this instruction far better than the 'non-language' children. When the subjects asked for feedback during testing, the experimenter said, 'In this part of the game, I must be very quiet. I think you are doing a good job of working on this'. This, then, was a potential source of adventitious reinforcement which would only be available to the 'language' groups. Finally, the stimuli were line drawings of animal-like figures, each coloured a different hue. These stimuli may have encouraged differential naming in the 'language' groups, and there is ample evidence to show that differential responding can considerably enhance the acquisition of matching tasks (Urcuioli, 1985). Any of the factors listed above may have contributed to faster learning of the test relations by the language-able subjects.

In their paper, Devany et al speculated about the possible relationships between language and equivalence formation:

> "It could be that the ability to form equivalence classes is a unique and distinct skill that itself is required for stimuli to be used symbolically. Conversely, language may be a distinct skill that in turn permits the formation of equivalence classes. Finally, it is possible that both the formation of equivalence classes and the acquisition of language are the result of other common processes.
>
> Further analyses of the performances of very young developing children might help clarify this issue. If, for example, performance on an equivalence test is excellent before the child has acquired any labels, the argument that the ability to form equivalence classes is distinct (e.g., Sidman, 1986) and may itself lead to language acquisition would be strengthened. Similarly, if successful language training also establishes equivalence-class formation in retarded

> children, the effect of language on equivalence classes would be implicated. If the two areas are essentially synonymous or if they both reflect common behavioral properties (such as the ability to respond in terms of arbitrary relations per se (e.g., Hayes, 1986; Hayes & Brownstein, 1985), training in equivalence-class formation or its presumed underlying behavioral process should assist in language acquisition, and vice versa" (Devany et al, 1986, p.254).

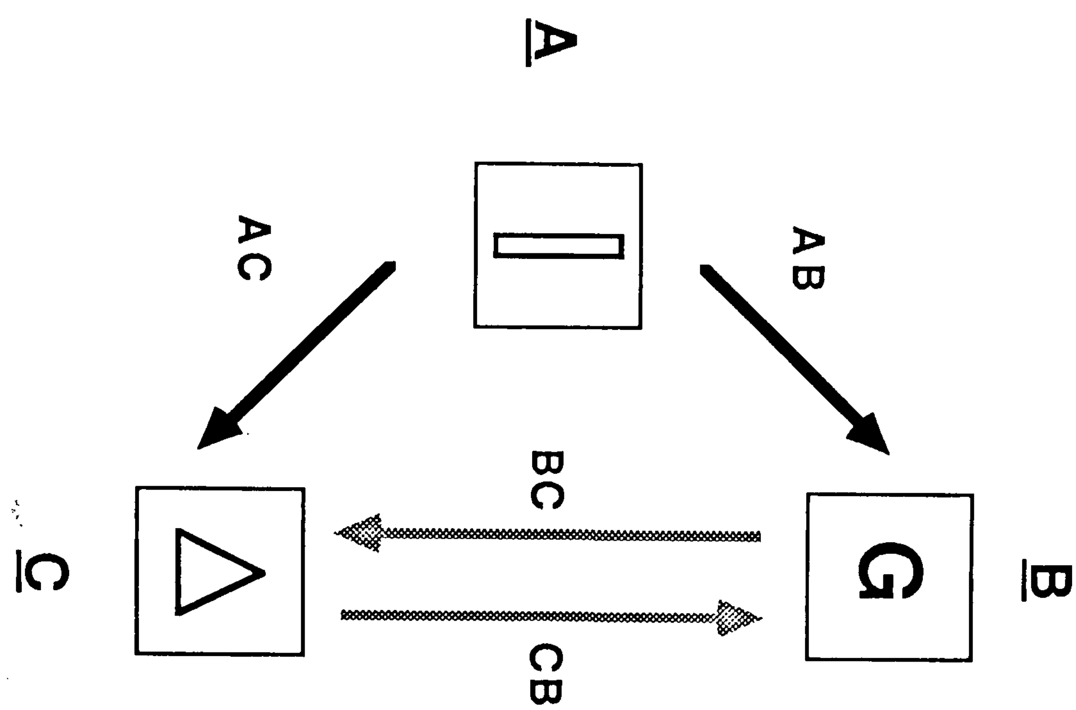
By the same reasoning, Beasty (1987) conducted a series of equivalence experiments with children in three age groups: 2-3, 3-4 and 4-5 year olds. Figure 1.9 depicts the paradigm he used. The stimuli were simple geometric shapes and colours, automatically presented on a 5-key response panel via a computer-controlled T.V. monitor. A screen isolated the subject from the experimenter to reduce the possibility of inadvertent cueing. In addition, all sessions were recorded on audio and videotape to capture any spontaneous verbal behaviour produced by the subject during the course of the experiment.

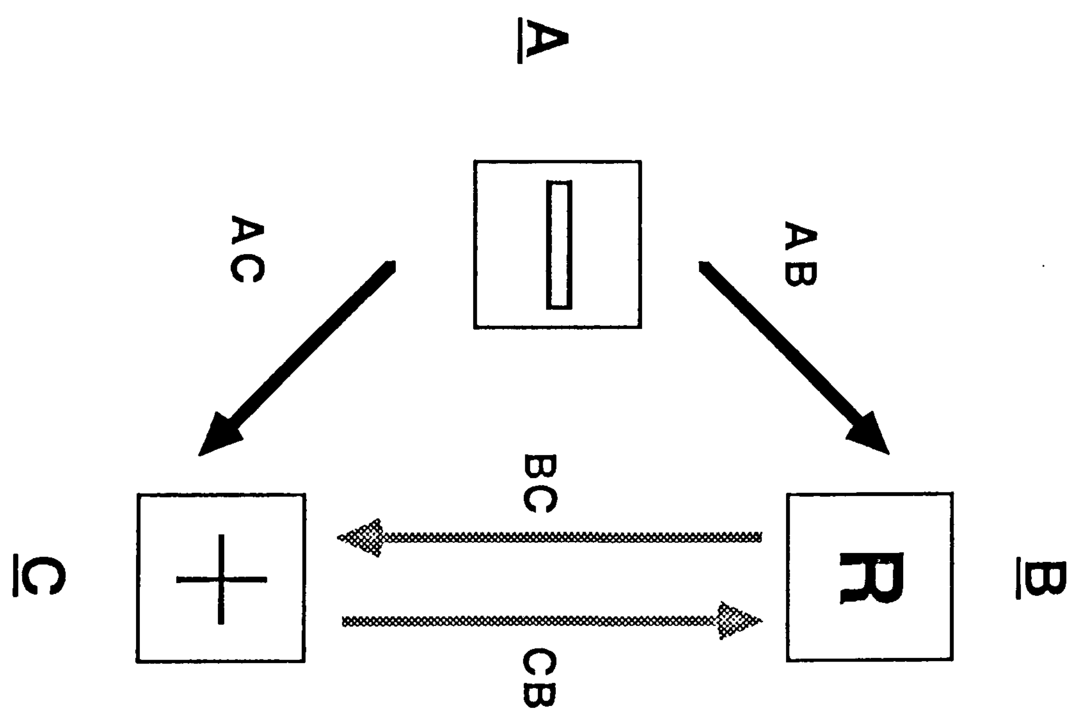
After the subjects had learned the AB and AC relations in Figure 1.9, the probability of reinforcement was gradually lowered to 0.2 so that only 1 in 5 correct responses was reinforced. Then, during test sessions, unreinforced test trials were interspersed among the sparsely reinforced AB and AC baseline trials. The first tests evaluated equivalence by presenting BC and CB trials. Further tests evaluated BA and CA symmetry. Each test lasted 4 sessions, and each session consisted of 24 baseline and 12 test trials.

The results indicated that equivalence has a developmental sequence. All ten 4-5 year olds passed the tests. In contrast half of the twelve 3 year olds and only one of the seven 2 year olds passed.

The results also suggested that equivalence is related to language. All of the subjects, including those who failed, were able

Figure 1.9 Equivalence paradigm adopted by Beasty (1987). Arrows point from samples to corresponding comparisons. Black arrows indicate trained relations and shaded arrows depict relations assessed during testing. Comparison always consisted of stimuli from the same set (A, B or C).


A
B
C
AB
AC
BC
CB
G


A
B
C
AB
AC
BC
CB
R

to name the stimuli spontaneously at some stage of the proceedings. So, naming per se was not sufficient for equivalence formation. It also seemed that common naming was absent; each subject appeared to give distinctive names to the stimuli. However, the manner in which the children spontaneously named the stimuli appeared to correlate well with their ability to pass the tests. The successful subjects often named the correct sample-comparison pairs in sequence during both baseline training and equivalence test trials. In contrast, sequential naming was absent from the repetoires of all of the subjects who failed equivalence tests.

Following the initial tests, the children who failed were taught the sequential naming routine which the others had used in training. The children were taught to name the correct sample-comparison pairs during baseline trials; they were required to say, for example, 'Up-Green' and 'Up-Triangle' for the AB and AC relations shown in the upper section of Figure 1.9, and 'Down-Red' and 'Down-Cross' for the AB and AC relations in the lower section. These were descriptions like those used spontaneously by the children who initially passed the tests.

The verbal intervention proved extremely effective. All of the subjects who learned sequential naming went on to pass the equivalence tests. One of the 2 year olds failed to learn sequential naming and he also failed equivalence, despite receiving repeated tests.

These data are significant for several reasons. First, they indicate that naming per se is not enough to bring about equivalence; the names are only effective if produced in an appropriate manner during the experiment (cf. Birge, 1941). Secondly, common naming may be one way in which stimuli can become equivalent, but it is not, as some have assumed, the only way (cf Lazar et al, 1984; Sidman et al, 1985). The studies by Beasty (1987) indicate that children may spontaneously employ other patterns of naming during arbitrary matching; sequential naming may be but one of many such patterns, all of which may be equally effective in promoting stimulus equivalence. Lowe and Beasty (1987) have speculated, in general terms, on why naming may be so effective:

> "It seems that animals, and non-verbal humans for that matter, are very much bound by the fixed-relations of the three-term contingency, e.g. 'In the presence of stimulus A a response to B will be reinforced' and this is why they fail equivalence. But when one names stimuli, and repeats the names to oneself, the relation between the names are freed of the spatial and temporal constraints that apply to the ordering of the stimuli in the environment" (Lowe and Beasty 1987, p.13.)

But, most importantly, the data presented above should be seen within the wider context of human operant behaviour. A considerable amount of experimental evidence has revealed major, and seemingly qualitative, differences between the performance of humans and animals on schedules of reinforcement (cf. Lowe, 1979; Matthews, Shimoff, Catania and Sagvolden, 1977). For example, on a Fixed-Interval (FI) reinforcement schedule, reinforcement is given for the first response after a fixed interval of time since the last reinforcer. Animal performance on FI schedules is characterised by a pause after reinforcement (the post-reinforcement pause or PRP) followed by a gradually accelerating response rate which ends when the next reinforcement is delivered (Branch and Gollub, 1974; Dews, 1978; Lowe and Harzem, 1977). This often produces a 'scalloped' pattern on cumulative records of responding. Furthermore, the overall response rate and the overall running rate (i.e. the response rate minus the PRP's) are declining functions of FI duration, whilst PRP's and succeeding inter-response times (IRT's) increase as the schedule value increases (Lowe and Harzem, 1977).

However, these orderly and replicable effects are not found when adult humans respond under conventional FI procedures. Human adult FI performance often takes one of two forms, either a high-rate pattern (a steady and high rate of responding throughout the interval) or a low-rate pattern (one or two responses at the end of the interval). These patterns, unlike those produced by animals, are often insensitive to changes in schedule value (Leander, Lippman and Meyer, 1968; Lowe, 1979; Weiner, 1969). Similar human-animal differences may also be found on Differential Reinforcement of Low Rate (DRL) and Fixed Ratio (FR) reinforcement schedules (Lowe, 1979). In addition, when adult humans are changed from one reinforcement schedule to another, they often show a 'rigidity' of performance that is uncharacteristic of animals (Bentall, Lowe and Beasty, 1983).

Lowe (1979; 1983) suggests that language may be the principal factor behind these performance differences:

> "Through participation in a verbal community humans acquire the skill of describing their environment and themselves, of formulating verbal rules, and of acting in accordance with these rules. This use of language is unique among living creatures and has profound effects upon much of human activity, including performance on schedules of reinforcement" (Bentall, Lowe and Beasty, 1985).

Lowe's hypothesis has gathered much support in recent years. In one study, Lowe (1979) noticed that when adult humans respond on FI schedules, they often reported counting out the interval between reinforcers. Those who produced a high-rate pattern seemed to be pressing the lever as they counted, whereas those who produced a low-rate pattern claimed that they pressed the lever only after counting out the interval. But when the subjects were supplied with a response-produced clock to attenuate counting, their lever presses came to resemble animal FI performance, both in terms of response patterning and sensitivity to schedule parameters.

Differences between human and animal performance are also found on more complex paradigms involving choice between two reinforcement schedules running concurrently. When animals are placed on these choice schedules their behaviour is so orderly that it can be readily predicted by mathematical equations. However, when adult humans are placed on such schedules, they often produce ideosyncratic and elaborate response sequences which appear to be determined by their verbal formulations of the contingencies (Lowe and Horne, 1985).

A number of studies have shown that the performance of pre-verbal infants on FI, FR and DRL schedules is indistinguishable from that of animals (see Bentall, Lowe and Beasty, 1983, 1985; Lowe, 1983; Lowe, Beasty and Bentall, 1983). Furthermore, there appears to be a developmental progression in FI performance. Children aged 5 years and older display the same high or low rate patterns as adult humans. However, children between 2 and 4 years of age, with less well developed verbal skills, produce patterns that are neither adult-like or animal-like but are seemingly intermediate, containing elements of both forms of responding. When the 2.5 - 4 year olds were taught to use their verbal behaviour in conjunction with lever pressing, their response patterns became the same as those produced by the older children and adults.

Taken together, this evidence suggests that humans can show conditioning effects like those observed with animals, but only during infancy, and before acquiring language. Childrens' behaviour alters radically as soon as they can articulate verbal descriptions of the contingencies. The data on human schedule performance is consistent with a growing body of literature on the role of language in the regulation of behaviour (cf. Bem, 1967; Luria, 1961; Risley, 1977; Vygotsky, 1962), and with Skinner's formulation of rule-governed behaviour and consciousness in humans (Skinner, 1974). The research reviewed above on naming and stimulus equivalence provides yet further support for the view that verbal behaviour plays a critical role in human development.

The view that naming is necessary for the emergence of equivalence raises a number of interesting questions. One question is best illustrated in conjunction with a recent study by Clarke, Remington and Light (1986). The subjects, three severely retarded children, were presented with two sets of stimuli. The two sets consisted of pictures which the subjects either could or could not choose conditional upon picture names spoken by the experimenter (we shall refer to these pictures as 'known' vs 'unknown' respectively). The subjects learned manual signs for the pictures, which were shown to them and named by the experimenter.

The children learned signs for the 'known' pictures much faster than for the 'unknown' pictures. This result suggested that the relations between 'known' pictures and signs were facilitated by the corresponding picture names spoken by the experimenter. Indeed, this is what one might expect. The 'known' picture names, spoken by the experimenter, already functioned to control each subject's choice of (and therefore attention to) the 'known' pictures. So perhaps the subjects also attended to the 'known' pictures as the experimenter named them during sign training. In contrast, the 'unknown' picture names, spoken by the experimenter, would not have acquired this 'attention cueing' function with respect to the unknown pictures. So perhaps the signs were learned faster for the 'known' pictures because the subjects attended to them better than the 'unknown' pictures. Alternatively, the 'known' picture names may have facilitated signing simply because those names already controlled the subjects' arbitrary matching performances in other contexts (i.e. name-picture matching).

The second finding involved the emergence of untrained relations in the subjects' repetoires. After the subjects had learned to produce signs conditional upon pictures, they were able to do the reverse - select pictures conditional upon signs. This suggested that the pictures and signs had become equivalent.

This finding poses certain questions. If naming is necessary for equivalence, then how do names become equivalent to their corresponding stimuli? One cannot argue that the names must be named because this may send one into an infinite regress of naming. Perhaps there is something fundamental about naming which allows us to escape this dilemma. Chapter 4 of this thesis presents new experimental data of relevance to this issue. Chapter 4 also addresses other questions. In the experiments by Beasty (1987), why was sequential naming so effective in promoting stimulus equivalence? Could common naming have the same dramatic effect on equivalence formation as sequential naming? Perhaps most importantly, how may we define naming?

But we will start where the present research programme began, with two equivalence experiments with animals. The first of these was unique, and therefore especially exciting, because it was designed to test the following possibility: if language is critical for equivalence formation then perhaps animals may pass equivalence tests after receiving extensive language training. The next chapter presents Experiment 1, which involved the first reported test of equivalence with language-trained chimpanzees.

CHAPTER 2

A SEARCH FOR STIMULUS EQUIVALENCE IN THE MATCHING-TO-SAMPLE PERFORMANCES OF LANGUAGE TRAINED CHIMPANZEES

1. Method
2. Results and Discussion
3. General Discussion

EXPERIMENT 1

METHOD

SUBJECTS

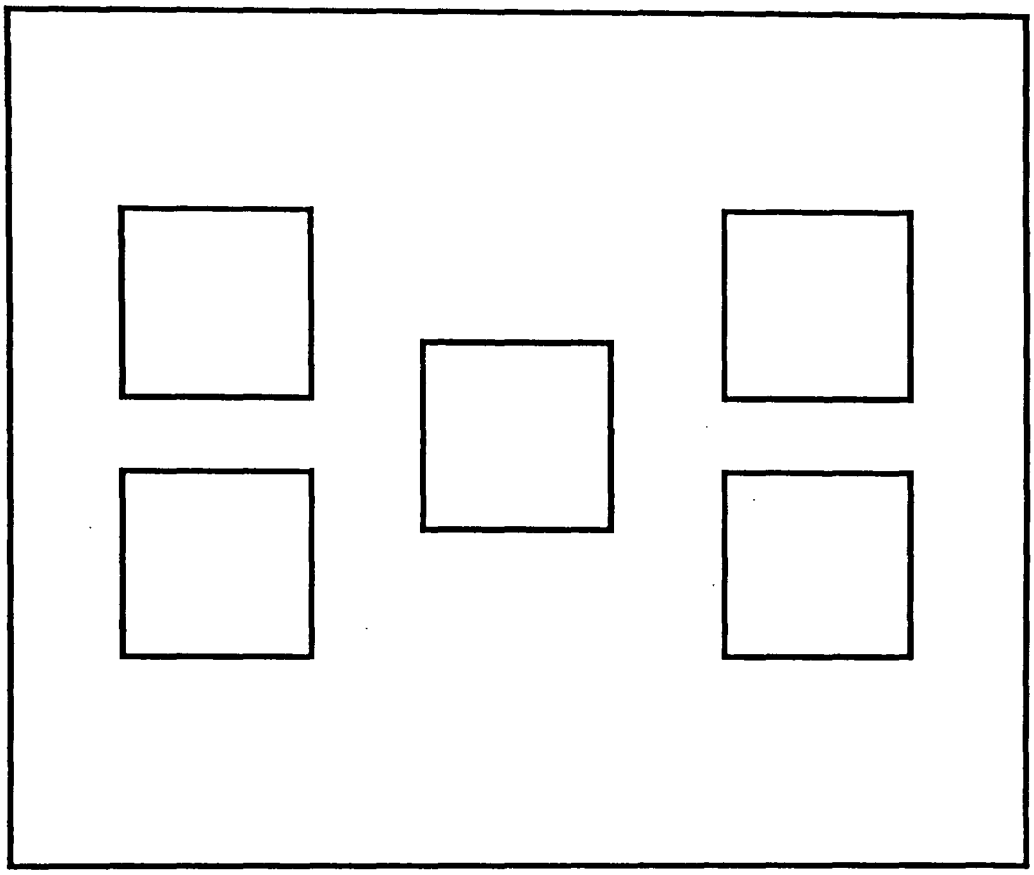
The subjects were three adult chimpanzees (Pan troglodytes), two males (Sherman and Austin) and one female (Lana). At the start of the experiment, Sherman and Austin were 13 and 12 years old respectively, and Lana was 16. All three chimps were subjects in an ape-language training programme at the Language Research Center in Atlanta, Georgia, U. S. A. The chimps had learned to communicate to others by pointing to lexigrams, which are visual-graphic stimuli arranged on a keyboard, each associated with an object, action or location. Details of their language training have been described elsewhere by their principal caretakers, Professors Duane Rumbaugh and Sue Savage-Rumbaugh (see Rumbaugh, 1977; Savage-Rumbaugh, 1986). It should be noted that the chimps did not have access to a lexigram keyboard during any time within the period in which the present experiment was conducted.

None of the chimps were food deprived in the general sense. However, highly preferred foods (e.g. candies, exotic fruits, yoghurt etc.) were reserved as reinforcers, and the chimps seldom had access to these at any time other than during experimental sessions.

APPARATUS

The experimental chamber was the middle room of the chimps' living quarters. A wall at one end of the chamber contained a five key stimulus-response panel (see Figure 2.1). Each key was 5cm square and

Figure 2.1 Schematic representation of the five key stimulus-response panel (see text for dimensions).



made of transparent perspex. Four of the keys were located at the corners of a 21 x 12 cm rectangle, with the fifth key in the rectangle's centre. The panel was mounted with the centre of the middle key at 76 cm above the floor, so that when the chimps were seated, their eyes would be approximately level with the middle key.

A colour monitor screen was placed directly behind the stimulus-response panel. The monitor was connected to an Apple microcomputer which programmed the sequencing and display of the stimuli, and recorded relevant key presses. The stimuli were two colours (red and green, each drawn as a 4 cm square) and four shapes (a 'Y',a 'zig-zag', a triangle and a cross, each drawn white on a black background to occupy a 4 cm square area). These stimuli were presented directly behind the keys, on the monitor screen.

A food chute was placed directly beneath the panel, approximately 30 cm from the floor. A variety of foods were dispensed down the chute via an automatic dispenser controlled by the Apple microcomputer.

PROCEDURE

Each chimp usually received at least one session per day for five days a week. Prior to beginning each experimental session, disturbances were minimised as much as possible by placing the subject alone in the experimental chamber and by locking the chamber doors.

During training sessions correct key presses produced a high-pitch tone from the computer and the delivery of food, whereas incorrect key presses produced a low-pitch tone and no food. After a correct or incorrect choice all stimuli were removed (i.e. the screen went blank) and a five second inter-trial interval followed, at the end of which the next stimulus appeared. The procedure was non-correction throughout; errors did not cause trials to be repeated.

Preliminary Training

Each subject began with one 48-trial session of preliminary training, which consisted of two stages. In the first stage, the chimps were taught to press the keys. On each trial, either a cross or a triangle appeared at random on any of the four outer keys. Pressing the lit key was designated correct and pressing any dark key was incorrect. When the subject had learned to press the lit key, and only the lit key, then the next stage commenced. Trials began with either the cross or the triangle appearing equally often, and at random, on the centre key. Pressing the centre key then produced an identical shape on any of the four outer keys, again at random. The centre stimulus remained on. Pressing the lit outer key was correct. In this and all subsequent sessions, pressing a dark key had no scheduled consequence. Once the subject was reliably pressing the lit centre key followed by the lit outer key, identity matching trials were presented (see below).

Identity Matching

In this stage, the triangle and cross appeared equally often as samples (centre key stimuli). When the sample was pressed it remained on and was joined by two comparisons (outer key stimuli). On each trial, the comparisons were the triangle and the cross, and their appearance was accompanied by an audible 'beep' from the computer. When the triangle was the sample, the triangle comparison was correct, and when the cross was the sample, the cross comparison was correct. The triangle and cross appeared equally often as samples in each 48-trial session.

In this and all subsequent matching-to-sample sessions, all trial-types (sample-comparison combinations) were presented at random with the following restrictions:

(a) No more than three trials with the same trial-type could occur consecutively.

(b) All four comparison keys had to be scheduled as correct before any could be correct again.

With these exceptions, all trial-types and correct keys appeared equally often in each session.

The subjects were presented with triangle and cross identity matching trials until they had reached a criterion of 90% correct responses per session. Following this, the Y-shape and zig-zag were introduced and the subjects were taught to identity match with these 'novel' stimuli. On each trial, the Y-shape and zig-zag appeared as comparisons and each shape appeared equally often as the sample. Sessions were continued until the 90% criterion was reached.

All animals finally had sessions in which the 'feedback' on each trial (i.e. the programmed consequence for a correct or incorrect response) was gradually reduced from a probability of 1.0 (feedback on every trial) to 0.2 (feedback, on average, every fifth trial). This procedure was a preparatory step toward subsequent testing (see below). For Sherman and Austin, sessions before testing consisted of 12 trials each of the triangle, cross, Y-shape and zig-zag identity matching trials. Lana, however, only received triangle and plus identity trials during this stage. All subjects proceeded to the next stage after maintaining criterion performance on the shapematching tasks at the 0.2 probability level.

Reflexivity Testing

During reflexivity test sessions, novel colourmatching trials were interspersed among the baseline shapematching trials. On colourmatching test trials the comparisons were the red hue and the green hue, and the sample appeared equally often as red or green. For Sherman and Austin, each 48-trial test session consisted of 8 trials each of the four shapematching baseline trial-types and the two colourmatching test trial-types. In each of her first five test sessions, Lana received 32 baseline shapematching trials with the triangle and cross (16 trials each) and 16 colourmatching test trials with the red and green hues (8 trials each). Thereafter, Lana's test sessions were the same as Sherman's and Austin's. On all test sessions, feedback on baseline trials was delivered according to the 0.2 probability schedule. Test trials, however, were unreinforced (i.e. correct and incorrect responses produced the inter-trial interval only, and neither food nor tones were delivered). At the end of testing, further identity matching sessions were presented, but this time with feedback on every trial.

AB Training (arbitrary matching)

The subjects were next presented with AB arbitrary matching trials. The Y-shape and zig-zag were Set-A samples and the green hue and red hue were Set-B comparisons. When the sample was a Y-shape, reinforcers were contingent upon choosing the green comparison, and when the zig-zag was the sample reinforcers were contingent upon choosing red (see Figure 2.2). These two trial-types appeared equally often in each 48-trial session.

In addition, a number of intervention procedures attempted to

Figure 2.2 Stimulus relations presented during AB training trials. Arrows point from sample stimuli (only one presented at a time) to corresponding comparison stimuli.

TRAIN AB

<u>A</u> SAMPLE		<u>B</u> CORRECT COMPARISON
Y	$\longrightarrow$	G
	$\longrightarrow$	R

accelerate learning of the AB task. Each 'intervention' is described below (further details (e.g. number of AB trials per intervention) appear in the Results section).

(a) Enlarging the baseline All of the chimps developed strong comparison preferences during early training sessions; they tended to choose one particular comparison colour on every trial. These comparison preferences may have arisen from, and may have been maintained by, the matching-to-sample contingencies on AB trials. Each AB session consisted of equal numbers of trials in which the red comparison and green comparison were scheduled as correct, but this did not prevent localised parts of a session from comprising of a greater proportion of trials with the same correct comparison. A subject may, for example, learn to repeatedly select the red comparison after being exposed to a number of neighbouring trials in which the red comparison is mostly correct. If the same contingencies were sufficiently recurrent then the preference, once learned, might be maintained. Although these possibilities were not subjected to a detailed analysis, they seemed likely enough to warrant 'evasive' action. If the baseline is expanded to include additional trial-types (i.e. additional comparison stimuli) then this necessarily reduces the probability of getting localised concentrations of trials with the same correct comparison. Consequently, a number of sessions were run, each consisting of an equal proportion of the two AB trial-types and two additional arbitrary matching trial-types (the latter consisted of Set-A samples and triangle and cross comparisons; sample-correct comparison combinations were Y-shape - triangle and zig-zag - cross). Enlarging the baseline did not, however, promote acquisition of the AB task.

(b) Interspersed identity matching trials It was possible that the chimps were not attending sufficiently to the stimuli on AB trials, and that this may have been why they were not acquiring the relation. Consequently, the chimps were given a number of sessions in which the AB trial-types were interspersed among identity matching trials with the Set-A and Set-B stimuli. The chimps were all able to identity match with the Set-A shapes and Set-B colours, thus demonstrating that they were at least attending to the same stimuli which also appeared on AB trials. Despite this, their performances on AB remained at or around chance level.

(c) Compound stimulus presentations Sherman received trials in which each Set-A sample shape was coloured the same as its corresponding Set-B comparison (i.e. the Y-shape was coloured green, and the zig-zag was coloured red). After 9 such sessions, Sherman was making no errors. Obviously, his correct choices may have been governed by the colour of the sample (he had already learned to match the colours), but it was hoped that this control would transfer, in an incidental fashion, to the shape of the sample. However, as soon as the colours were removed from the samples (i.e. as soon as normal AB trials were presented), Sherman's performance fell to chance level.

(d) Fading Intervention (c) may have been successful if the colour had been gradually faded from the sample shape, but this was technically very difficult to achieve with the Apple microcomputer. Nevertheless, a different fading programme was eventually devised in order to capitalise on the chimpanzees' pre-existing colourmatching skills. Each sample shape was initially drawn on a background colour which matched the colour of the corresponding comparison. So the Y sample was drawn on a green background, and this was to be matched to the green comparison, whereas the zig-zag sample was drawn on a red background, and this was to be matched to the red comparison. At the start of the session, the sample background colour-cue was identical in size and shape to each comparison colour (i.e. 4cm square). As the session progressed, correct responses produced a gradual reduction in the size of the sample colour-cue, whereas incorrect responses increased it in size (up to a maximum of 4 cm square). For each trial-type, six consecutive correct responses resulted in the temporary disappearance of the colour-cue from the sample shape. The fading therefore proceeded from colourmatching to AB arbitrary matching, and fading was implemented to a degree determined entirely by the subject's performance.

Sherman and Lana eventually learned AB matching via the fading procedure (see Results). Austin, however, did not; he therefore received a number of additional interventions, described below.

(e) Interspersed zero-delay identity matching trials Here, sessions were the same as in (b) above, except that all identity matching trials were zero-delay i.e. when the sample was pressed it disappeared and was followed immediately by the presentation of the comparisons. The zero-delay procedure was meant to encourage Austin to attend more to the Set-A and Set-B stimuli. However, it did not affect his AB performance, which remained at or around chance level.

(f) Delayed cueing In AB sessions involving delayed cueing, the incorrect comparison was removed immediately after it was presented. This forced the subject to select the correct comparison. Then, the time between comparison onset and removal of the incorrect comparison was gradually increased across trials. This contingency permitted two possible outcomes; either the subject made a choice prior to the removal of the incorrect comparison (and thus the choice could be correct or incorrect) or the subject waited for the incorrect comparison to disappear, and then made a correct choice by default. Other studies (e.g. McDonagh, McIlvane and Stoddard, 1985; McIlvane, Withstandley and Stoddard, 1985) have succeeded in teaching matching-to-sample via similar delayed cueing techniques. However, delayed cueing did not improve Austin's AB performance. On the majority of trials he waited for the incorrect comparison to disappear before responding, and on the few trials in which he did not wait, he responded equally often to the correct and incorrect comparison.

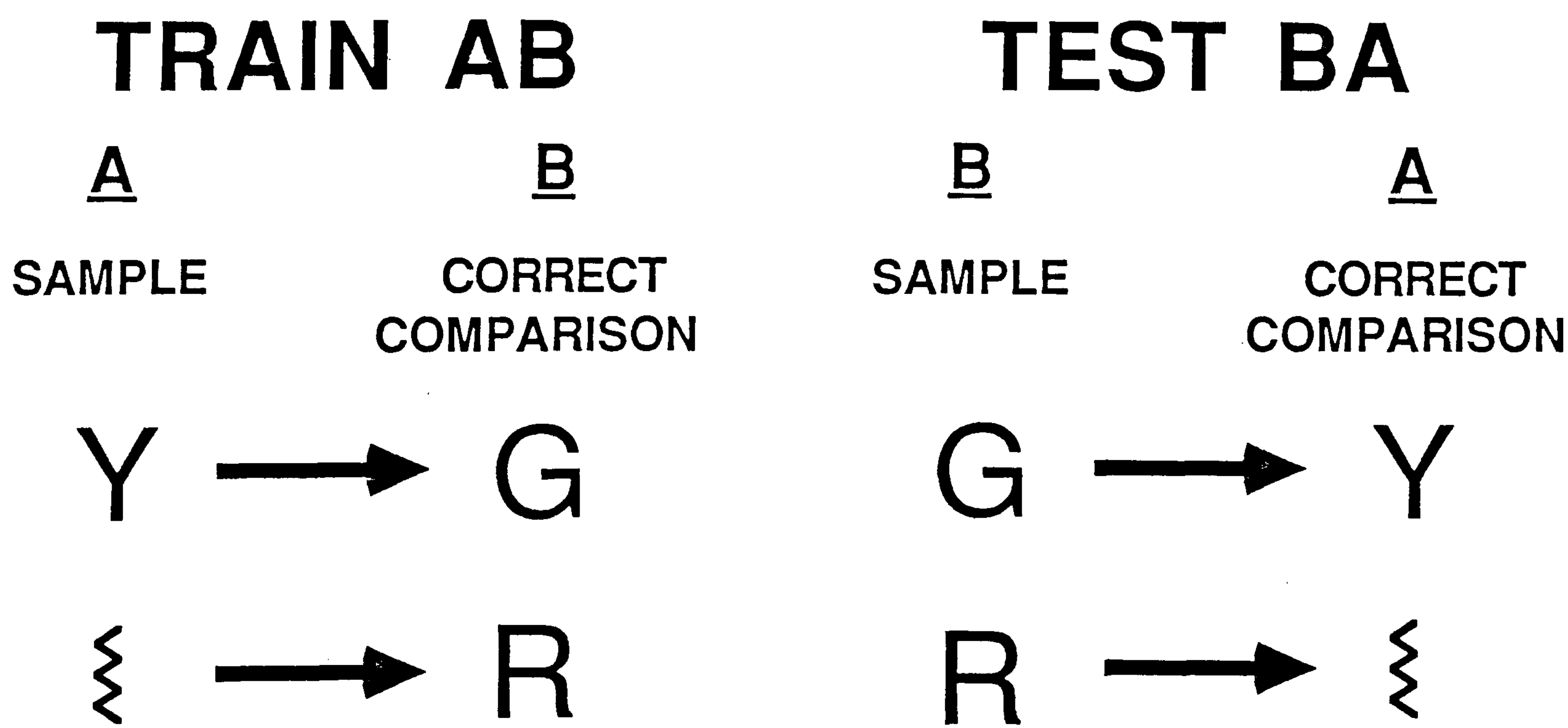
(g) Differential sample schedules In order to produce the comparisons, Austin was required to press the zig-zag sample five times or to wait three seconds after the Y sample appeared before pressing it (presses before three seconds reset the interval). Teaching the subject to respond differentially to the samples has often resulted in faster acquisition of matching tasks (e.g. Cohen, Looney, Brady and Aucella, 1976; Sidman et al, 1982; Urcuioli, 1985). However, sample schedules were not effective with Austin; in fact, he often managed to satisfy the schedule contingencies even when he was clearly not looking at the sample. On these occasions, he began by pressing the sample repeatedly; if this did not produce the comparisons (and the audible beep which accompanied their presentation) then he stopped responding, paused for longer than three seconds, and then resumed pressing. This tactic guaranteed production of the comparisons irrespective of the sample presented.

Summary Lack of time precluded further attempts to teach Austin AB matching. Sherman and Lana, however, learned the AB relation, so they received additional sessions in which the reinforcement probability was gradually lowered from 1.0 to 0.2 in preparation for the subsequent test phase (see below). During probability reduction, each AB session still began with sample colour-cues, even though both chimps were now capable of choosing correct comparisons without these cues. After reaching the 0.2 level, additional sessions were presented, all of which began with no sample colour-cue (although the cue re-appeared on the first trial following an incorrect choice). Finally, AB sessions were run at the 0.2 level with the colour-cues totally absent.

Symmetry Testing

Symmetry test sessions always included AB baseline trials and BA symmetry test trials (see Figure 2.3). On symmetry test trials, the stimuli which were formerly samples and comparisons were interchanged i.e. a green or red sample was presented with Y and zig-zag comparisons. To pass the test, the chimps would have to select the Y-shape when the sample was green, and the zig-zag when the sample was red, although they had never been explicitly trained to do so. Other trial-types were also included in the baseline at various stages of testing, and details of these are presented in the Results section, along with other relevant information regarding the number of trials per trial-type, and the reinforcement contingencies on baseline and test trials.

Figure 2.3 Stimulus relations presented during symmetry test sessions (see text).


TRAIN AB
A
B
SAMPLE
CORRECT COMPARISON
Y
G
R
TEST BA
B
A
SAMPLE
CORRECT COMPARISON
G
Y
R

RESULTS AND DISCUSSION

Identity Matching

Table 2.1 shows the number of trials each subject required in order to reach criterion on the first two identity matching tasks. All three chimps learned each identity matching task as quickly as (and often faster than) the children from the experiments by Beasty (1987) (and these children were in fact trained on colourmatching tasks which generally take less time to learn than shapematching tasks like those presented to the chimps; see Carter and Werner, 1978). Sherman took approximately twice as long to learn the second task than the first, whereas Austin and Lana needed exactly the same number of trials to learn each task. So, although each chimp had initially learned to identity match with one set of shapes (the triangle and cross), their matching performances did not transfer immediately to a novel set of shapes (the Y and zig-zag). These results, however, do not necessarily indicate an absence of reflexivity in the chimpanzees' identity matching performances. Many of the children in Beasty's experiments matched some novel shapes to criterion but failed to match others, in reinforced trials which followed colourmatching training (see Beasty, 1987). Perhaps then, it would be unwise to draw firm conclusions from a subject's performance on a single novel identity matching task.

Following acquisition, the probability of reinforcement on the baseline shapematching trials was gradually lowered to 0.2 in preparation for testing (see Procedure). This reduction was accomplished in three sessions with Sherman, and in five sessions with both Austin and Lana. All three chimps maintained their shapematching performances during this time.

Table 2.1 The number of trials each subject required to reach criterion on the first two identity matching tasks of Experiment 1 (**Δ** = triangle, + = cross, Y = Y-shape, Z = zig-zag).

SUBJECT	IDENTITY MATCHING TASK	
	I (Δ, +)	II (Y, Z)
AUSTIN	144	144
LANA	432	432
SHERMAN	288	528

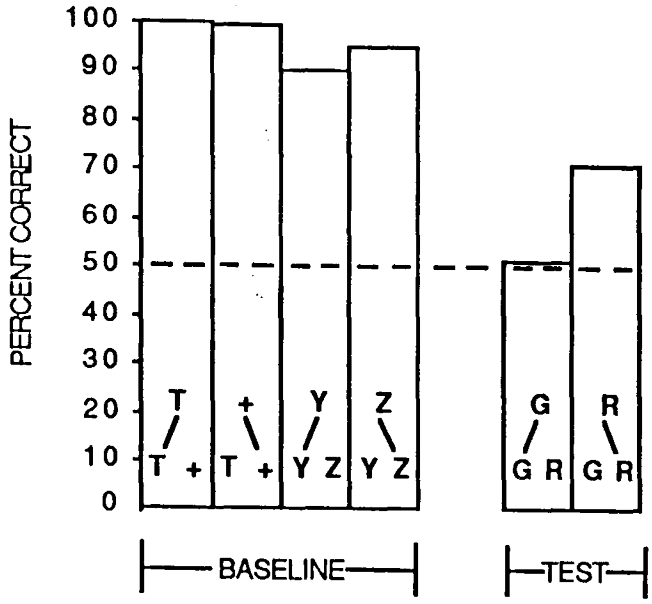
Reflexivity Testing

Figure 2.4 shows each subject's overall performance on reflexivity tests. The figure depicts the percentage of correct responses for each trial-type, averaged over the test sessions. The stimuli for each trial-type are identified at the bottom of the bars. Sample stimuli are placed above comparisons, and a line connects each sample to its corresponding comparison. The bars to the left show that the subjects' performances on the sparsely reinforced baseline shapematching trials were at or around 90% correct during testing. Despite this, their overall scores on the unreinforced colourmatching test trials were relatively poor, typically at or around chance level or (50% correct), which is depicted by the dotted line. Figures 2.5 - 2.7 show each chimp's scores on unreinforced colourmatching test trials for each individual test session. Sherman responded below chance on the red-red trial-type for the first six test sessions, but on the 7th and 9th sessions accuracies on both colourmatching trial-types rose above 80% correct (see Figure 2.5). A similar pattern emerged in Austin's test sessions (Figure 2.6). Austin's test trial scores reached a peak of around 80% correct on his 6th and 7th test session. In Lana's first five test sessions, her baseline did not include identity matching trials with the Y-shape and zig-zag, and her scores on the unreinforced colourmatching trials typically remained at or below chance level (see Figure 2.7). Then, from the 6th test session onwards, the Y-shape and zig-zag identity trial-types were added to the baseline, and one session later (session number 7), her colourmatching performance peaked above 80% correct, declining thereafter.

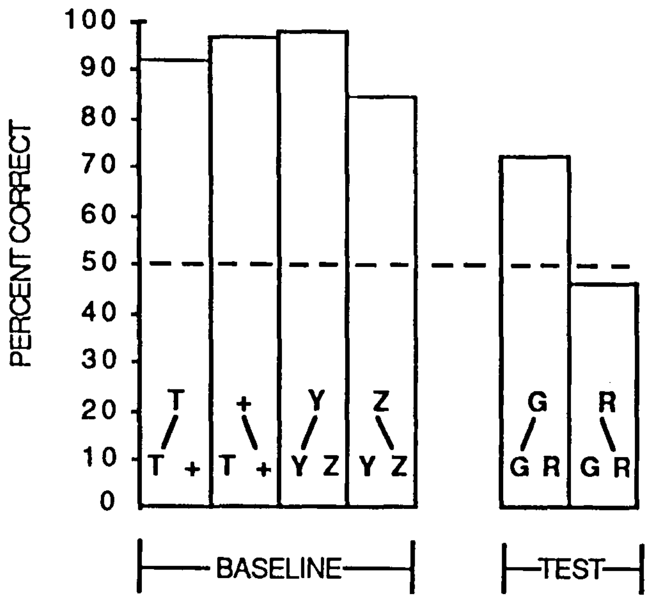
At the end of testing, the subjects received an additional identity matching session which differed from the preceding sessions

Figure 2.4 Overall results (percentage of correct responses) produced by the three chimpanzees during their reflexivity tests (T = triangle, + = cross, Y = Y-shape, Z = zig-zag, G = green, R = red; see text for further details).

LANA



AUSTIN



SHERMAN

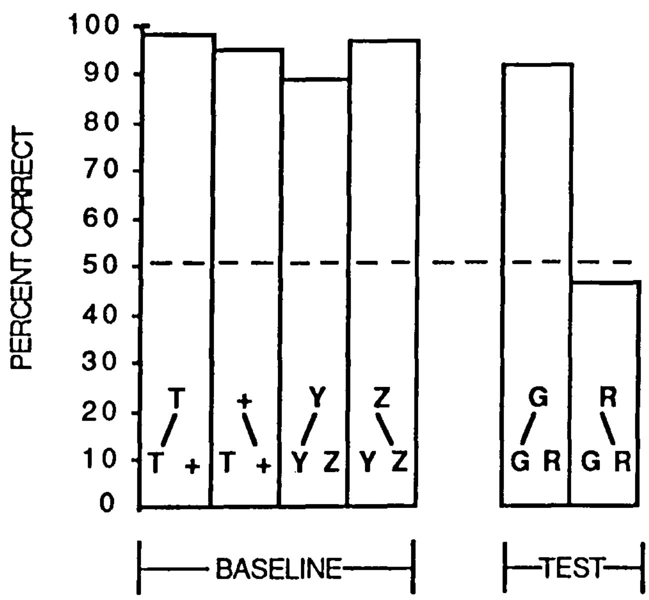


Figure 2.5 Sherman's performance on unreinforced colourmatching test trials during each individual reflexivity test session.

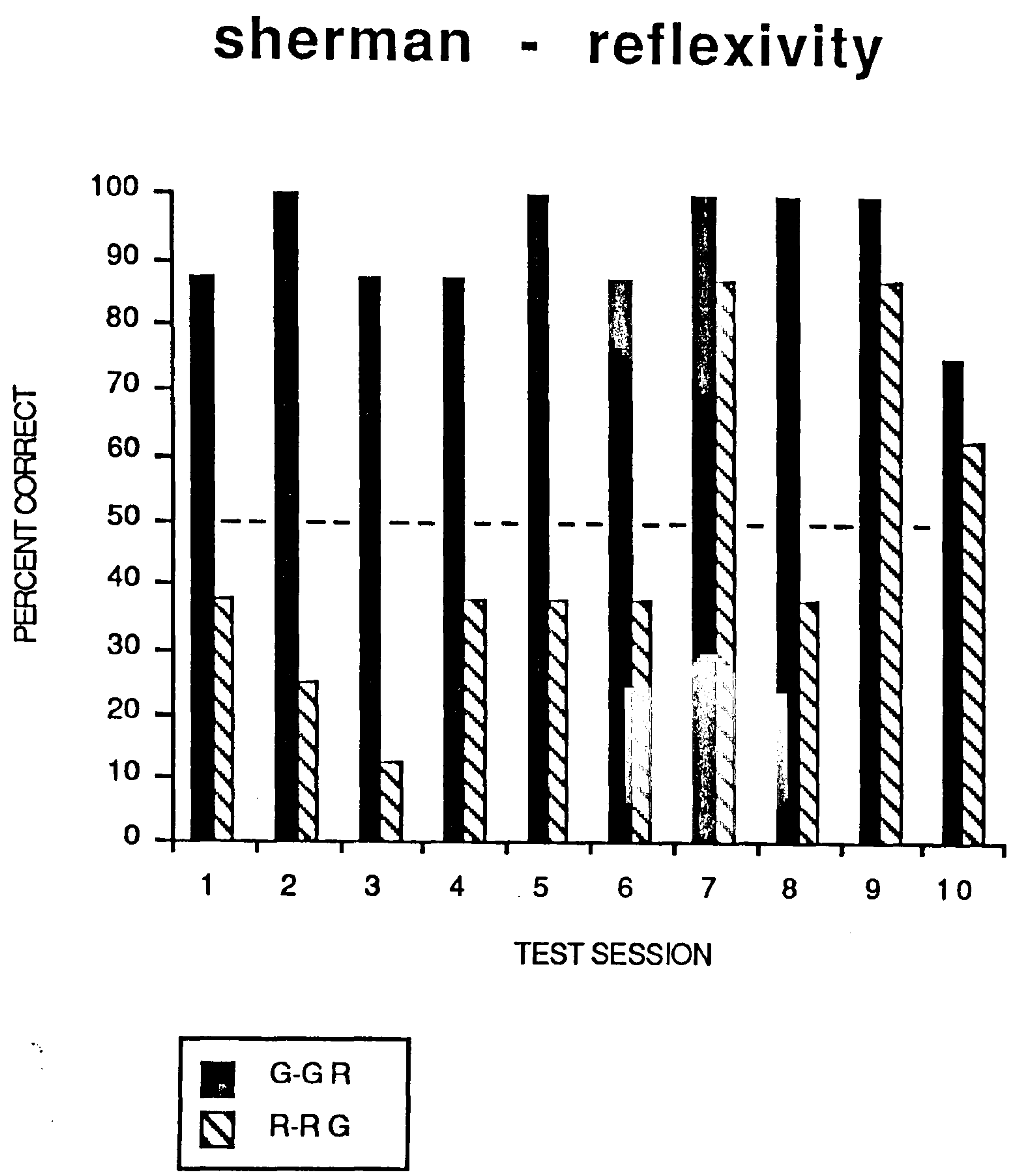

sherman - reflexivity
PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
TEST SESSION
G-G R
R-R G

Figure 2.6 Austin's performance on unreinforced colourmatching test trials during each individual reflexivity test session.

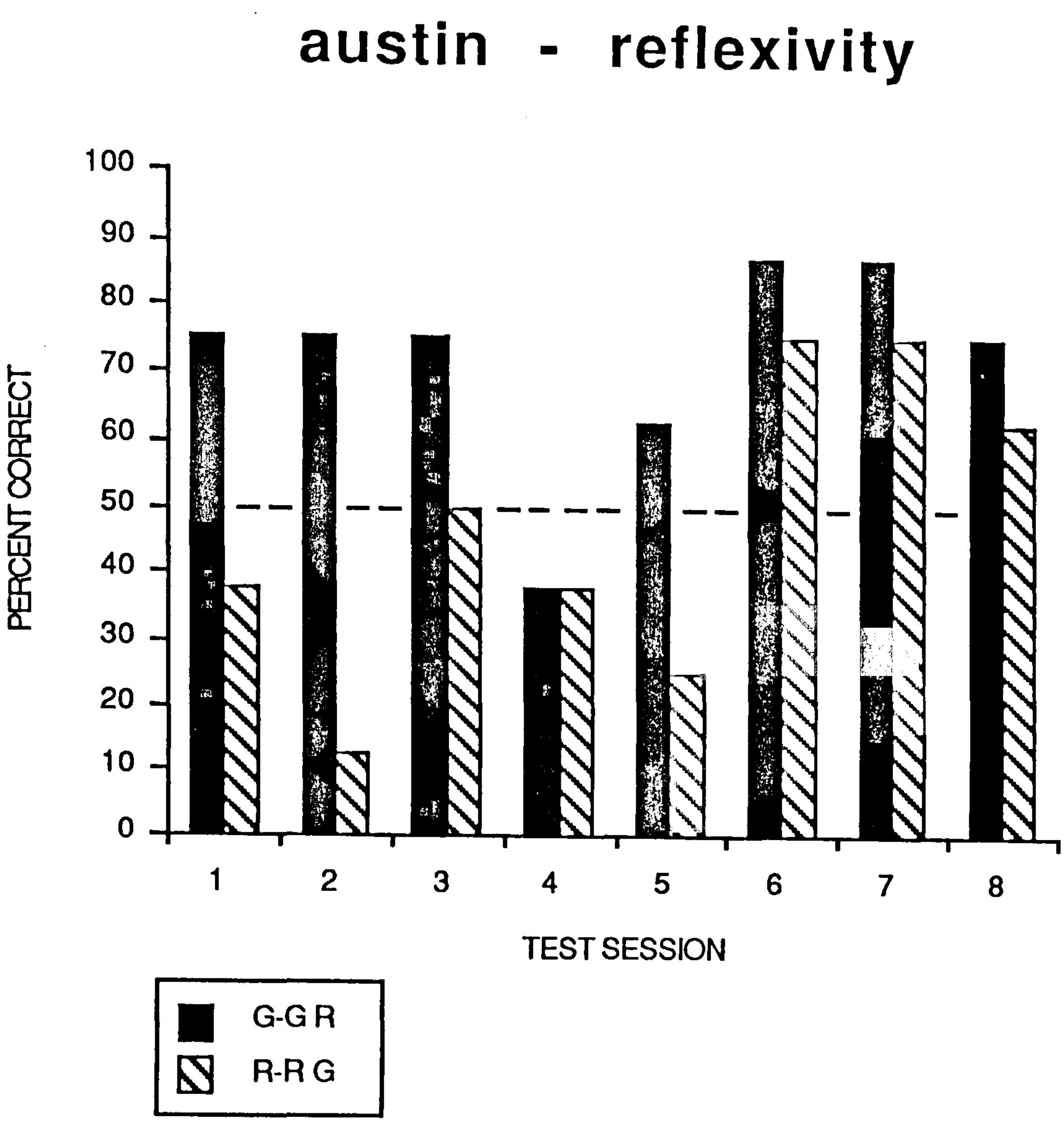
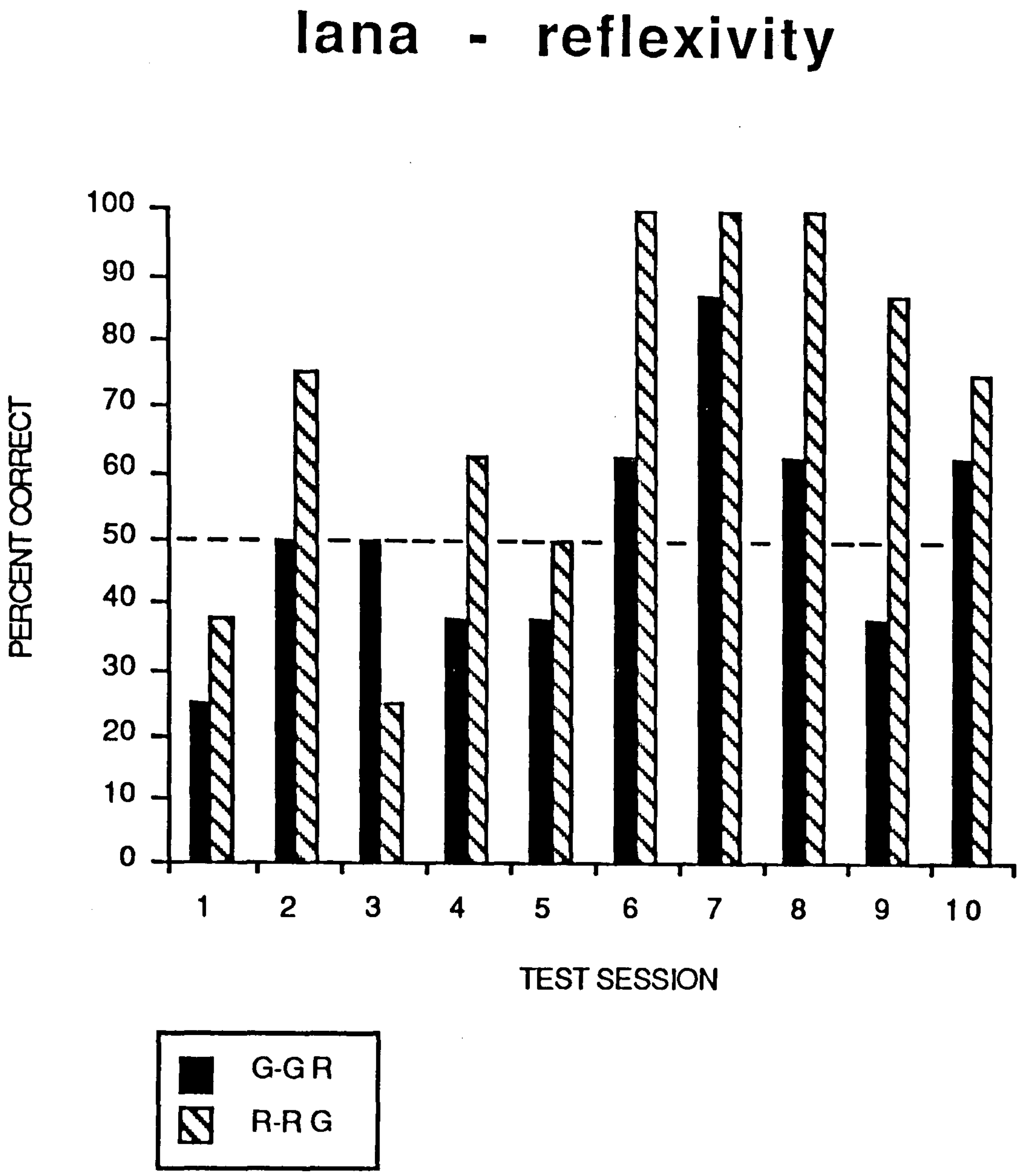

austin - reflexivity
PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
TEST SESSION
G-G R
R-R G

Figure 2.7 Lana's performance on unreinforced colourmatching test trials during each individual reflexivity test session.


lana - reflexivity
PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
TEST SESSION
G-G R
R-R G

only in that reinforcers were available for correct responses on all identity matching trials. The chimps made only one error each in 48 reinforced trials.

There are two reasons why these results are difficult to interpret. Firstly, the results do not necessarily indicate an absence of reflexivity in the chimpanzees' matching performances. All three chimps failed the tests overall, but their scores on unreinforced test trials did reach a peak toward the latter stages of testing. This may be important given that novel relations do not always emerge immediately in unreinforced test sessions; sometimes several such sessions are required before a subject begins to respond correctly to the test trials (see Devany, Hayes and Nelson, 1986; Sidman, Kirk and Willson-Morris, 1985; Spradlin, Cotter and Baxley, 1973).

However, it is important to note that while these results do not necessarily indicate an absence of reflexivity, they do not necessarily indicate its <u>presence</u> either. Reflexivity, or generalised identity matching, is only demonstrated when a subject is able to match completely novel stimuli (i.e. stimuli the subject has not already learned to pair via identity matching trials). But, just prior to testing, it became obvious that these chimps were already able to identity match not just with the red and green stimuli used in the present experiment, but with a whole host of colours. The chimps demonstrated this ability not by pressing keys on a panel but by moving a cursor, under the control of a joystick, to screen positions occupied by a sample colour and its identical comparison. This task was devised by others not involved with the present experiment, and was occasionally offered to the chimps as a 'time filling' task (N.B. although the chimps were presented with other such tasks, none of these tasks used the same stimuli as in the present experiment).

Although the experiment had thus far provided no substantial evidence for or against equivalence in these language-trained chimpanzees, it had at least ensured that they could identity match with the same stimuli, and in the same experimental context, as would be used in later testing. As one shall see, these identity matching skills became particularly significant in the subsequent phases of the experiment.

AB Training

Table 2.2 lists the number of trials each chimpanzee received on each intervention procedure during AB training. Sherman and Lana eventually reached criterion via the fading programme. The same fading procedure, however, did not work with Austin, despite the fact that he received approximately three times more fading trials then Lana, and about four times more than Sherman. In fact, Austin never reached criterion on AB, even after more than 6,000 AB trials spread over a number of intervention procedures. He took no further part in the experiment. Sherman and Lana, however, were next taken through a series of stages in preparation for symmetry testing. These stages are outlined in Table 2.3, along with the number of trials per stage. Lana received 8 more fading sessions (384 trials) in which the reinforcement probability was gradually lowered from 1.0 to 0.2. This was followed by two further fading sessions at the 0.2 probability level, each beginning with no sample colour-cue (see Procedure). Finally, Lana received two standard AB sessions at the 0.2 level. At no stage did Lana's AB performance deteriorate, so she therefore proceeded to symmetry testing. Sherman went through the same three preparatory stages as Lana, but his AB performance deteriorated in the final stage. In two additional sessions, Sherman's AB performance was reinstated by

Table 2.2 Number of trials per AB intervention for each subject in Experiment 1. Interventions are listed in the order in which they appeared in the experiment. The 'normal' category refers to standard AB matching trials; all other categories are explained in the procedure section. Each asterisk denotes the point at which a chimp achieved criterion on the AB task (Austin never reached criterion).

SUBJECT	INTERVENTION	NUMBER OF AB TRIALS PER INTERVENTION
SHERMAN	Normal	480
	Compound stimuli	432
	Normal	768
	Enlarged baseline	144
	Interspersed identity trials	96
	Fading	1104 *
	TOTAL	3024
LANA	Normal	192
	Enlarged baseline	192
	Interspersed identity trials	96
	Fading	1440 *
	TOTAL	1920
AUSTIN	Normal	768
	Enlarged baseline	240
	Interspersed identity trials	144
	Fading	3963
	Normal	528
	Delayed cueing	240
	Zero-delay identity trials	192
	Sample schedules	336
	TOTAL	6384

Table 2.3 The number of trials required by Lana and Sherman in each stage between AB acquisition and symmetry testing (see text for further detail).

SUBJECT	NUMBER OF TRIALS PER PHASE				TOTAL
	I	II	III	IV	
LANA	384	96	96	-	576
SHERMAN	288	240	96	96	720

KEY

CATEGORY	SESSION TYPE	REINFORCEMENT PROBABILITY
I	AB with fading (sessions begin with maximum colour-cue)	Gradually reduced from 1.0 to 0.2
II	AB with fading (sessions begin with no colour-cue)	0.2
III	AB (normal)	0.2
IV	AB (normal)	1.0

changing the reinforcement probability to 1.0, so that every correct response was reinforced. Sherman then proceeded directly to symmetry testing because, by now, Lana had received several test sessions, the results of which indicated that it would not be necessary to reduce the reinforcement probability on Sherman's AB baseline trials.

Symmetry Testing

Lana had three distinct test phases. In her first test phase, correct responses on BA symmetry test trials were not reinforced. These unreinforced BA test trials were interspersed amongst the sparsely reinforced AB baseline trials. Since test trials were not reinforced, the reinforcement probability on AB baseline trials was increased sufficiently to maintain the overall probability at 0.2. Each test session consisted of 32 AB baseline trials (16 per trial-type) and 16 BA symmetry test trials (8 per trial-type). Lana was given 12 test sessions, her overall performance on which is represented in Figure 2.8. Lana's AB baseline performance was around 90% correct but, despite this, her performance on the critical symmetry test trials was at or around 50% correct, or chance level. Figure 2.9 depicts Lana's performance on each individual test session, and shows that her BA scores remained at chance level for all 12 test sessions.

So, in her first test phase, Lana had apparently failed the symmetry test. She gave no evidence that her training had established equivalence between samples and corresponding comparisons.

Now it could be argued that Lana failed the symmetry test for reasons other than a lack of symmetry in her baseline relations. There are two justifications for this argument. Firstly, Lana's correct responses on test trials were never reinforced, whereas those on

Figure 2.8 Lana's overall performance during her first symmetry test phase (symmetry test sessions 1-12). The two left-hand bars depict the AB baseline trial-types and the two right-hand bars depict the unreinforced BA symmetry test trial-types. Each baseline bar represents 192 trials and each test bar represents 96 trials.

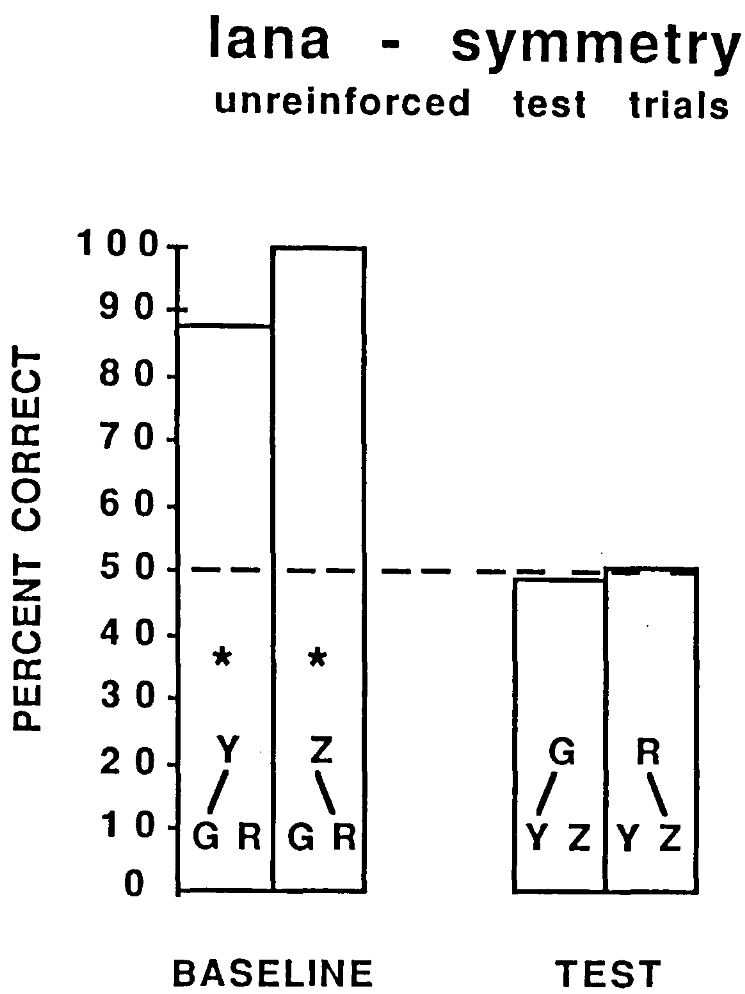
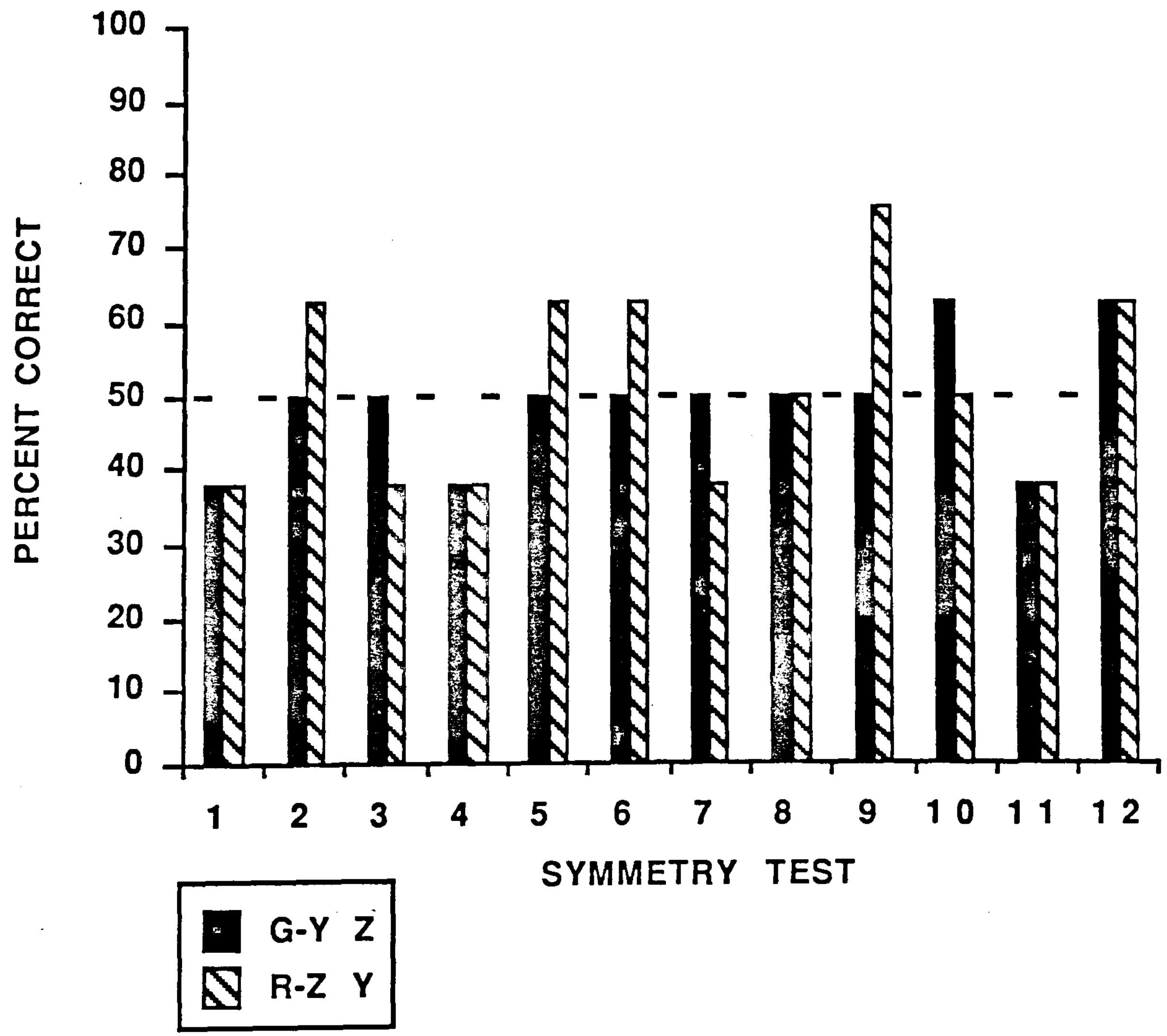

lana - symmetry
unreinforced test trials
PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
*
Y
G R
*
Z
G R
G
Y Z
R
Y Z
BASELINE
TEST

Figure 2.9 Lana's performance on the unreinforced BA symmetry trial-types during each session of her first symmetry test phase. Each bar represents 8 trials.

lana - symmetry

unreinforced test trials



baseline trials were reinforced, albeit only occasionally. So maybe she had learned not to attend to the stimuli on symmetry test trials. Secondly, on symmetry test trials Lana was presented with the colours as samples and the shapes as comparisons for the very first time. Perhaps this novelty may have somehow disrupted her test performance. So two changes were made before Lana's second test phase. First, reinforcement became available for all correct responses not only on baseline but also on test trials. Secondly, identity matching trials were added to the baseline. Prior to the resumption of testing, Lana was required to match not only Set-A samples to Set-B comparisons but also each Set-B colour and each Set-A shape to itself. Two 72-trial sessions were presented, each consisting of 24 trials each of the AB trial-types and 6 trials each of the identity matching trial-types. The identity trials ensured that Lana had experienced sample colours and comparison shapes before they were presented on BA symmetry test trials. Lana made only one error on the identity matching trials, and her overall score on AB was 92.7% correct.

Subsequent test sessions (test numbers 13-20) consisted of 12 trials each of the AB baseline trial-types, 6 trials each of the identity matching trial-types and 6 trials each of the BA symmetry test trial-types. Lana's performance, averaged over all 8 test sessions, is depicted in Figure 2.10. The left-hand group of bars once again depict her baseline performance. The asterisks denote the AB trial-types which were tested for symmetry and the other four baseline bars denote the additional identity trial-types. All the baselines were above 90% correct. In contrast, Lana's performance on one of the symmetry test trial-types was still at chance level, namely the one with green as a sample. Figure 2.11 shows her BA scores for each individual test session. As testing progressed, Lana generally became more likely to choose the zig-zag comparison on each test trial. In the last two

Figure 2.10 Lana's overall performance during her second symmetry test phase (symmetry test sessions 13-20). The bars containing asterisks depict her performance on the AB baseline trial-types, and the other four baseline bars represent her scores on the identity matching trial-types which were added to the baseline. All correct baseline and test trials were reinforced. The bars with asterisks represent 96 trials each, and all other bars represent 48 trials each.

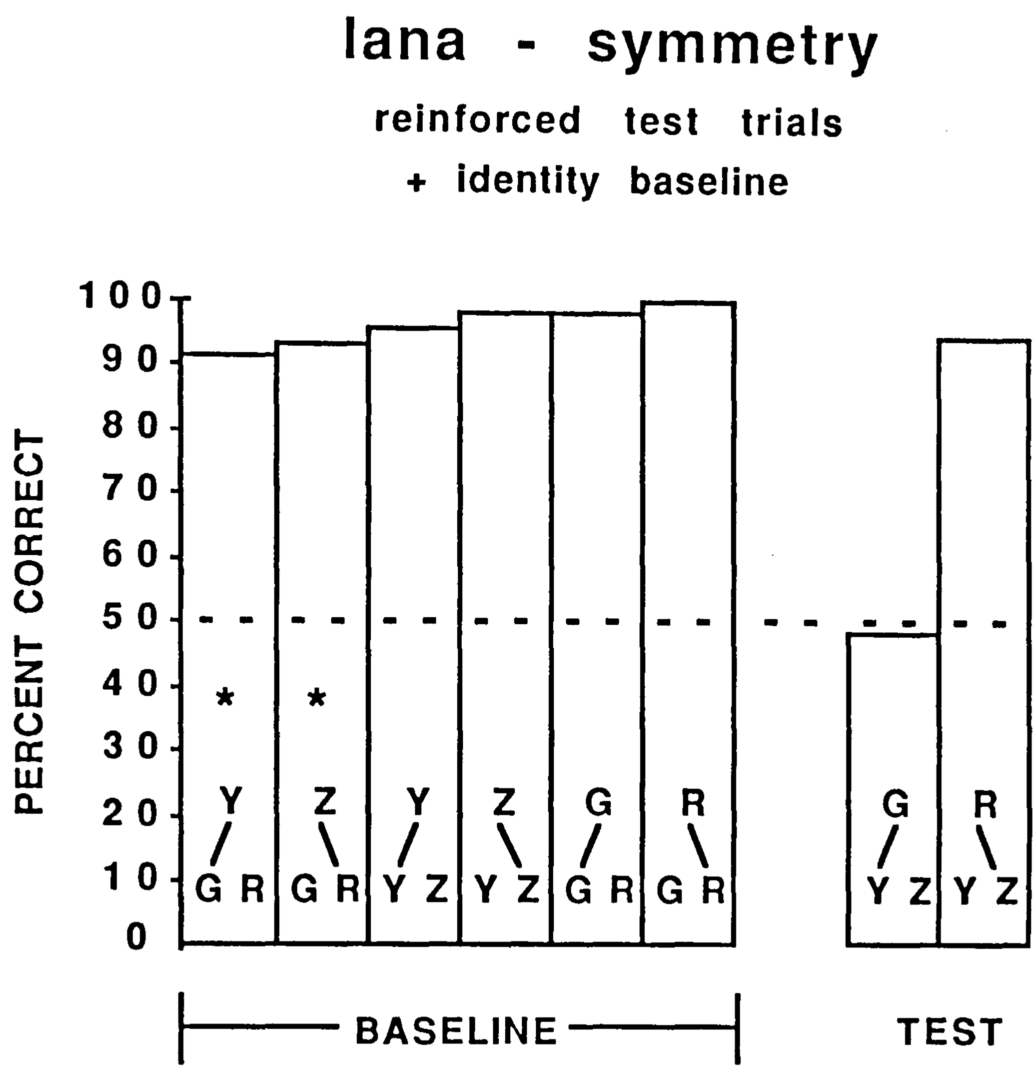
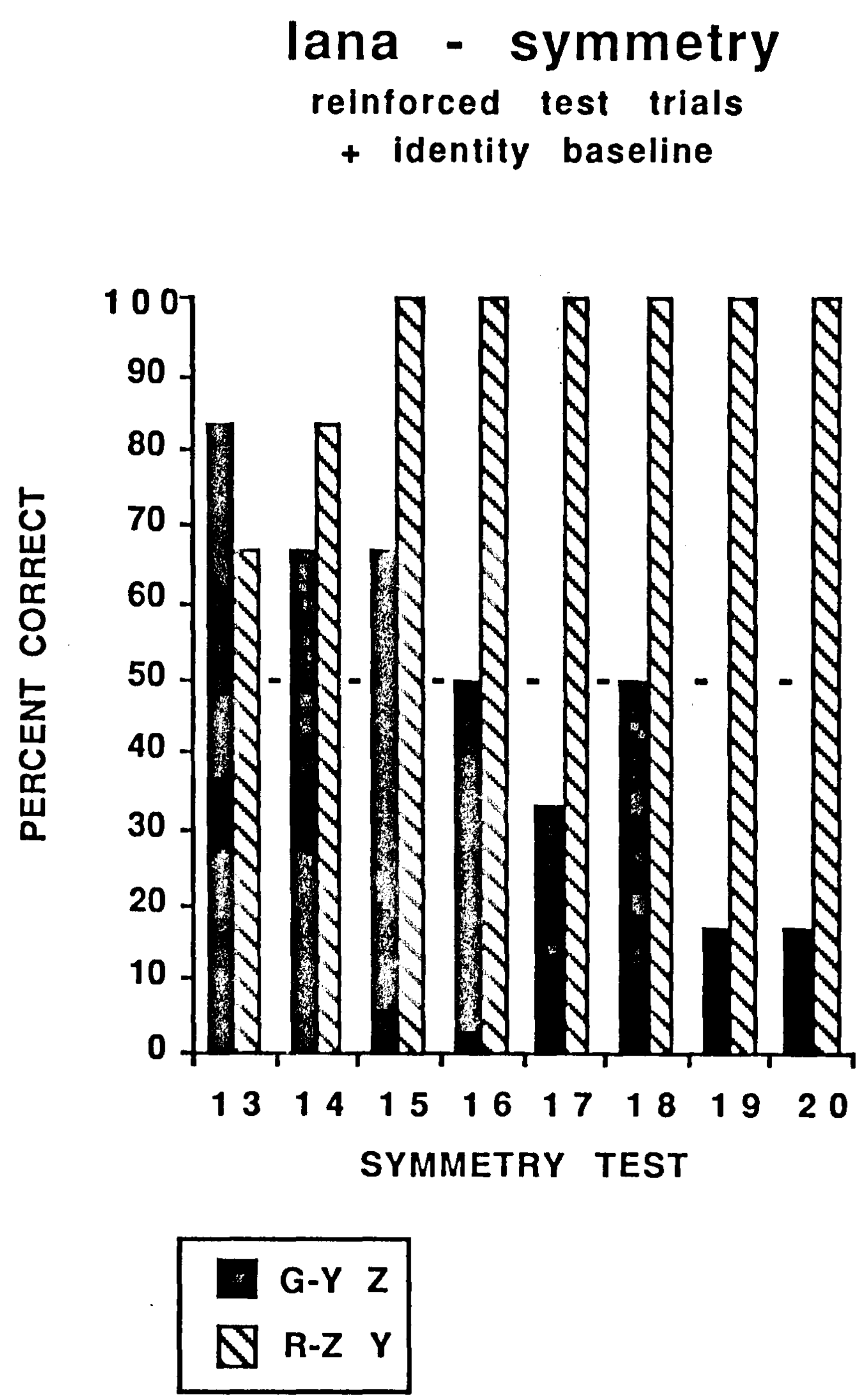

lana - symmetry
reinforced test trials
+ identity baseline
PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
*
*
Y
/
G R
Z
\
G R
Y
/
Y Z
Z
\
Y Z
G
/
G R
R
\
G R
G
/
Y Z
R
\
Y Z
BASELINE
TEST

Figure 2.11 Lana's performance on the reinforced BA symmetry trial-types during each session of her second symmetry test phase. Each bar represents 6 trials.


lana - symmetry
reinforced test trials
+ identity baseline
PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
13 14 15 16 17 18 19 20
SYMMETRY TEST
G-Y Z
R-Z Y

sessions (19 and 20), Lana chose the zig-zag on all but two of the 24 BA trials. So, Lana had failed the symmetry test even with reinforced test trials and identity matching controls.

Nevertheless, to be absolutely sure, Lana received further symmetry test sessions in a third and final test phase. Once again the baseline was modified; this final modification was dictated by a combination of factors, described below.

On symmetry test trials, samples were red or green, and comparisons were the Y-shape and zig-zag. So, among the pre-requisites for symmetry were a successive discrimination between red and green samples and a simultaneous discrimination between Y and zig-zag comparisons. If these basic discriminations were absent from Lana's repetoire then she would be unable to pass the symmetry test, even if her training had established stimulus equivalence. In Lana's first two test phases, the procedures were such that they did not guarantee the presence of these pre-requisite discriminations. It thus became important to re-examine Lana's response patterns on BA test trials, because this might help to determine whether the discriminations were in place.

In her first series of tests (test numbers 1-12), Lana's performance was at chance level on the two BA trial-types (see Figures 2.8 and 2.9). This pattern unfortunately tells us nothing about the status of the pre-requisite discriminations; the same pattern could be produced equally easily with or without discriminations between sample colours and/or comparison shapes. Lana's response patterns from her second test phase (test numbers 13-20) were, however, much more informative. Figure 2.10 shows that Lana was behaving discriminatively to the test trial stimuli; when the sample was red she almost invariably chose the zig-zag comparison, and when the sample was green comparison shape equally often. This pattern strongly indicates the presence of the pre-requisite discriminations for symmetry. It is highly unlikely that such a pattern could be produced without a successive discrimination between red and green samples and a simultaneous discrimination between Y and zig-zag comparisons.

It is worth emphasising that although one can safely infer the presence of the pre-requisite discriminations from Lana's response patterns during her second test phase, one cannot do so from her concomitant baseline identity matching performance. For example, Lana could match each colour to itself on baseline identity trials, but this only required a simultaneous discrimination between the colours. On baseline identity trials, the red or green sample could be viewed together with the red and green comparisons i.e. green and red could be compared directly. However, this could not be done on symmetry test trials, because green and red were never presented together; they always appeared successively as samples. For this reason, the colourmatching task did not necessarily require the subject to learn the particular successive discrimination required on symmetry test trials. Similarly, although Lana could match each shape to itself on baseline identity trials, this performance need not necessarily involve the same discriminations as are required on symmetry test trials. On identity matching trials, the subject could respond correctly by discriminating between the shape on the sample key and the shape on each comparison key. This discrimination would not be available on symmetry test trials; rather, the test trials required discriminations between shapes located only on the comparison keys.

In summary, thus far no steps had been taken procedurally to establish the presence of the pre-requisite discriminations for symmetry. However, a retrospective examination of Lana's test trial response patterns showed that these pre-requisites were at least intact during her second test phase. Nevertheless, it is not a good tactic to rely on response patterns as the <u>only</u> evidence for pre-requisite discriminations. What if the response pattern changes? Some patterns, as we have seen, tell us nothing about the status of sample or comparison discriminations. For example, an exclusive preference for one of the comparisons may tell us that the subject can discriminate between the comparisons, but it tells us nothing about discrimination of the samples. Figure 2.11 shows that, in her second test phase, Lana's preference for the zig-zag comparison was incrementally strengthening across symmetry test sessions. If this trend continued, one could no longer be confident that all of the pre-requisite discriminations were in place. Consequently, Lana's final test series included baseline trials to provide an independent assessment of the pre-requisite discriminations for symmetry.

This time the baseline included <u>zero-delay</u> identity matching trials. Up until this point, all trials had involved simultaneous matching; when the sample was pressed it stayed on and was accompanied by the comparisons. But on zero-delay trials the sample disappears when it is pressed, and then the comparisons appear. When the subject chooses between the comparisons, the sample is no longer present. So, on zero-delay identity matching trials, the subject is no longer able to respond correctly by directly comparing samples with comparisons. Zero-delay identity trials therefore demand the same discriminations as symmetry test trials. If Lana responds correctly on the zero-delay colourmatching trials then we would know she was still discriminating between temporally successive instances of green and red stimuli (the samples). The same successive discrimination between red and green samples is one of the pre-requisites for correct responding on symmetry test trials. And if Lana responds correctly on the zero-delay n we would know she was still discriminating between the Y-shape and zig-zag when they could only be seen simultaneously as comparisons. The same simultaneous discrimination between Y and zig-zag comparisons is the other pre-requisite for symmetry.

So, Lana's final test series included zero-delay identity trials, to provide an ongoing assessment of the pre-requisite discriminations for symmetry. These final test sessions (15 in all) were identical in composition to the last series, except for the zero-delay on identity trials; on all other trials, the sample remained on when it was pressed. Once again, all correct responses were reinforced. Figure 2.12 shows that Lana scored well above chance on each baseline trial-type. Her good performance on the zero-delay identity trials showed that the pre-requisite discriminations for symmetry were still intact. Despite this, Lana's symmetry test performance remained virtually the same as before; overall, the green-Y BA trial-type was still only marginally above chance level. Furthermore, the test trial averages depicted in Figure 2.12 were generally representative of Lana's performance in each individual session. As Figure 2.13 shows, Lana produced the same pattern of chance level performance on green - Y test trials in the majority of her test sessions. Lana, in fact, never achieved the within-session criterion of 90% correct on each test trial-type. By the end of testing, her scores on each BA trial-type had risen to 83.3% correct, and there is little doubt that they would have risen further with additional reinforced test trials. However, this of course cannot be taken as evidence of symmetry; we have no reason to appeal to symmetry if the performance involved can readily be shown to occur due to simpler processes of stimulus control. It seemed that Lana was responding correctly to the BA trials not because the Set-A and Set-B stimuli were equivalent, but because she had gradually

Figure 2.12 Lana's overall performance during her third and final symmetry test phase (symmetry test sessions 21-35). The bars containing asterisks represent 180 trials each. All other bars represent 90 trials each.

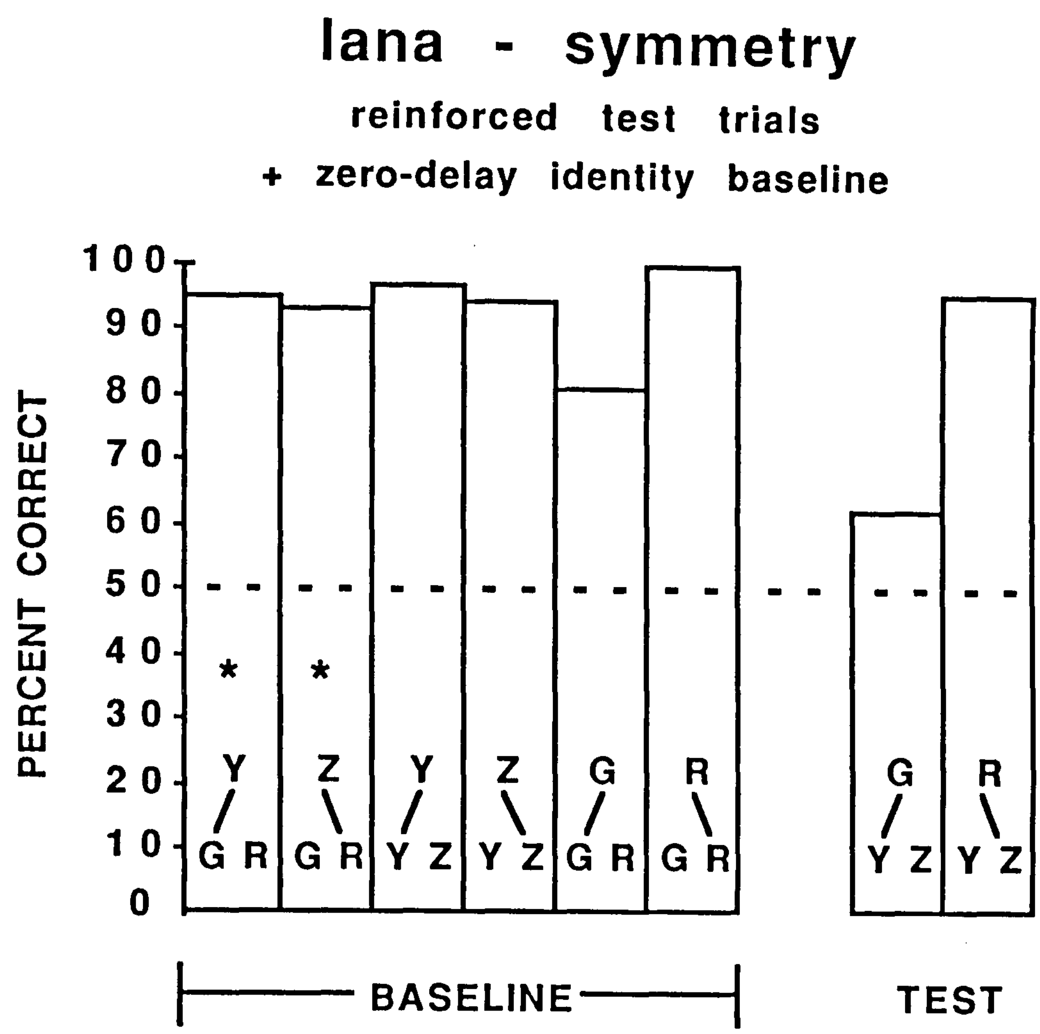
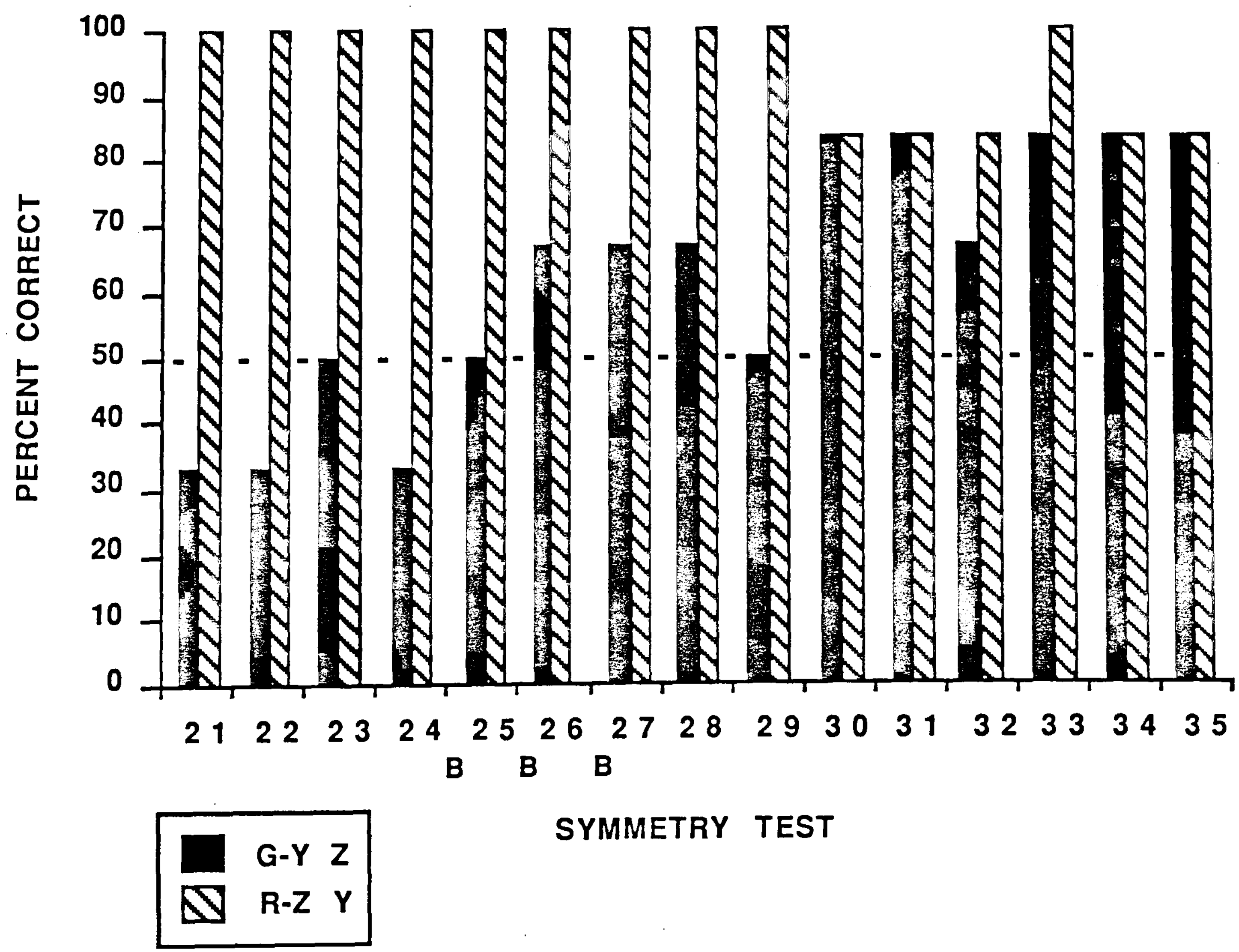

lana - symmetry
reinforced test trials
+ zero-delay identity baseline
PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
*
*
Y / G R
Z \ G R
Y / Y Z
Z \ Y Z
G / G R
R \ G R
G / Y Z
R \ Y Z
BASELINE
TEST

Figure 2.13 Lana's performance on the reinforced BA symmetry trial-types during each session of her third and final symmetry test phase. Each bar represents 6 trials. The capital-B's on the horizontal axis depict points at which baseline sessions intervened between symmetry test sessions (see text).

lana - symmetry

reinforced test trials
+ zero-delay identity baseline



learned to do so simply as a consequence of the differential reinforcement on test trials.

One should note that the test sessions in Lana's final series were not all presented consecutively. The capital - B's on the horizontal axis of Figure 2.13 depict points in which additional baseline sessions intervened between symmetry test sessions. The interruptions to testing provided additional evidence against symmetry, as will be revealed later. For the time being we may conclude that Lana had given no evidence of symmetry despite receiving 35 test sessions and a grand total of 468 test trials.

Now we come to Sherman. Sherman received 12 test sessions, all of which were the same as the last series presented to Lana. All correct responses were reinforced. Figures 2.14 and 2.15 show that Sherman also failed the symmetry test. Although his baseline scores were well above chance level, his performance on the critical symmetry test trials remained at or near chance level throughout. Sherman, just like Lana, had failed the test, despite also being exposed to conditions which maximised his chances of success.

Once again, the capital B's in Figure 2.15 denote the points at which symmetry tests were interrupted with the presentation of baseline trials only. These additional baseline sessions were administered in order to counteract performance decrements which appeared on particular baseline trial-types during the course of the symmetry tests. Baseline deterioration occurred not only in Sherman's test sessions but also in Lana's final test series.

Figure 2.16 once again shows Lana's scores from her final test series, and Sherman's scores from his single test series. Figure 2.16 is actually a conglomerate of Figures 2.12 and 2.14 shown earlier, except that this time the shading highlights those baseline trial-types

Figure 2.14 Sherman's overall performance during his 12 symmetry test sessions. The bars containing asterisks represent 144 trials each; all other bars represent 72 trials each.

sherman - symmetry

reinforced test trials

+ zero-delay identity baseline

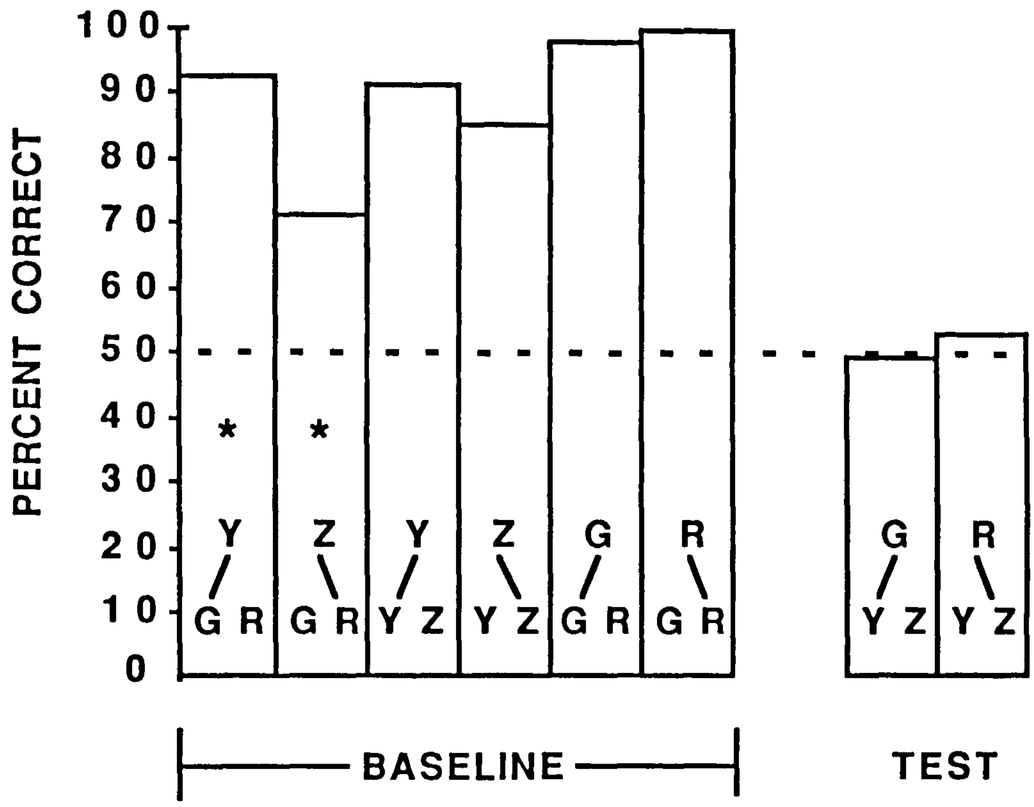
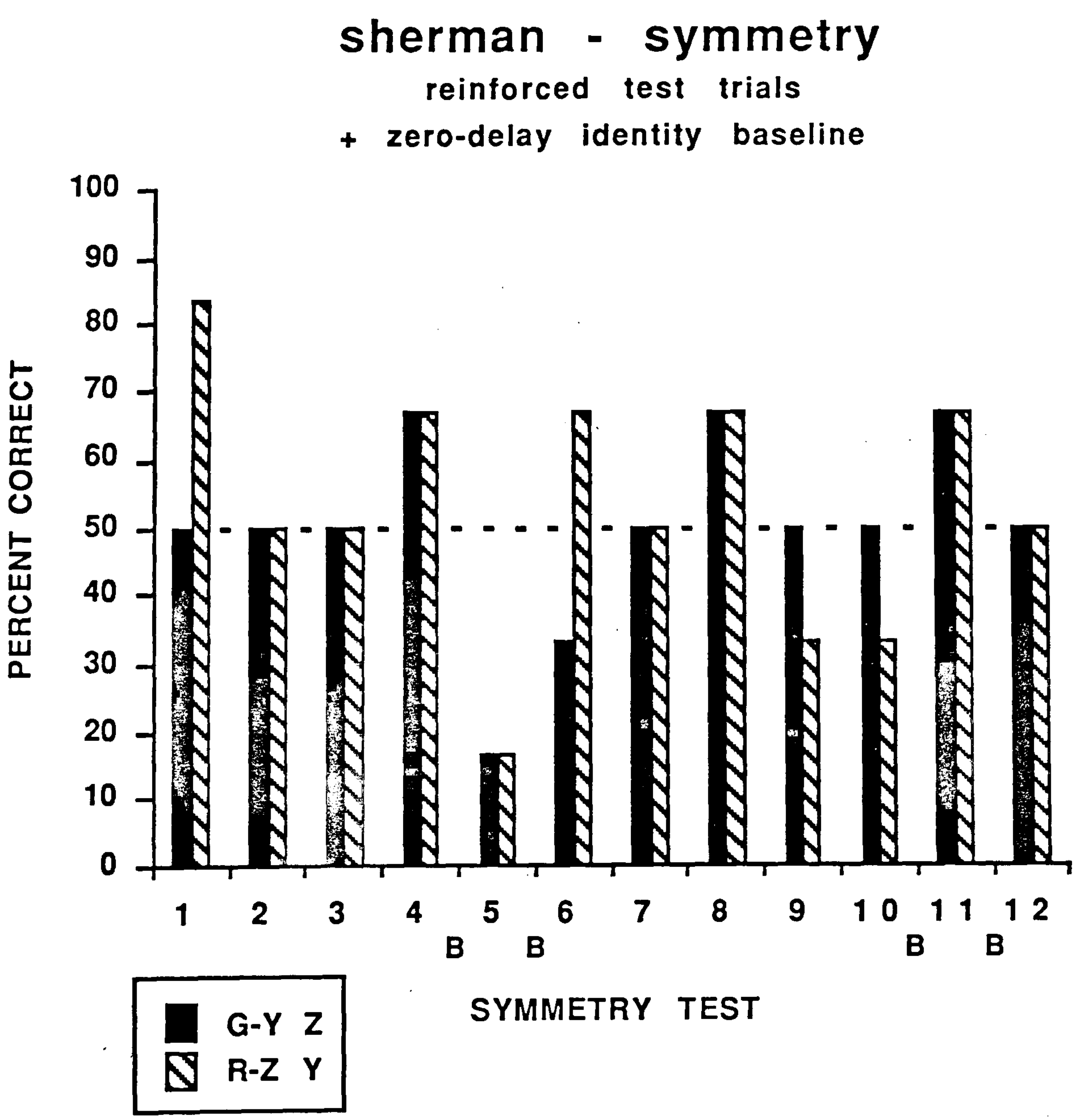


Figure 2.15 Sherman's performance on reinforced BA symmetry trial-types during each individual test session (see text).

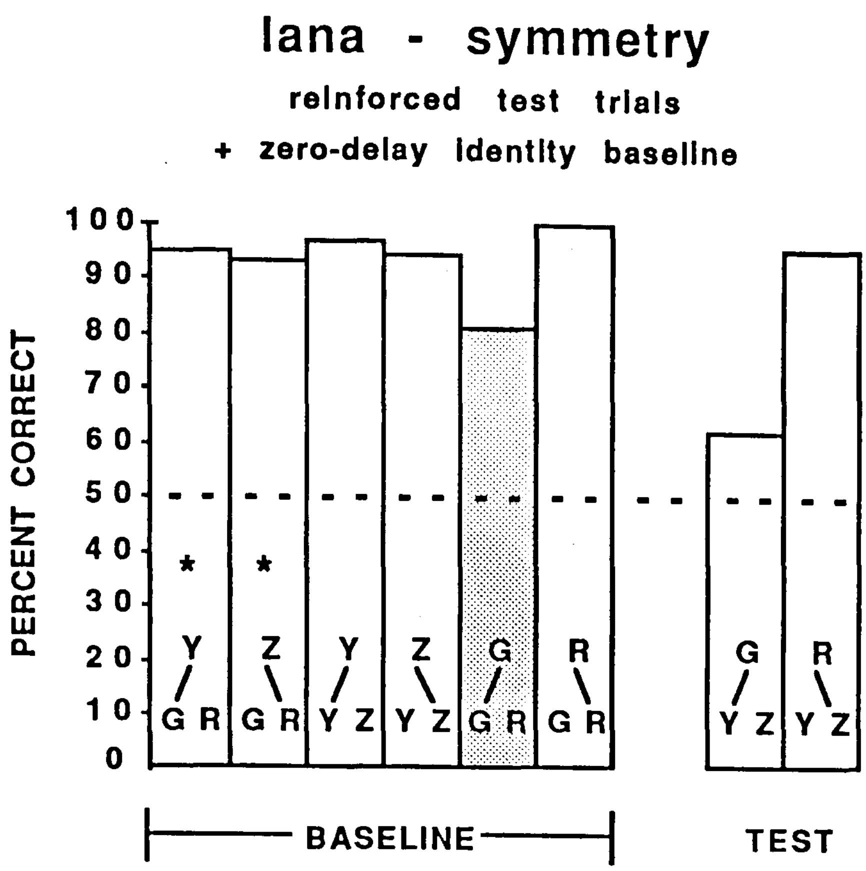

sherman - symmetry
reinforced test trials
+ zero-delay identity baseline
PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
1 0
1 1
1 2
B
B
B
B
SYMMETRY TEST
G-Y Z
R-Z Y

testing. The figure shows that Lana averaged only 80% correct on the green - green identity matching trial-type, and Sherman averaged only 70% correct on the zig-zag - red arbitrary trial-type, during the aforementioned test phases.

Now it could be argued that symmetry failed to emerge because of these somewhat diminished baselines i.e. one could claim that the baseline trials affected the test trials. However, a functional analysis of the baseline deterioration suggested quite the opposite - that is, the test trials somehow affected the baselines. The points on Figure 2.17 represent Lana's performance on the green - green identity matching trial-type which was disrupted during her final test phase. Each instance of disruption is included. The graph is divided into phases by the vertical dotted lines. These phases are marked as either Plus-T or Minus-T. Plus-T phases are those in which symmetry test trials were presented alongside the baseline trials i.e each Plus-T session denotes a symmetry test session. In Minus-T phases, the symmetry test trials were terminated. Minus-T phases correspond with the capital - B's in Figure 2.12 i.e each Minus-T session consisted of baseline trials only. The advantage of Figure 2.17 is that it enables one to clearly see what happened to Lana's performance on the green - green baseline as a function of the presence or absence of symmetry test trials. Between the 21st and 24th symmetry test session, Lana's performance on the green - green trials gradually broke down. After the 24th test session, and before the 25th, the symmetry test trials were removed and the green - green baseline immediately recovered. This pattern of baseline deterioration in the presence of symmetry test trials and baseline recovery in the absence of symmetry test trials was repeated over the sessions that followed.

In similar fashion, the points on Figure 2.18 represent Sherman's performance on the zig-zag - red baseline which deteriorated during his

Figure 2.16 Sherman's and Lana's performances during symmetry test phases in which some baseline trial-types (i.e. those depicted by the shaded bars) were disrupted.


lana - symmetry
reinforced test trials
+ zero-delay identity baseline
100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
*
*
Y / G R
Z \ G R
Y / Y Z
Z \ Y Z
G / G R
R \ G R
G / Y Z
R \ Y Z
BASELINE
TEST

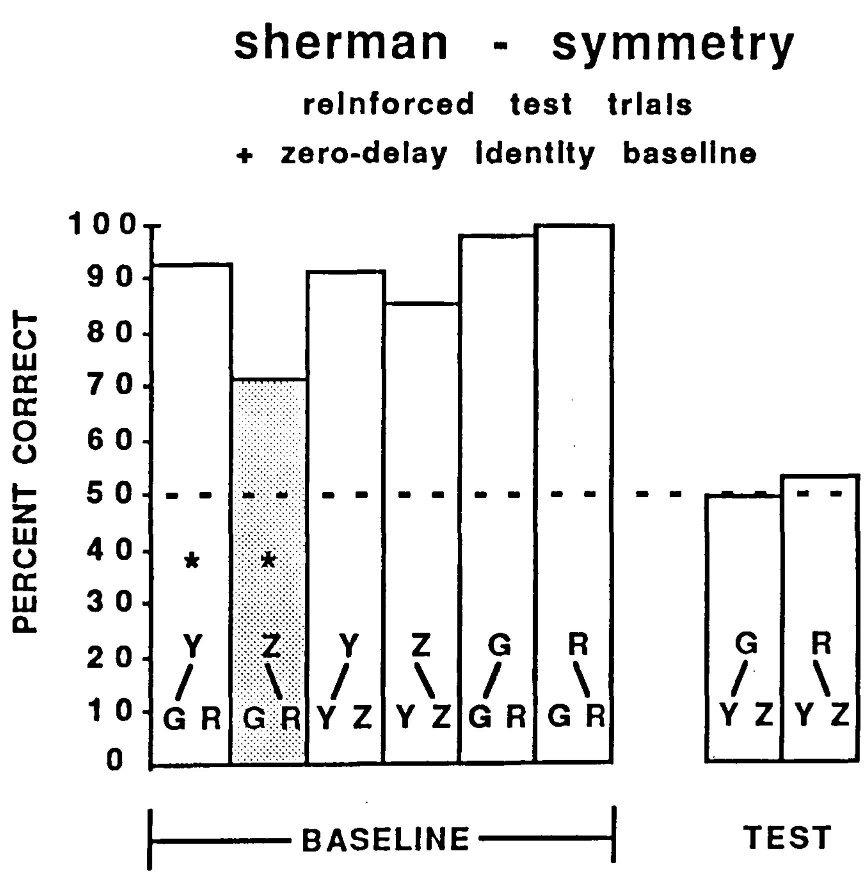

sherman - symmetry
reinforced test trials
+ zero-delay identity baseline
100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
*
*
Y / G R
Z \ G R
Y / Y Z
Z \ Y Z
G / G R
R \ G R
G / Y Z
R \ Y Z
BASELINE
TEST

Figure 2.17 Lana's performance on the green-green identity matching baseline as a function of the presence (+T) and absence (-T) of symmetry test trials (see text).

lana

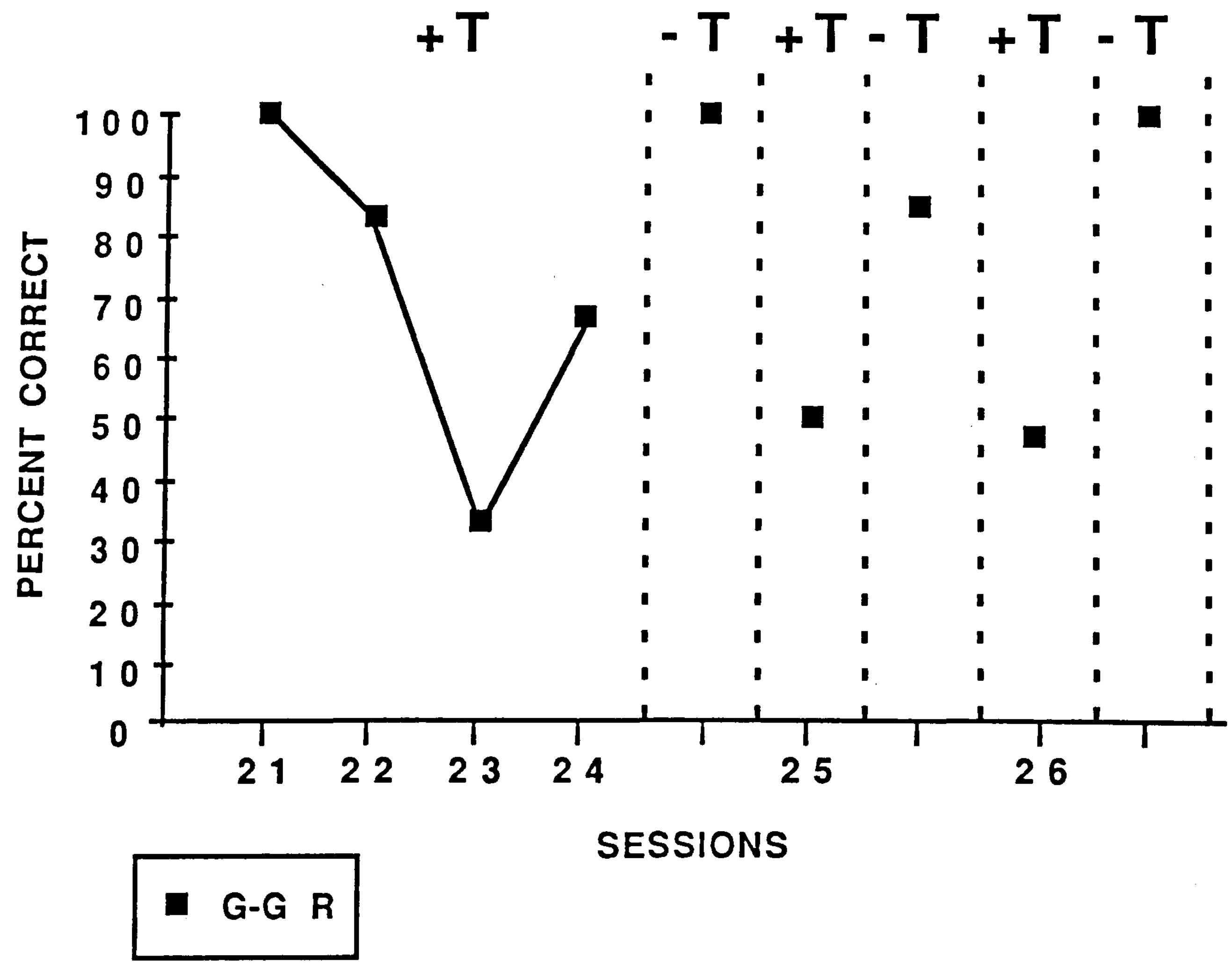
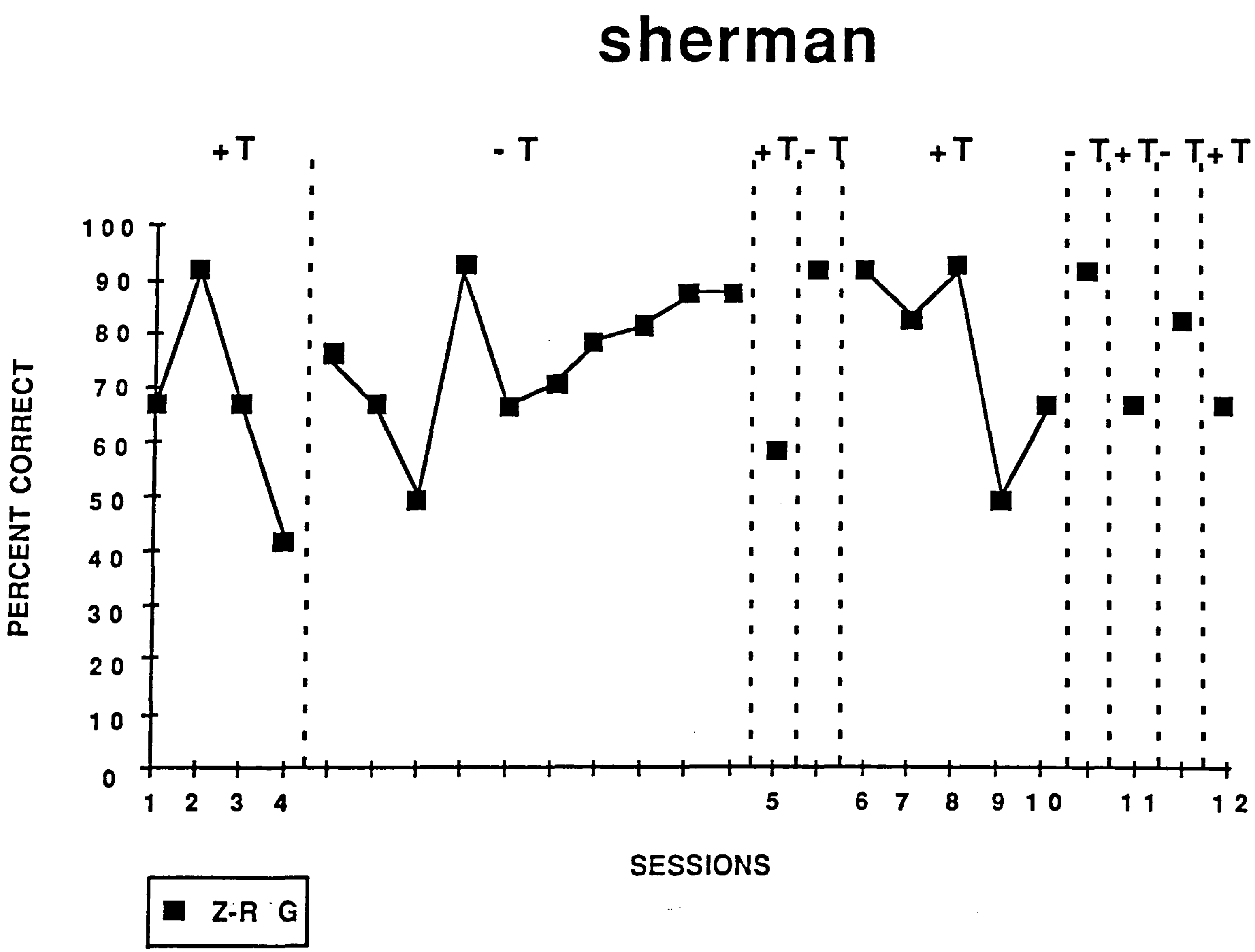

+T
- T
+T
- T
+T
- T
100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
21
22
23
24
25
26
SESSIONS
G-G R

Figure 2.18 Sherman's performance on the zig-zag - red arbitrary matching baseline as a function of the presence (+T) and absence (-T) of symmetry test trials (see text).


sherman
+T
- T
+T
- T
+T
- T
+T
- T
+T
100
90
80
70
60
50
40
30
20
10
0
PERCENT CORRECT
1
2
3
4
5
6
7
8
9
10
11
12
SESSIONS
Z-R G

symmetry tests (this time, the Minus-T phases correspond with the capital - B's in Figure 2.15). Again, when symmetry test trials were present (Plus-T) the baseline deteriorated, and when they were removed (Minus-T) the baseline recovered.

It is important to note that the patterns depicted in Figures 2.17 and 2.18 occurred even though correct responses on symmetry test trials were reinforced. It seems reasonable to assume that reinforcement on symmetry test trials would not have disrupted the baselines if the Set-A and Set-B stimuli were equivalent. If the stimuli were equivalent, then reinforcement on symmetry test trials should, if anything, strengthen the baseline relations, and certainly should not weaken them. If one reinforces choosing Y in the presence of green, and if green and Y are equivalent, then choosing green in the presence of green should also be strengthened. Similarly, if zig-zag and red are equivalent then reinforcing correct responses on red - zig-zag trials should result in improved performance on zig-zag - red trials. But the first of these baselines (green - green) deteriorated during Lana's test sessions and the second (zig-zag - red) deteriorated during Sherman's, in spite of the test trial reinforcement. For all their language training, Sherman and Lana behaved the same as many of the 'ordinary' alinguistic rhesus monkeys and baboons that were tested for symmetry by Sidman et al (1982). Like Sherman and Lana, these animals not only failed the symmetry tests but also suffered baseline disruption correlated with test trial reinforcement. In the present context, the data on 'baseline disruption' is perhaps the strongest evidence against stimulus equivalence in the arbitrary matching performance of these language-trained chimpanzees.

GENERAL DISCUSSION

In this experiment, symmetry tests were applied to a single arbitrary matching problem. Within this context, and within the time available, everything possible was done to bias the results in the chimpanzees' favour. Despite this, these language-trained chimpanzees failed the symmetry tests. Because symmetry is a necessary property of stimulus equivalence, its absence is sufficient to show that each sample and its corresponding comparison had not formed a class of equivalent stimuli. In as much as equivalence may be taken as a prerequisite of symbolic behaviour (see Devany et al, 1986; Sidman, 1977), the chimpanzees' behaviour on matching-to-sample trials could not be classed as symbolic.

Experiments reviewed in Chapter 1 of this thesis provided strong evidence for, and no convincing evidence against, the role of language in stimulus equivalence. The chimp data therefore leaves one wondering: if these chimps are linguistically accomplished then why did they fail a standard symmetry test? At this stage, no attempt shall be made to address this question. Rather, the issue shall be deferred on the grounds of expediency; we may be better able to discuss the specific implications of the chimp data at a later stage of this thesis, when additional experiments have further examined the behaviour of children on equivalence tests.

But what are the general implications of the chimp data? Firstly, the results from this experiment contribute significantly to the corpus of data on stimulus equivalence. Previous reports of the failure of animals on symmetry tests have been confirmed and extended to yet another animal species, namely chimpanzees. If one of man's closest relatives in the animal kingdom is unable to pass these tests, then what chance is there for man's other, more distant, relatives?

More significant still, the chimpanzees tested here are unique among <u>all</u> animals, in that they have had a rich and complex history of training. I doubt that any animals are as 'test-wise' as these chimpanzees. Before this experiment they must surely have ranked among the favourites for the first animals to pass a standard symmetry test. Yet they did not pass the tests.

When Sidman et al (1982) were unable to demonstrate symmetry with rhesus monkeys and baboons, they had this to say:

> "It is of course impossible to prove by failures alone that conditional discrimination procedures are incapable of establishing symmetric relations for any organism." (Sidman et al, 1982, p.42).

Now this statement may appear rather damning, although it is not (and I doubt if it was intended to be). Although undeniably true, the statement lacks any real relevance. Obviously, failures alone cannot prove that a subject is completely unable to pass equivalence tests, because a modification to the training and/or test procedure might possibly turn the failure into success. But this is not the real issue. Failures do not prove anything, but they do <u>support</u> the <u>hypothesis</u> that the subject is unable to pass equivalence tests. While failures alone are not proof of an absence of equivalence they are sufficient as <u>evidence</u> of such an absence; but then they would have to be, because presumably one cannot present anything other than failures as evidence <u>against</u> equivalence. Furthermore, the quotation above misses the most important point of all; that although it is impossible to prove a human-animal difference in equivalence formation, thus far <u>nobody has proven otherwise</u>. There is, to date, no convincing evidence that animals (not even those as 'sophisticated' as Sherman and Lana) are capable of stimulus equivalence.

In their paper, Sidman et al also list a number of procedural modifications which might yield positive evidence of symmetry in animals. This list is undoubtedly valuable - suggestions for future research are an essential first step toward progress in any area of science. No doubt others will add to the list of suggestions which, in principle, could be supplemented ad infinitum. But, in so doing, it is important not to let the list assume more than its true value. The data from Experiment 1 have been presented, in part, at a number of conferences (Dugdale, 1988; Dugdale and Lowe, 1987 (a); 1987 (b)). Of those in the audience, some (albeit only a few) have appeared reluctant to accept the chimp data, not because of any methodological flaws in the experiment or unsound reasoning on the author's part, but because of the mere possibility that procedural modifications might yield 'positive' results. These people appear to find their own suggestions sufficient for postponing judgment on the issue, or, worse still, for convincing themselves that animals can pass the tests. But suggestions alone are empirically worthless unless, that is, they are 'cashed in' as actual experimental data. One can always think of more experiments to do, but this should not excuse one from acknowledging what has already been done.

Given more time with the chimps, Experiment 1 may have been extended in accordance with one of Sidman et al's suggestions:

> "Perhaps the most relevant experience to provide would be additional symmetry tests, with initial test failures being followed by explicit reinforcement of the desired performance ...... if the animals were taught consecutive pairs of conditional discriminations, with the second of each pair always the symmetric version of the first, would the subjects eventually perform a symmetry test accurately the

> first time they encountered it? Would they learn the general principle, 'Sample and correct comparison are interchangeable'?" (Sidman et al, 1982, pp42-43).

The hypothesis is that, unlike humans, animals do not ordinarily experience enough exemplars for symmetry to emerge. Experiment 2 tested this hypothesis. The language-trained chimps were no longer available, so pigeons were used instead. Pigeons might be seen as a poor substitute for chimpanzees, but their use as subjects in the present context could be advantageous. If pigeons respond positively to the exemplar training outlined above, then one would anticipate the same result from other animal species higher up the phylogenetic scale.

CHAPTER 3

THE EFFECTS OF EXEMPLAR TRAINING ON THE SYMMETRY TEST PERFORMANCES OF PIGEONS

1. Method
2. Results
3. Discussion

EXPERIMENT 2

METHOD

SUBJECTS

The subjects were four adult male homing pigeons (Columbia livia). Two of the birds (Exocet and Falcon) were experimentally naive, whereas the other two (Ben and Eric) had served as subjects in a previous experiment on concurrent schedules. Throughout the present experiment, all four birds were individually housed and maintained at 85% - 90% of their free-feeding weights, with water freely available in their home cages.

APPARATUS

The experiment was conducted in a standard Lehigh Valley pigeon test chamber, equipped with a three key response panel. Pecks to the keys with a force of at least 0.15 N registered as responses. Each key was made of transparent plastic mounted behind a 2.5 cm diameter hole in the front panel of the chamber. The keys were horizontally aligned 8 cms apart from centre to centre, and were 25 cm above the chamber floor. Each key was illuminated from behind via an IEE in-line multiple stimulus projector.

A special set of stimuli were custom designed and prepared as suggested by McConnell (1966). Forms were drawn white on a black background to occupy a 2 cm square area. In addition, the keys could be fully lit with two hues, red and green, produced by projecting light through Kodak Wratten filters 26 and 61 respectively.

Above the centre key was a 2.5 W houselight providing dim illumination in the chamber, and below the key was a food magazine or hopper which was illuminated when food (mixed grain) was dispensed. The pigeon chamber was located in a sound-attenuating room to isolate subjects from extraneous noises, and additional masking was provided by white noise played through a speaker on the front panel. The programming of experimental events and data collection were handled by a Gemini mini-computer.

PROCEDURE

Preliminary Training

All birds were initially given one session of preliminary training, which taught them to respond to the stimuli that would later appear during the first arbitrary matching task. In the first stage, the birds received magazine training. The houselight remained off throughout, and the hopper was raised (and lit) at irregular intervals. Once the birds had learned to feed quickly and reliably from the lit hopper, they were taught, via autoshaping, to peck at illuminated keys. On each trial the houselight remained off, and a cross or a circle appeared on the left or right key. Each combination of stimuli and side keys occurred equally often and at random. Pecks to the lit key were designated correct and pecks to either dark key were incorrect. In these and all subsequent trials (unless otherwise noted), correct responses turned off the stimuli and produced 2 seconds access to the lit grain hopper, followed by a 5 second inter-trial interval. Incorrect responses turned off the stimuli and produced the inter-trial interval only. The next stage began once the pigeon had learned to peck the lit key, and only the lit key. The houselight still remained off, and each trial began with a dot or a wavy line appearing equally often, and at random, on the centre key. A peck to the lit centre key produced a stimulus on one of the side keys, again at random. The centre key stayed on. If the dot was the centre key stimulus (sample) then the circle appeared as the side key stimulus (comparison). Conversely, if the sample was a wavy line, then the cross appeared as the comparison. Pecks to the comparison were correct, whereas pecks to the dark key were incorrect. Training continued until the birds learned to peck the sample and comparison in sequence for six consecutive trials per trial-type (sample-comparison pairing). Finally, the houselight was switched on, and the birds continued to peck the sample and corresponding comparison without error for a further 12 trials (6 per trial-type). Preliminary training then ended, and the subjects proceeded to the next phase.

Exemplar Training (arbitrary matching)

Following pre-training, all birds were taught a succession of arbitrary matching problems (see Figures 3.1 and 3.2). All sessions began with the illumination of the houselight, which remained on until the session ended. The birds initially received AB training, as represented in Figure 3.1. Each trial began with a Set-A sample, either the dot or the wavy line. A peck to the sample produced the Set-B comparisons, the circle and the cross. Once the comparisons appeared, further pecks to the sample (which remained on) had no scheduled effect. When the dot was the sample, pecks to the circle were correct, and when the wavy line was the sample, pecks to the cross were correct. The stimuli were presented at random, with the restriction that all 4 sample-comparison configurations had to occur

Figure 3.1 Stimulus relations taught in the first four stages of Experiment 2. The dotted lines separate the stimuli in each stage, and the training sequence is determined by reading from left to right and down the page. Arrows point from samples to corresponding comparisons.

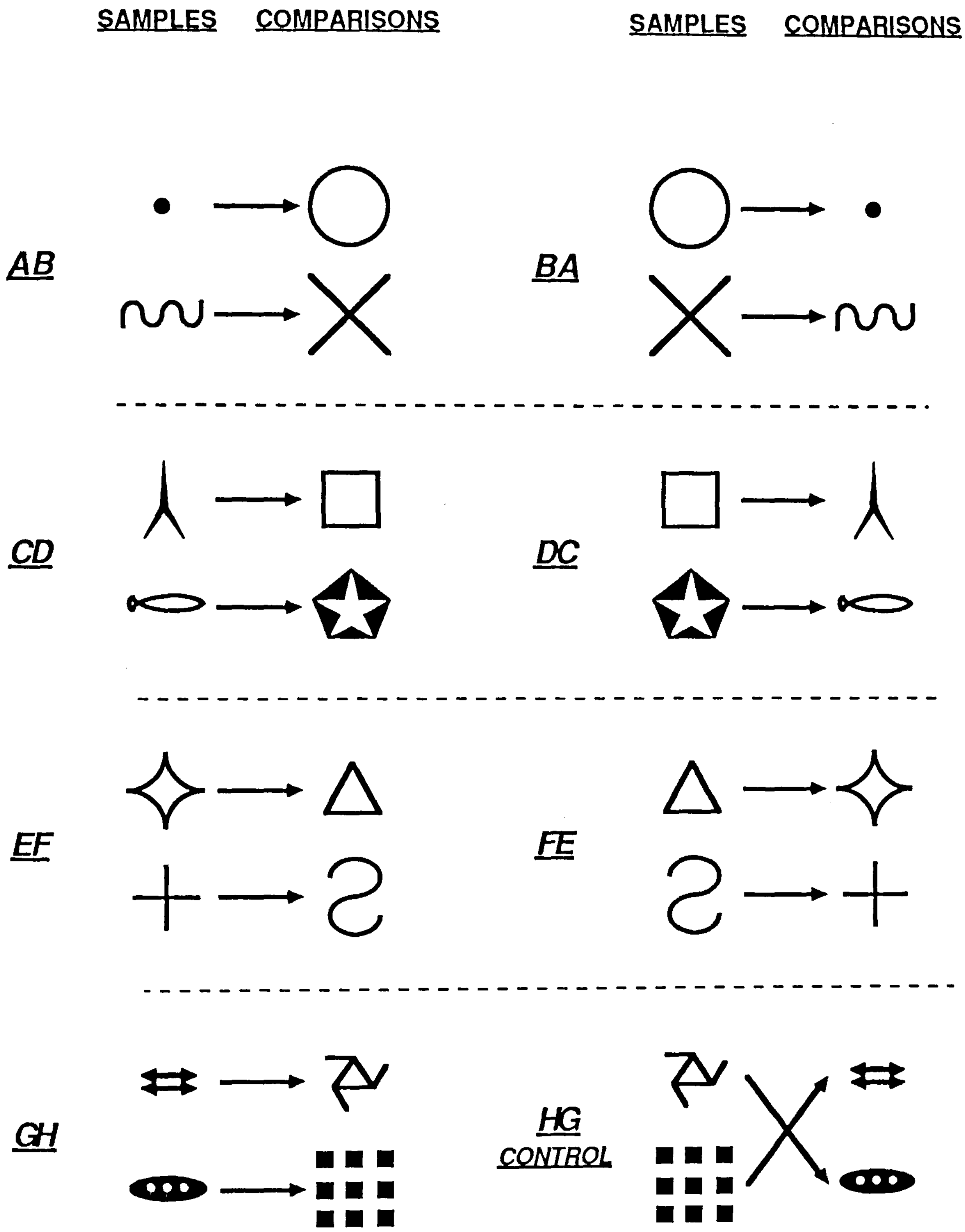
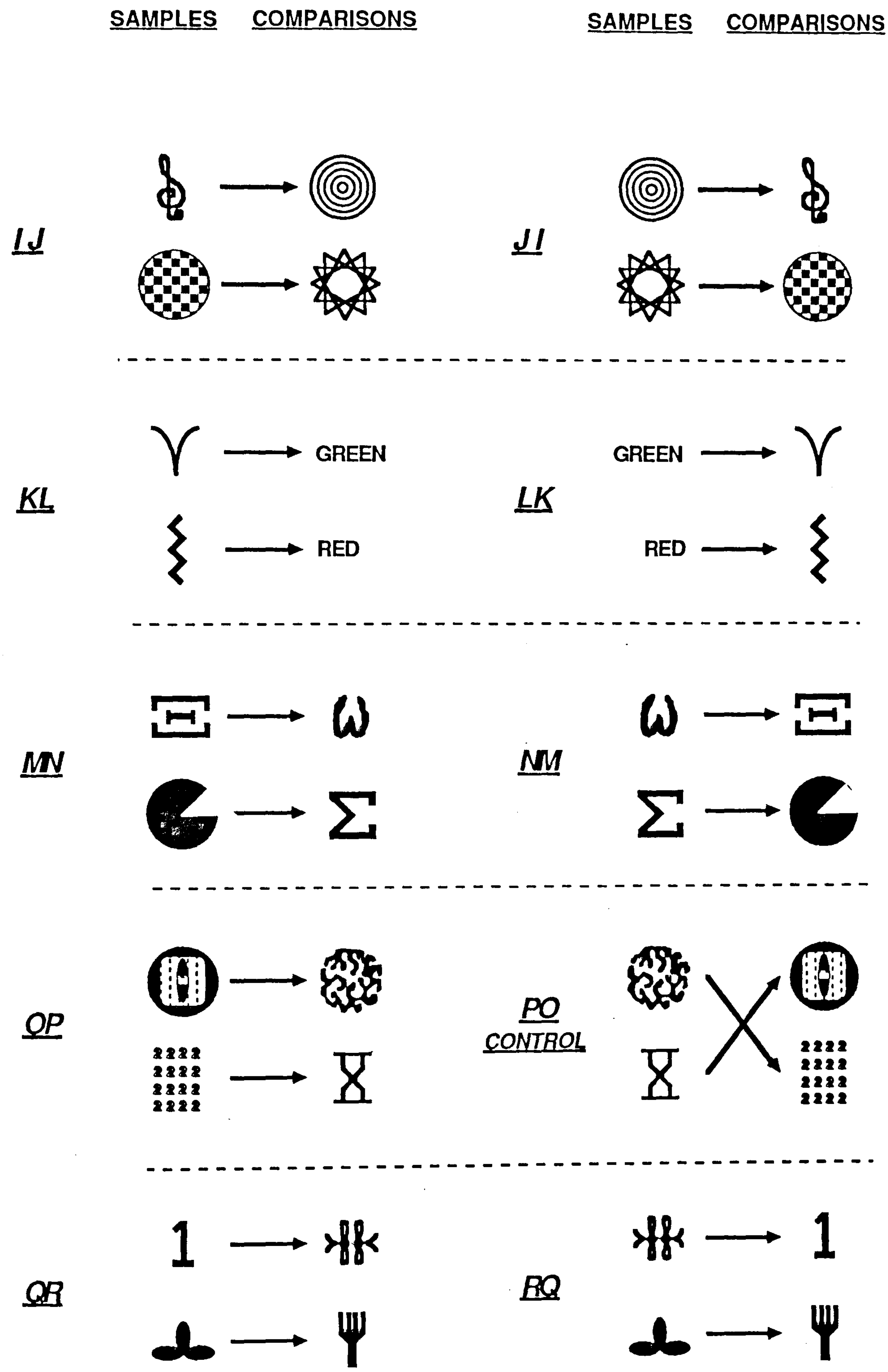

SAMPLES
COMPARISONS
SAMPLES
COMPARISONS
AB
BA
CD
DC
EF
FE
GH
HG
CONTROL

Figure 3.2 Stimulus relations taught in the last five stages of Experiment 2.


SAMPLES COMPARISONS
SAMPLES COMPARISONS
IJ
JI
KL
GREEN
RED
LK
GREEN
RED
MN
NM
OP
PO
CONTROL
QR
RQ

before any could repeat.

In the first AB session, a single peck to the sample produced the comparisons. In the second session, three sample pecks produced the comparisons. Then, in the third session, the sample response requirement was raised to five pecks, where it remained throughout the experiment. Previous studies have shown that more than one required response to the sample may yield accelerated acquisition and improved accuracies on matching-to-sample tasks (see, for example, Sacks, Kamil and Mack, 1972).

Initial AB sessions consisted of 96 non-corrected trials (48 per trial-type). However, after 5 sessions with Exocet, 6 with Falcon, 9 with Eric and 10 with Ben, a correction procedure was implemented. Errors caused trials to repeat, and each session ended when the bird had responded correctly to 48 trials (24 per trial-type). Several factors led to the implementation of correction. Firstly, introducing correction meant that although the number of trials per session varied, the number of reinforced trials remained a constant 48 per session. Because each bird's food intake was more or less constant across sessions, it became far easier to keep their weights within 85% - 90% of their free-feeding levels. Secondly, all four birds developed strong preferences for either the left or right key during non-corrected trials. Changing the procedure to correction meant that on half the trials the birds were forced to respond to their non-preferred side key in order to secure reinforcers. Presumably, correction would eventually weaken the position preference and would also help attenuate other comparison-based response tendencies, thus enabling faster acquisition of arbitrary matching. Finally, correction may be beneficial not only in the acquisition of matching, but also in its subsequent transfer to novel test situations (Catania, personal communication; Zentall, Edwards and Hogan, 1984). The general approach to the current experiment was the same as in Experiment 1; the procedures were designed to maximise the subjects' chances of success during transfer phases, without invalidating the tests.

AB training was continued until the subjects reached a criterion of 85% correct responses per trial-type for two consecutive sessions. Then, in the next stage, sessions consisted of 48 reinforced trials (24 per trial-type) of the BA task depicted in Figure 3.1. The BA task was the symmetrical counterpart of the trained AB relation. Reinforcers were contingent upon choosing the dot comparison in the presence of the circle, and the wavy line comparison in the presence of the cross. BA training was continued until the subject reached the 85% learning criterion, at which point the former procedures were repeated with a new pair of tasks, namely CD and DC. The CD task was trained to criterion, followed by its symmetrical counterpart, DC (see Figure 3.1). Then the EF and FE relations were trained.

So, in the first three stages, the subjects were trained on consecutive pairs of arbitrary matching tasks, the second task of each pair being the symmetric version of the first. The question was would this exemplar training result in the formation of a general principle of interchangeability, or equivalence, with respect to the samples and comparisons from each pair of related tasks? If so, one would expect the second task of each pair to be learned much faster than the first, and that such savings in learning would perhaps increase proportionally with each successive pair trained. The difficulty with this is that one would expect the same pattern of results even if the birds had not learned to respond according to equivalence. Savings were anticipated on the second task of each related pair simply because the stimuli at this stage were no longer completely novel - they had all been encountered before, albeit in different locations, during training of the first task of the pair. Others have commented on the notorious 'novelty aversion' of pigeons (Lombardi, Fachinelli and Delius, 1984; Zentall, Edwards and Hogan, 1984) and have noted how it may account for the less than optimal transfer of pigeons' performances on reflexivity tests using completely novel stimuli. In the present context, however, any 'novelty aversion' would serve to increase rather than decrease savings in learning. Furthermore, there are a number of other reasons why a subject's performance may improve with each successive matching problem e.g. general habituation to the apparatus and procedures; suppression of control by incidental or irrelevant stimulus features, such as the position of correct comparisons; learning to attend to the stimuli on the keys; and so on. These factors are common to all matching problems and can be learned independently of stimulus equivalence.

In conclusion, a novelty aversion, coupled with general 'learning to learn', may produce exactly the same incremental savings in learning as one might expect from an emergent tendency to respond according to stimulus equivalence. What is needed, then, is some control procedure for establishing whether savings are equivalence related or not. The fourth pair of tasks provided this necessary control condition (see Figure 3.1). After training GH to criterion, HG control sessions were presented, in which the former samples and comparisons were interchanged so as not to form a symmetrical counterpart; rather, the opposite relationships held. If one denotes the initial sample - comparison associations as G_1 - H_1 and G_2 - H_2, then the new associations taught after establishing GH were H_1 - G_2 and H_2 - G_1. If any previous savings were based on general factors unrelated to equivalence then the subject would presumably show a similar (if not a larger) degree of savings on the control task. If, however, previous savings were equivalence related, then the control trials probably would _retard_ the subject's matching performance. If so, then any savings probably would be less than one might anticipate from preceding trends, and the subject might even take longer to learn the second task than the first. The same design has been used before by D'Amato, Salmon, Loukas and Tomie (1985) and Hogan and Zentall (1977), and it apparently provides an extremely sensitive measure of equivalence within the context of exemplar training with pigeons.

After acquisition of the HG control task, the training sequence was repeated with a new series of stimuli (see Figure 3.2). The subjects were to be taught three more pairs of symmetrical counterparts (IJ and JI; KL and LK; MN and NM), followed by one more control set (OP and PO control). The final pair to be trained were the symmetrical counterparts, QR and RQ.

In addition to the main training sessions outlined above, the birds were presented with baseline maintenance sessions. Baseline maintenance sessions were presented prior to each main training session and included trials from tasks already learned. For example, when subjects were learning BA matching, each main BA session was preceded with a maintenance session in which the AB performances were reviewed. Similarly, when CD was being learned, the maintenance sessions reviewed the subjects' AB and BA performances. The review procedure consisted in presenting blocks of 12 reinforced trials (6 per trial-type) of each baseline, or previously learned task, followed by a repeat block of any baseline that fell below the 85% criterion during the first block. Thus, while birds learned a new task, criterion performances were maintained on tasks already learned. Initially, _all_ previously learned tasks were to be included in the maintenance sessions, but this strategy could not be sustained. Each time the bird learned a task, at least one more block of reinforced trials had to be added to the maintenance sessions. This meant that the pigeon's daily food intake increased with each successive task learned, and thus it became increasingly likely that the bird might eventually exceed 90% of its free-feeding weight (the upper limit for this experiment). Conditioned reinforcers were occasionally used in an attempt to prevent the birds' weights from rising excessively (i.e. on some trials the hopper light came on after a correct response but the hopper was not raised). Conditioned reinforcement was only ever given during baseline maintenance sessions, never during main training sessions. This helped to suppress body weights, but only temporarily. Eventually, each bird's weight came consistently close to exceeding its upper limit, and at that point the maintenance sessions were altered. Subsequently, maintenance sessions were presented only while the subject learned the second task of a given pair, and then they only included trials from the first task of that pair. This meant that the review sessions were constant throughout the experiment with respect to maintaining the first task of each pair while the second task was being learned.

Further details of the maintenance sessions, such as their exact composition at particular stages of the experiment, can be found in the Results section below.

RESULTS

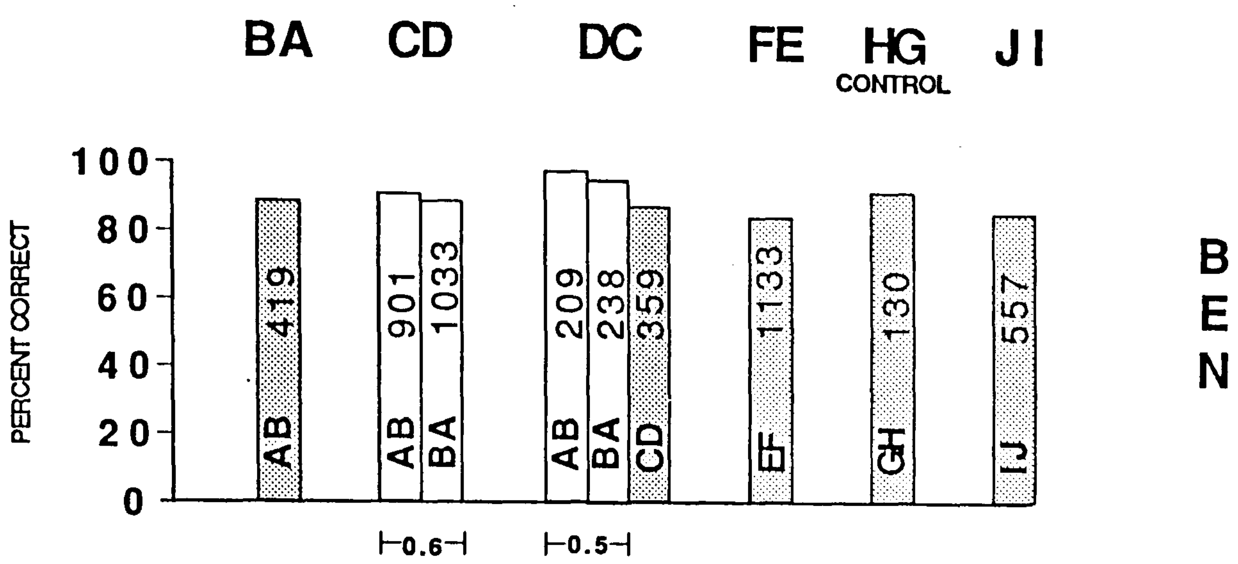
Baseline Maintenance

Each row of bars in Figures 3.3 and 3.4 depict one pigeon's overall performance during successive baseline maintenance phases. Each column heading (BC, CD, DC etc.) refers to the main task on which the subjects were trained, while the bars grouped below each heading represent the subjects' performances on the accompanying baseline tasks. Each baseline is identified by the letters within the bars, and the numbers within the bars show how many trials the subject received per baseline. The shaded bars highlight related baselines i.e. those using the same stimuli as their accompanying main training task. The unshaded bars depict unrelated baselines i.e. those that used different stimuli than the main training tasks which they accompanied. If conditioned reinforcers were used on baseline tasks, then this is noted by the numbers along the horizontal axes. These numbers refer to the probability of receiving primary reinforcers; for example, a value of 0.6 means that primary reinforcers were scheduled for 60% of the trials, and conditioned reinforcers for 40%.

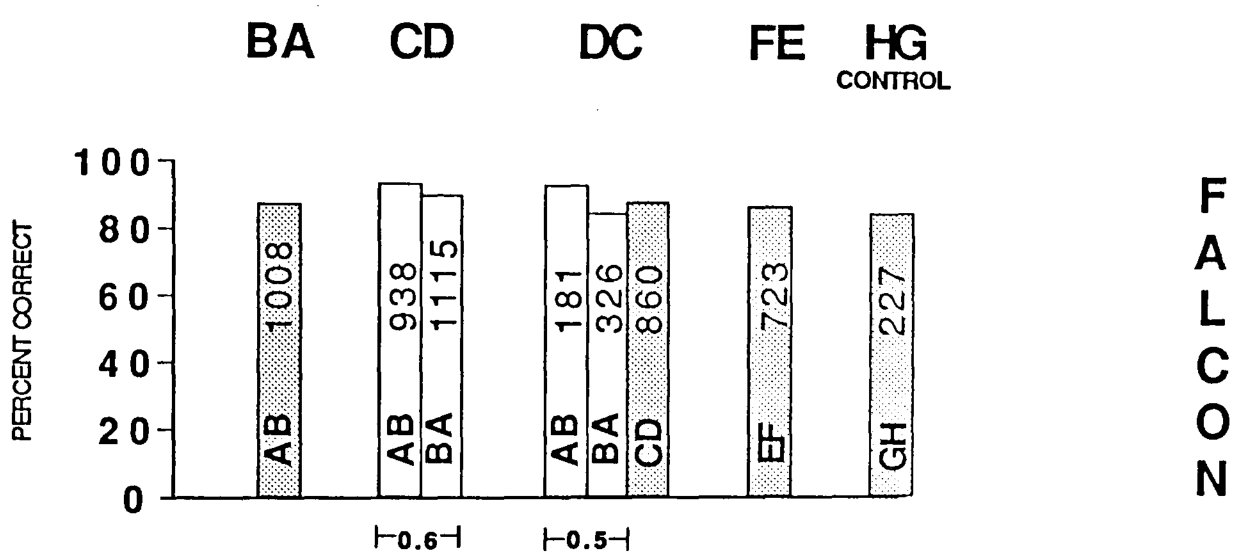
Several points are illustrated by Figures 3.3 and 3.4. Firstly, each bird reached a different stage of the experiment. Exocet was the only bird to complete all stages, learning tasks AB through to RQ. In contrast, Eric learned all but the last pair of tasks (QR and RQ), Ben learned up to and including task JI, and Falcon learned up to and including the HG control task. These differences reflected the speed with which each bird learned the arbitrary matching tasks; generally, Exocet was the fastest learner, followed by Eric, Ben and Falcon in decreasing order.

The birds also differed according to the point at which unrelated (unshaded) baselines were dropped from the maintenance phases. With Ben and Falcon (Figure 3.3) the unrelated baselines were dropped after they had learned the DC task, whereas with Exocet and Eric (Figure 3.4) the unrelated baselines were dropped much later, after they had learned the HG control task. The unrelated baselines were removed when a bird came consistently close to exceeding 90% of its free-feeding weight. Beyond this point, continuation of the unrelated baselines probably would have pushed weights over the prescribed limit; these baselines

Figure 3.3 Ben and Falcon's overall performances during successive baseline maintenance phases (see text for further details).

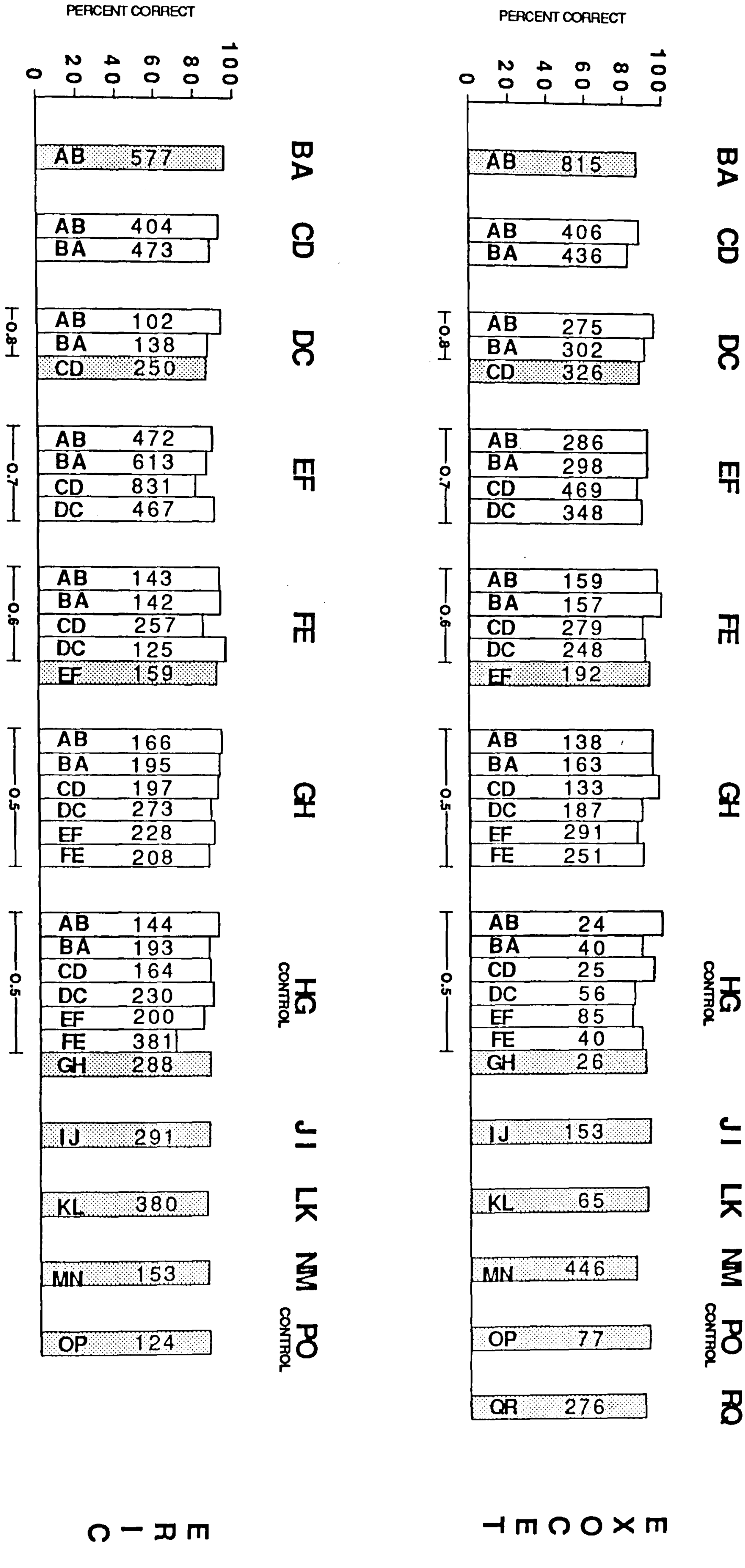


BA
CD
DC
FE
HG
CONTROL
JI
PERCENT CORRECT
100
80
60
40
20
0
AB 419
AB 901
BA 1033
AB 209
BA 238
CD 359
EF 1133
GH 130
IJ 557
0.6
0.5
BEN



BA
CD
DC
FE
HG
CONTROL
PERCENT CORRECT
100
80
60
40
20
0
AB 1008
AB 938
BA 1115
AB 181
BA 326
CD 860
EF 723
GH 227
0.6
0.5
FALCON

Figure 3.4 Exocet and Eric's overall performances during successive baseline maintenance phases (see text for further details).



ERIC
PERCENT CORRECT
0 20 40 60 80 100
BA: AB 577
CD: AB 404, BA 473
DC: AB 102, BA 138, CD 250
EF: AB 472, BA 613, CD 831, DC 467
FE: AB 143, BA 142, CD 257, DC 125, EF 159
GH: AB 166, BA 195, CD 197, DC 273, EF 228, FE 208
HG CONTROL: AB 144, BA 193, CD 164, DC 230, EF 200, FE 381, GH 288
JI: IJ 291
LK: KL 380
NM: MN 153
PO CONTROL: OP 124
0.8
0.7
0.6
0.5
0.5
EXOCET
PERCENT CORRECT
0 20 40 60 80 100
BA: AB 815
CD: AB 406, BA 436
DC: AB 275, BA 302, CD 326
EF: AB 286, BA 298, CD 469, DC 348
FE: AB 159, BA 157, CD 279, DC 248, EF 192
GH: AB 138, BA 163, CD 133, DC 187, EF 291, FE 251
HG CONTROL: AB 24, BA 40, CD 25, DC 56, EF 85, FE 40, GH 26
JI: IJ 153
LK: KL 65
NM: MN 446
PO CONTROL: OP 77
RQ: QR 276
0.8
0.7
0.6
0.5
0.5

represented a significant proportion of the birds' dietary intakes. Ben and Falcon's unrelated baselines were removed much sooner than Eric and Exocet's simply because the former pair threatened to exceed the critical weights much sooner than the latter.

The number of trials per baseline task varied according to two main factors. Firstly, since baselines always accompanied main training tasks it followed that the longer a bird took to learn the main task then the more trials it received on each baseline within that particular phase. Secondly, within each phase, some baselines required fewer daily trials than others, simply because a bird's performances on the former were of a consistently high and stable level that additional trials seemed unnecessary. The variation in trial numbers therefore partly reflected the experimenter's insistence in maintaining accuracies without detriment to the birds' overall weights.

But most important of all, Figures 3.3 and 3.4 clearly show that each bird's baselines were remarkably stable throughout the experiment. All of the related baselines stayed at or above the 85% criterion level, and only one unrelated baseline fell significantly below criterion (i.e. Eric's FE baseline (69% correct) when learning the HG control task; see Figure 3.4). Although baseline disruption was a noticeable feature of Experiment 1, it was virtually absent in the present experiment. The reason for this disparity is not clear at present, but one possibility is that disruption was enhanced in Experiment 1 because baseline trials were interspersed with test trials, whereas in the present experiment all tasks were presented in successively separate blocks. Whatever the reason, if, in the present experiment, stimulus equivalence is shown to be absent from the pigeons' arbitrary matching performances, then this would be in spite of their excellent baselines, and not because of any baseline disruption.

Analysis of Savings

Savings in learning the second task of each consecutive pair were calculated using the following formula:

$$\text{Percentage savings} = \frac{X - Y}{X} \times 100$$

where X = trials to criterion for the first task and Y = trials to criterion for the second.

Tables 3.1 - 3.3 show the trials to criterion for each task learned and the percentage savings achieved on each task pair. In addition, the percentage savings for each bird are represented pictorially in Figure 3.5.

There are two main issues to address. Firstly, were there any savings in learning on task pairs consisting of symmetrical counterparts i.e. symmetrical pairs? Secondly, if so, how did these savings compare with those achieved on control pairs, in which the second task was not a symmetrical counterpart of the first?

Firstly, savings from symmetrical pairs were far from systematic for each bird. For example, Falcon's savings on the first two symmetrical pairs increased from 46 to 58%, but then decreased to 10% on the third symmetrical pair. Furthermore, Falcon's savings then increased to 29% on the subsequent control pair. Ben produced a similar pattern. His savings on the first three symmetrical pairs initially increased from 43 to 69%, but then dropped to 3%. Ben then produced increased savings of 38% on the subsequent control pair, followed by a decrease to 22% on his final symmetrical pair.

Table 5.1 Table depicting the number of trials to criterion for each task learned by Falcon and Ben, and the percentage savings they achieved on each successive pair of tasks.

PIGEON	TASK	TRIALS TO CRITERION	PERCENT SAVINGS
FALCON	AB	5956	
	BA	3189	46%
	CD	5881	
	DC	2451	58%
	EF	2222	
	FE	2007	10%
	GH	977	
	HG control	696	29%
BEN	AB	2683	
	BA	1528	43%
	CD	4017	
	DC	1228	69%
	EF	2853	
	FE	2757	3%
	GH	831	
	HG control	512	38%
	IJ	1814	
	JI	1418	22%

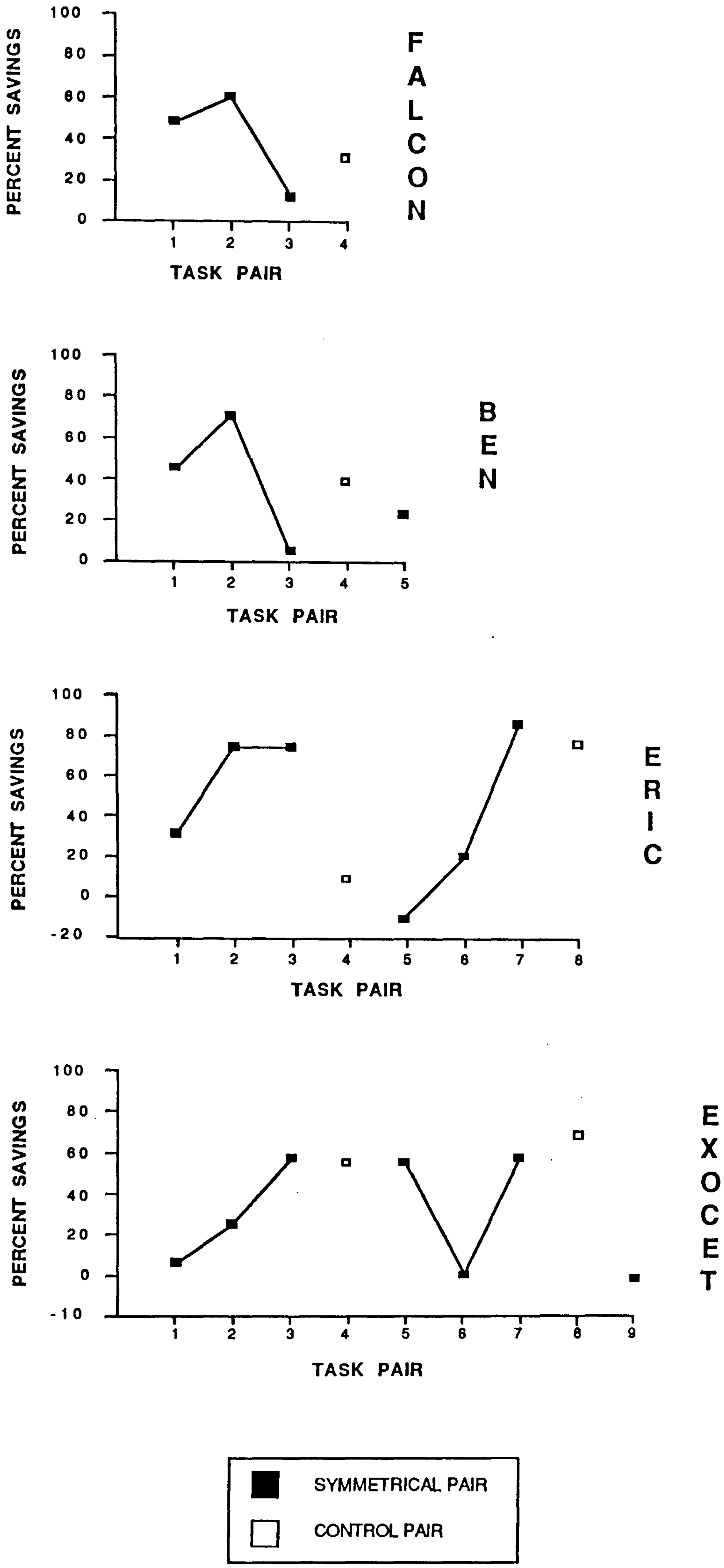
Table 5.2 Table depicting the number of trials to criterion for each task learned by Eric, and the percentage savings he achieved on each successive pair of tasks.

PIGEON	TASK	TRIALS TO CRITERION	PERCENT SAVINGS
ERIC	AB	1773	
	BA	1233	30%
	CD	2103	
	DC	580	72%
	EF	2235	
	FE	632	72%
	GH	1211	
	HG control	1187	2%
	IJ	882	
	JI	980	-10%
	KL	1725	
	LK	1392	19%
	MN	2336	
	NM	381	84%
	OP	1540	
	PO control	407	74%

Table [illegible] – Table depicting the number of trials to criterion for each task learned by Exocet, and the percentage savings he achieved on each successive pair of tasks.

PIGEON	TASK	TRIALS TO CRITERION	PERCENT SAVINGS
EXOCET	AB	2037	
	BA	1907	6%
	CD	1674	
	DC	1249	25%
	EF	1898	
	FE	823	57%
	GH	981	
	HG control	445	55%
	IJ	789	
	JI	359	55%
	KL	795	
	LK	791	0%
	MN	2541	
	NM	1448	57%
	OP	1091	
	PO control	369	66%
	QR	1102	
	RQ	1118	-1%

Figure 3.5 Percentage savings achieved on each successive task pair, for all four subjects in Experiment 2.


PERCENT SAVINGS
TASK PAIR
FALCON
BEN
ERIC
EXOCET
SYMMETRICAL PAIR
CONTROL PAIR

Both Ben and Falcon had therefore produced control-pair savings which were well within the range of savings produced on symmetrical pairs. This is hardly what one would expect if the birds were learning that the samples and corresponding comparisons were interchangeable, or equivalent. If equivalences were emerging as a result of the exemplar training, then one would at least expect greater overall savings on symmetrical than control pairs.

In the first half of the experiment, Eric's pattern of savings seemed indicative of equivalence (see Figure 3.5). Eric began by achieving 30% savings on the first symmetrical pair, after which savings rose to 72% on the next two symmetrical pairs, before dropping to 2% on the following control pair. However, although the pattern thus far suggested equivalence, subsequent events suggested otherwise. Although the next pair of tasks (pair 5) were symmetrical counterparts, Eric produced 'savings' of -10% i.e. he took 10% longer to learn the second task than the first. Once again, this is not what one would predict if equivalence were emerging. Although savings increased over the next two symmetrical pairs (rising first to 19% and then to 84%), Eric ended by achieving 75% savings on a control pair (pair 8). He, like Ben and Falcon, had therefore produced control pair savings which were well within the range produced on symmetrical pairs.

In spite of the exemplar training, Exocet, too, gave no evidence of equivalence. His savings gradually increased to 57% over the first three symmetrical pairs, but this figure was matched by the subsequent control pair (55%). Thereafter, Exocet's performance on the symmetrical pairs became somewhat erratic - savings went from 55 to 0 to 57% over the next three symmetrical pairs, and then increased to 66% on a control pair, before falling to -1% on the final symmetrical pair (pair 9). Overall, then, Exocet's performance was variable, but his

Table 3.4 The mean percentage savings achieved on symmetrical and control pairs for each subject in Experiment 2.

SUBJECT	MEAN PERCENTAGE SAVINGS	
	SYMMETRICAL PAIRS	CONTROL PAIRS
BEN	34	38
ERIC	44	38
EXOCET	28	60
FALCON	38	29
Overall Means	36	41

savings from the two control pairs were in one case better than, and in the other approximately equal to, his asymptotic level of savings from the symmetrical pairings.

Finally, the mean percentage savings from symmetrical and control pairs were calculated for each bird (see Table 3.4). A two-tailed correlated samples t-test (Robson, 1973) showed no significant difference in savings obtained on symmetrical and control pairs ($t(3) = 0.66$, $p < 0.05$).

DISCUSSION

In this experiment, all four birds received extensive amounts of symmetry exemplar training. The number of trials each bird received on symmetry exemplar training may be calculated by adding the number of trials to criterion for the second task of each symmetrical pair to the number of trials received on symmetrical pairs during baseline maintenance sessions. (N.B. - the number of trials to criterion for the first task of each symmetrical pair perhaps should not be included; when one is teaching the initial task of each pair it is not necessarily obvious that one is reinforcing responding in accordance with a symmetrical relation). Calculating thus, Exocet received 15,064 trials of exemplar training, Eric received 13,745, Falcon 12,798, and Ben 11,780. There were occasional periods of control pair training (which perhaps could be best described as the opposite of symmetry exemplar training), but these periods were minimal; the ratio of 'symmetry' training trials to control training trials was approximately 16:1 for Exocet, 7:1 for Eric, 14:1 for Falcon and 18:1 for Ben.

Despite the extensive exemplar training, none of the birds gave any convincing evidence for the emergence of a general principle of sample-comparison interchangeability, or symmetry. The birds instead appeared to learn a set of simple conditional or 'if-then' relations (i.e. unidirectional sample-comparison response chains), and in this respect the present results agree with the corpus of data on pigeon matching-to-sample (see Carter and Werner (1978), and Wilson, Mackintosh and Boakes (1985), for reviews). Although there were savings in learning on symmetrical pairings, these savings did not appear to be related in any simple systematic way with length of exemplar training. Rather, savings on symmetrical pairs fluctuated throughout the experiment, and savings on control pairs fell squarely within the range of these fluctuations. The birds' performances during symmetrical and control pairings were, 'savings-wise', indistinguishable, and all savings were most likely due to general learning factors, unrelated to stimulus equivalence (see Procedure).

The search for equivalence in the arbitrary matching behaviour of animals will undoubtedly continue, but the present results, taken together with previous failures to generate symmetry with non-humans, warn against relying exclusively on studies of animals if one wishes to learn where stimulus equivalence comes from. The present research programme consequently took to another, potentially more informative, path. Although few in number, studies of children have already provided some interesting data on the role of language in general, and of naming in particular, in the formation of stimulus equivalence (see Chapter 1). The next series of studies further examined naming and stimulus equivalence in children, with the aim of learning something more about the mechanisms through which equivalence may emerge.

CHAPTER 4

THE ROLE OF COMMON NAMING IN THE FORMATION OF STIMULUS EQUIVALENCE: STUDIES OF 4-5 YEAR OLD CHILDREN

INTRODUCTION

This final series of experiments involved 4-5 year old children and focused on the role of common naming in the formation of stimulus equivalence. The experiments, however, were not all designed at the same time, to be executed in some pre-determined sequence; rather they all evolved from each other. Each successive experiment was suggested by the results of the experiment which preceded it.

This 'evolutionary progression' stemmed from a common origin, the raw material of which was a 'pool' of thirty children. All thirty were initially presented with the same AB matching task as was earlier presented to the chimps (see Experiment 1). Eight of the children learned the AB task relatively quickly, with the possible help of standard intervention procedures (e.g. fading). Their data is presented as Experiment 3. In this experiment the AB relation was tested for symmetry, and further training and testing eventually led to an evaluation of equivalence. The remaining 22 children did not learn AB matching through standard procedures, but the results of Experiment 3 suggested alternative methods for establishing the relation. These children were distributed between four more experiments (4(a), 4(b), 5 and 6) each of which involved an attempt to establish the AB relation (and its symmetrical counterpart, BA) indirectly, through common labelling of the Set-A and Set-B stimuli.

Before presenting each experiment in detail, a General Method section outlines the common features shared by all.

GENERAL METHOD

The general subject, apparatus and procedure specifications apply to all of the following experiments unless exceptions are noted.

SUBJECTS

Thirty 4-5 year old normal children were selected at random from those attending the nursery section of the Our Lady Roman Catholic School, Bangor. None of the children had previously participated in any psychology experiments.

APPARATUS

The experiments were conducted in a specially adapted room at the school. While experiments were in progress, the room was used for no other purpose. At one end of the room a table held an Apple microcomputer connected to a colour t.v. monitor. A five key stimulus response panel, identical to the one used in Experiment 1, was held in front of the monitor by mounting the panel in the centre of a larger wooden screen. This wooden screen was itself attached to the front edge of the table on which the equipment sat. The Apple computer programmed the sequencing and display of the stimuli, which appeared on the monitor screen behind any of the five clear perspex response keys. Any of eight possible stimuli could be presented, two colours (red and green, each drawn as a 4cm square) and six shapes (each drawn white on a black background to occupy a 4cm square area; precise details of the shapes are given in each experiment).

An additional wooden panel was attached to the left-hand side of the main wooden screen. This additional panel separated the experimenter from the child and also held a plastic matrix down which plastic 'coins' (token reinforcers) could be dropped. The matrix held and displayed up to 42 'coins' in seven columns of six. Also, at the top of the main wooden screen and directly above the response panel there was a glass fronted channel behind which small gift boxes could be displayed to the child. Figure 4.1 is a photograph of the apparatus as seen from the subject's viewpoint.

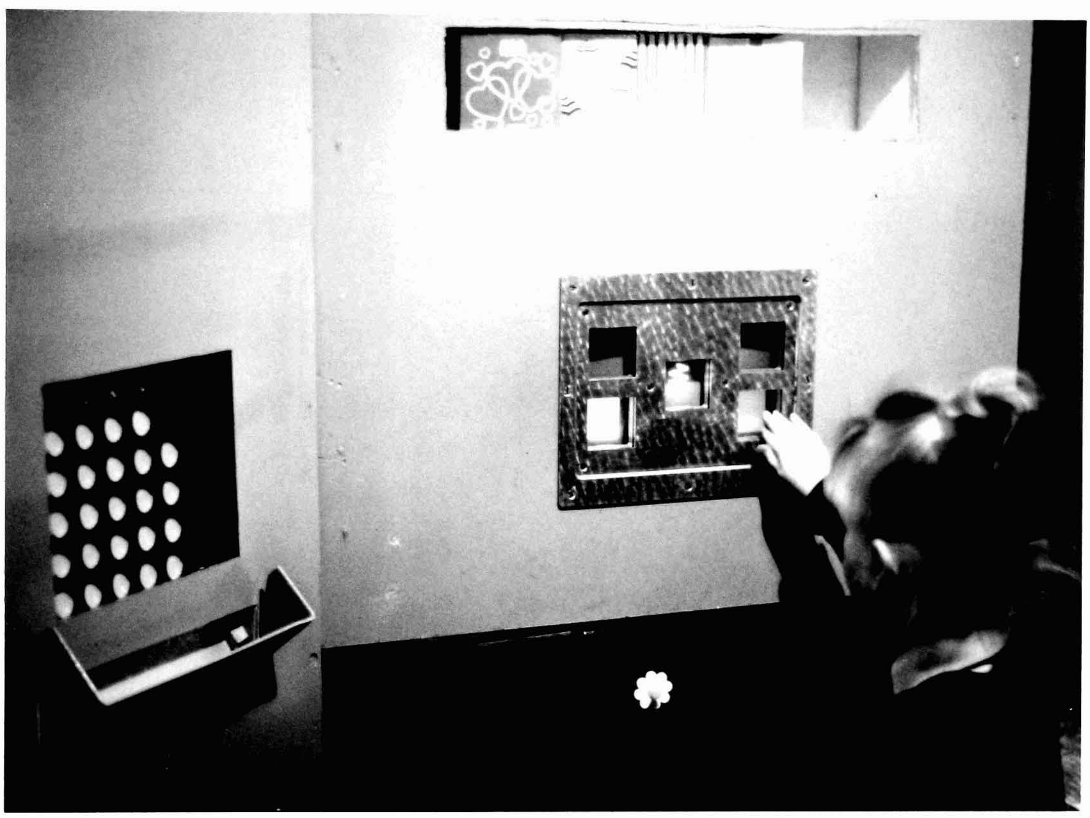
Although the experimenter was hidden from the subject's view during all but the first few trials of the introductory session, he could observe the child via a remote t.v. monitor situated behind the wooden partition. This monitor was itself connected to a video recorder and camera mounted at the opposite end of the room to the response panel. During test sessions the video link was disconnected and the experimenter's monitor was instead connected to the computer. This minimised the chances of the experimenter unwittingly affecting the subject's performance during testing. Although the experimenter could see the stimuli which were being displayed, he could not see the subjects or their responses during test sessions.

Finally, in one corner of the room there was a large box containing various gift-wrapped presents.

Procedure

The experimenter spent at least one day in the nursery with each of the subjects before introducing them individually to the experimental room. Once in the room, the subjects were sat in front of the response panel with their eyes level with the centre window. The experimenter (E) sat next to the subject for the first few trials of

Figure 4.1 A subject chooses a comparison during an arbitrary matching trial.



the introductory session and explained the general procedure. The children were told they were going to play a game, and the following instructions were given:

'In a little while you will see something come on here (E points to the middle window of the response panel; a few seconds elapse and the sample appears). When it comes on you must press it, like this (E presses the sample, and comparisons appear on two of the four outer windows). See, I pressed it and two more things have come on (E points to the two comparisons). If you press the right one, this will happen (E presses the correct comparison and drops a yellow token into the left-most column of the token display matrix). See, I got it right, so I got one of these tokens (E points to the token). You only get a token if you press the right one. If you press the wrong one you will not get a token. But if you keep getting it right, you will keep getting these tokens, like this (E drops four more yellow tokens into the left-most column of the matrix, leaving one blank space at the top of the column). Now, when you get to the top you will get a special token, like this (E drops a red token into the top space of the column). Every time you get one of these (E points to the red token) a box will come in here (E puts one gift box into the glass fronted channel above the response panel). There is a little present for you inside that box. When you have finished you can see what it is. You do some now. Try to get them right so you can get some more tokens and boxes.'

For the next few trials, the experimenter guided the child through the sequence of pressing the sample and correct comparison. Then, the experimenter said:

'You carry on now. I'm going to sit behind here.' (The E then sat behind the side panel, out of the subject's view, and watched the proceedings via the remote monitor).

For the remainder of the first session (and only the first session) the experimenter gave verbal feedback following each trial e.g. 'Yes, that was right, you get a token' or 'No, that was wrong, try again.'

Training Procedure

During training sessions, pressing the correct comparison produced a high-pitch tone from the computer and the delivery of a token. Incorrect comparison presses produced a low-pitch tone and no token. Once comparisons had appeared, all other key presses had no scheduled consequence. After a correct or incorrect choice the stimuli were removed (i.e. the screen went blank) and a five second inter-trial interval followed, at the end of which the next sample appeared. The computer selected trials on a random basis with the following restrictions:

(a) No more than three trials with the same sample could occur consecutively.

(b) The window occupied by the correct comparison on any given trial was left blank on the following trial.

With these exceptions all trial-types (sample-comparison combinations) and all correct windows were equally probable on successive trials.

Most trials were non-corrected i.e. errors did not cause trials to be repeated. However, correction was occasionally introduced in an attempt to break comparison preferences. If the subject began to select the same comparison on every trial, then incorrect responses repeated the trial. This contingency forced the subject to eventually switch to the non-preferred comparison in order to secure a reinforcer. Correction was only ever used during training.

At the end of each training session (normally 48 trials), the tokens dropped out of the matrix into a tray beneath. The subject was required to take the red tokens, which were then swapped for the boxes. The subject was allowed to open each box, which either contained a small toy (which the child could keep) or a small coloured star. The stars were glued into an exercise book and when the subject had filled the page with stars (approximately twelve stars to a page), the big toy box was opened and the subject was allowed to select a wrapped present. The children opened these presents and were allowed to play with them for a little while, but did not take them home until the end of the experiment or the end of school term, whichever came first.

Training was continued on each task until the subject was correct on at least six consecutive trials per trial-type. Once criterion was reached and maintained on each task, the subjects were taken through the next stage to prepare them for testing.

Reducing the reinforcement probability

In preparation for testing, the immediately preceding training trials were presented in the absence of reinforcement i.e. correct and incorrect responses were followed by the inter-trial interval only, and neither tokens nor tones were delivered. The subjects were given the following instructions prior to reducing the reinforcement probability to zero:

'This time the computer won't make any noises to tell you if you're right or wrong, and you won't see any tokens come down here (E points to the token matrix). Instead, I will keep your tokens behind the screen and you will get them all when you have finished. Just keep doing your best, and keep trying to get them right.'

At the end of this and all subsequent unreinforced sessions, the subjects were given a fixed number of red tokens (usually five) which they could then swap for the gift boxes. At no stage did the experimenter give feedback of any kind regarding the subject's performance on unreinforced trials.

All subjects maintained high levels of accuracy during this final stage before testing, so once reinforcement was terminated it did not need to be re-introduced in order to re-establish a subject's performance.

Test Procedure

All testing was conducted in the absence of reinforcement, and trial presentations were totally random. Details of specific stimulus presentations appear in the procedure sections of each experimental report.

Naming Tests

Each subject received at least one naming test during the course of the experiment. The stimuli were presented one at a time in the centre window and the child was asked a variety of questions aimed to evoke a verbal response e.g. 'What is it?', 'What is it called?', 'What is its name?', 'What do you call that?', etc. Normally, only one or two question forms are used but, given that responses to naming tests can often be idiosyncratic (see Hird and Lowe, 1985), it was felt that a larger variety of questions would give a more accurate impression of the subject's ability to name the stimuli. After each response the experimenter, seated behind the wooden screen, pressed a switch to initiate the inter-trial interval and the next presentation. The subjects were not required to press the windows, and no tokens or tones were delivered during the test.

<u>Procedures for analysing the children's spontaneous verbal behaviour</u>

During each session, the subject's behaviour (both verbal and non-verbal) was recorded on videotape. Although a great deal of spontaneous verbal behaviour was recorded, not all is presented in this thesis. The analysis was necessarily restricted to those aspects of verbal behaviour which were directly related to the subject's matching performance e.g. the childrens' comments regarding either the stimuli presented or the relations between those stimuli.

<u>Post-experimental language test</u>

At the end of the experiment the subjects were tested on the Reynell Developmental Language Scale (Reynell, 1977) which provided a measure of their language comprehension and production skills. Their scores on the test were expressed as 'age equivalents' (i.e. the age at which the majority of subjects from a standardised population achieved a given test score). The Reynell was administered at the end of the experiment so that the subjects would be more likely to perform to the best of their abilities, unimpeded by shyness or other characteristics which might be accentuated by a lack of familiarity with the experimenter.

EXPERIMENT 3

SUBJECTS

Four male and four female children took part. Table 4.1 lists the subjects by their chronological age at the time of testing, and by their equivalent language age from the Reynell Developmental Language Scale.

APPARATUS (See General Method)

PROCEDURE

The general sequence of training and testing is represented in Table 4.2. Each stage is explained more fully below (see the General Method section for additional details).

Train AB

The AB training trials were the same as those presented in Experiment 1 (see the left-hand side of Figure 4.2). Reinforcers were contingent upon choosing the green comparison in the presence of the Y sample, and the red comparison in the presence of the zig-zag sample. All subjects began with the same fading programme which apparently helped establish AB matching for two of the chimps in Experiment 1. If fading did not appear to work with the children, then any of three other intervention procedures were implemented. These were:

(a) Delayed cueing (see Experiment 1 for details)

(b) Instructions Some subjects were given the following instructions

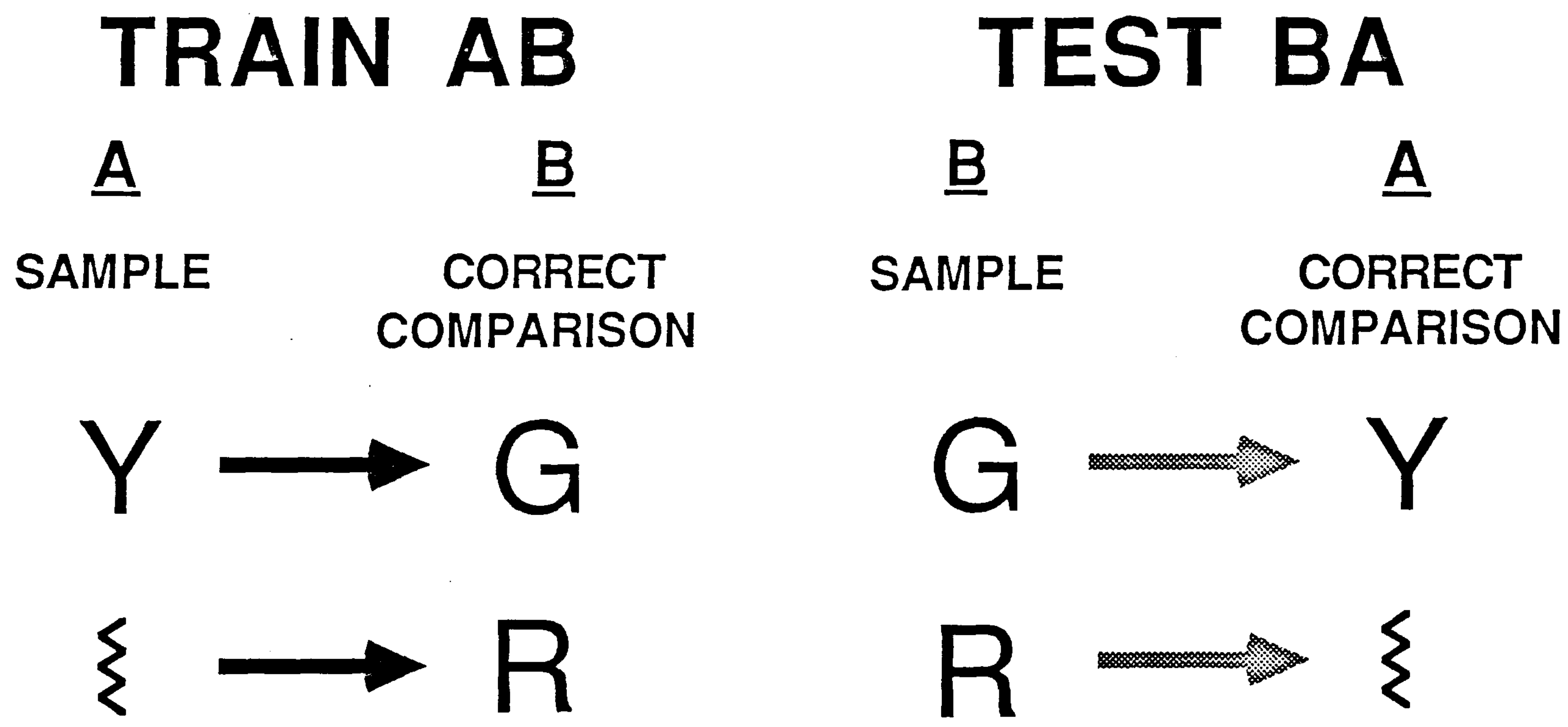
Table 4.1 Chronological age (in years-months) and equivalent language age (from the Reynell Developmental Language Scale) for the eight subjects in Experiment 3.

SUBJECT	CHRONOLOGICAL AGE	EQUIVALENT LANGUAGE AGE	
		VERBAL COMPREHENSION	EXPRESSIVE LANGUAGE
ANTONY	4 - 6	5 - 6	7 - 0
AMY	4 - 1	7 - 0	7 - 0
BECKY	4 - 4	5 - 6	4 - 11
GEMMA	4 - 2	5 - 8	7 - 0
HELEN	4 - 8	5 - 6	5 - 2
MATTHEW	4 - 10	5 - 6	6 - 6
MICHAEL W.	4 - 10	7 - 0	4 - 1
NICHOLAS	4 - 6	5 - 3	5 - 6

Table 4.2 Sequence of training and testing in Experiment 3.

1. Train AB
2. Reduce reinforcement probability
3. Test BA (symmetry) in baseline of AB
4. Train CA
5. Combine CA and AB
6. Reduce reinforcement probability
7. Test BC (equivalence) in baseline of CA and AB
8. Test CB (transitivity) in baseline of CA and AB
9. Test AC (symmetry) in baseline of CA
10. Test A, B, and C naming

Figure 4.2 Stimulus relations presented during AB training and BA test trials (see text).


TRAIN AB
A
B
SAMPLE
CORRECT COMPARISON
Y
G
R
TEST BA
B
A
SAMPLE
CORRECT COMPARISON
G
Y
R

both prior to and during standard AB sessions:

'The shape tells you which colour to choose'.

(c) Increased inter-trial interval (ITI) It was possible that some subjects were getting confused simply because the rate of trial presentations was too fast for them. So, where necessary, this rate was slowed down by increasing the ITI from 5 to 10 seconds.

Eventually, all subjects reached criterion on the AB task and proceeded to the next stage.

Reduce reinforcement probability

Having learned AB, the reinforcement probability was reduced to zero on AB trials (see General Method) to prepare the subjects for their first phase of testing.

Test BA (symmetry)

In this phase, the AB relation formed a baseline which was assessed for symmetry by the inclusion of BA test trials. On BA test trials, the original samples and comparisons were interchanged (i.e. a green or red sample was presented with Y and zig-zag comparisons, as depicted on the right of Figure 4.2). Two test sessions were given, each consisting of 32 AB baseline trials (16 per trial-type) and 16 BA symmetry test trials (8 per trial-type). All trials were unreinforced.

Train CA

After symmetry testing, the subjects were taught CA matching (see

Figure 4.3 Relations taught during CA training.

TRAIN CA

C SAMPLE

A CORRECT COMPARISON

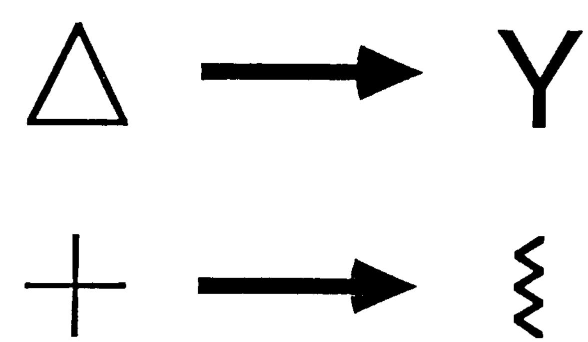


Figure 4.3). Reinforcers were contingent upon choosing the Y-shape comparison in the presence of the triangle sample, and the zig-zag comparison in the presence of the cross sample. All trials were standard; intervention procedures were not needed during CA training.

Combine CA and AB

During this phase, the CA and AB arbitrary matching trials were combined within the same 48-trial session (24 trials per task). Combined sessions continued until the subject met or exceeded criterion on both tasks.

Reduce reinforcement probability

The reinforcement probability was reduced to zero prior to presenting the final combined session of AB and CA trials. The subjects were required to maintain their AB and CA matching performances at criterion levels, and in the absence of differential reinforcement, before proceeding to the next stage.

Test BC (equivalence)

Each of two equivalence test sessions consisted of 32 baseline trials (16 AB and 16 CA) and 16 BC equivalence test trials. The test trials assessed equivalence because all three properties of reflexivity, symmetry and transitivity were required if BC was to emerge from the AB and CA baseline trials. The emergence of BC first required symmetry of the trained relations (AB and CA producing BA and AC). Then, provided the Set-B and Set-C stimuli could function in their new positions as samples and comparisons respectively (thus demonstrating reflexivity), the derived BA and AC relations could yield BC via transitivity (BA and AC producing BC).

Test CB (transitivity)

After equivalence was tested, two transitivity test sessions were administered. Each session consisted of 32 baseline trials (16 CA and 16 AB) and 16 CB transitivity test trials.

Test AC (symmetry)

Final conditional discrimination tests assessed symmetry of the trained CA relation. Once again two tests were given, each consisting of 32 CA baseline trials and 16 AC symmetry test trials.

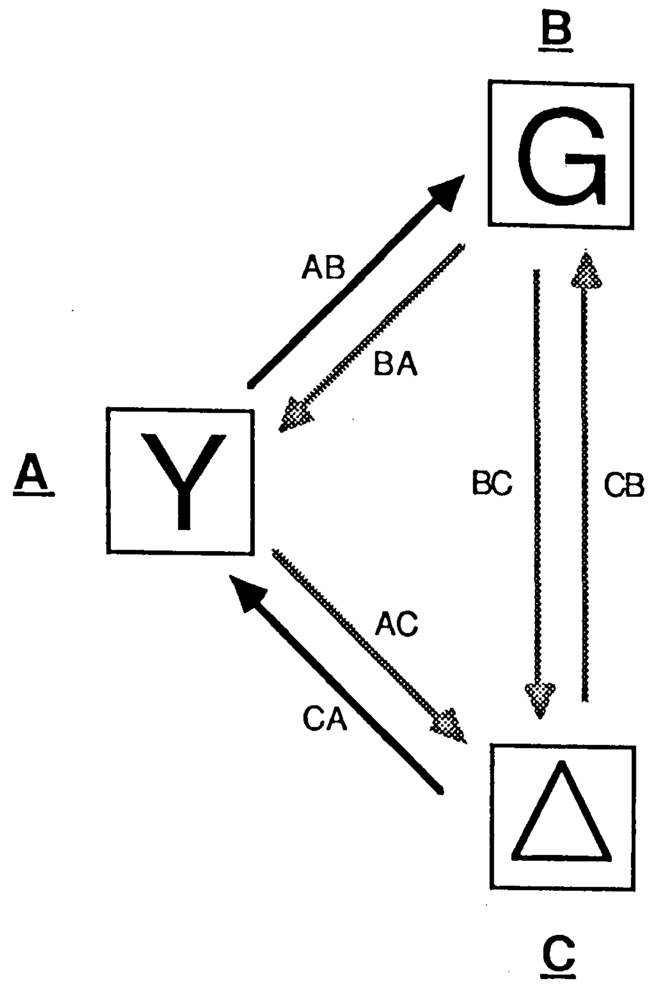
Summary

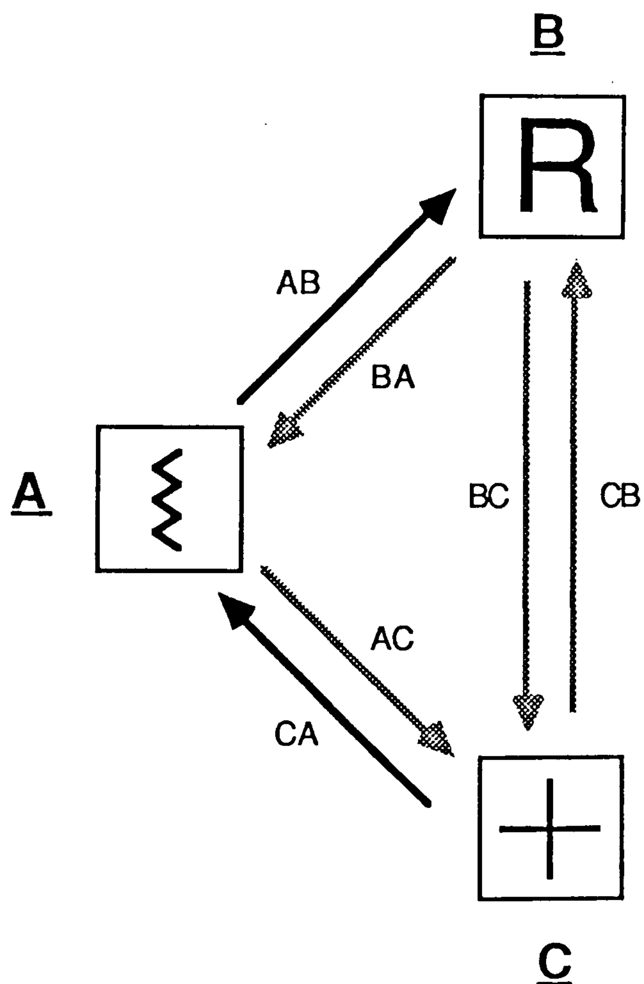
Figure 4.4 depicts all of the conditional relations which were trained and tested during the experiment. The solid arrows depict relations taught by explicit reinforcement whereas the shaded arrows represent relations which might emerge during unreinforced test sessions.

Naming Tests

After completing all of the above tests, each subject was asked to overtly name the visual stimuli. Each 24-trial naming test consisted of four presentations of each of the A, B, and C stimuli (see General Method for further detail).

Figure 4.4 The equivalence paradigm adopted in Experiment 3. The black arrows represent relations that were explicitly taught to the subject, and the shaded arrows represent relations that were tested after others had been explicitly taught.


B
G
AB
BA
A
Y
BC
CB
AC
CA
C


B
R
AB
BA
A
BC
CB
AC
CA
C

RESULTS

AB Training

All the subjects learned the AB relation. Table 4.3 lists the number of trials each subject received for each AB intervention. Only fading and instructions seemed to be effective in promoting learning of the AB task. The success of the fading programme depended upon whether the subject could initially match the sample colour-cue to its identical comparison colour. If so, and with further correct responses, the colour-cue would gradually shrink in size from around the sample shape until only the shape itself would be left to indicate which comparison colour was correct. Fading therefore proceeded from colourmatching to arbitrary (AB) matching. Two of the subjects, Gemma and Nicholas, were unable to benefit from the fading schedule because they did not even learn to colourmatch, despite receiving 168 and 251 trials respectively. At that stage the two children were transferred to another intervention procedure. The six other children all learned to colourmatch. Becky and Helen began colourmatching immediately, as did Antony, who made only one error in the first 12 trials. Matthew and Amy required only 48 trials to colourmatch, whereas Michael W. needed 147 trials. Of these six subjects, only Amy did not proceed to learn AB matching via the fading programme. Amy continued to match correctly up to and including the penultimate fading step, on which the sample colour cue was at its smallest. However, her responding fell to chance level on the final step of fading, on which no sample colour cue appeared.

Although fading did not prove effective for Amy, Gemma and Nicholas, all three quickly learned the AB relation after instructions were administered. (i.e. after the experimenter told them: 'The shape

Table 4.3 Number of trials per AB intervention for each subject in Experiment 3.

SUBJECT	TOTAL NO. AB TRIALS	TRIALS / INTERVENTION
MATTHEW	103	103 Fading
ANTONY	126	126 Fading
MICHAEL W.	203	203 Fading
BECKY	194	194 Fading
HELEN	218	218 Fading
AMY	255	230 Fading 25 Instructions
GEMMA	242	168 Fading 36 Increased ITI 38 Instructions
NICHOLAS	585	251 Fading 192 Delayed cueing 67 Normal 48 Increased ITI 27 Instructions

tells you which colour to choose').

Having learned the AB relation, all eight children received one AB session without reinforcement. All of the subjects continued to respond correctly in the absence of differential reinforcement, thus demonstrating that they were ready for their first test phase.

Test BA (symmetry)

Figure 4.5 shows the subjects' performances during symmetry tests. All subjects passed the symmetry test; they performed well above chance on both AB baseline and BA symmetry test trials. Overall, few errors were made.

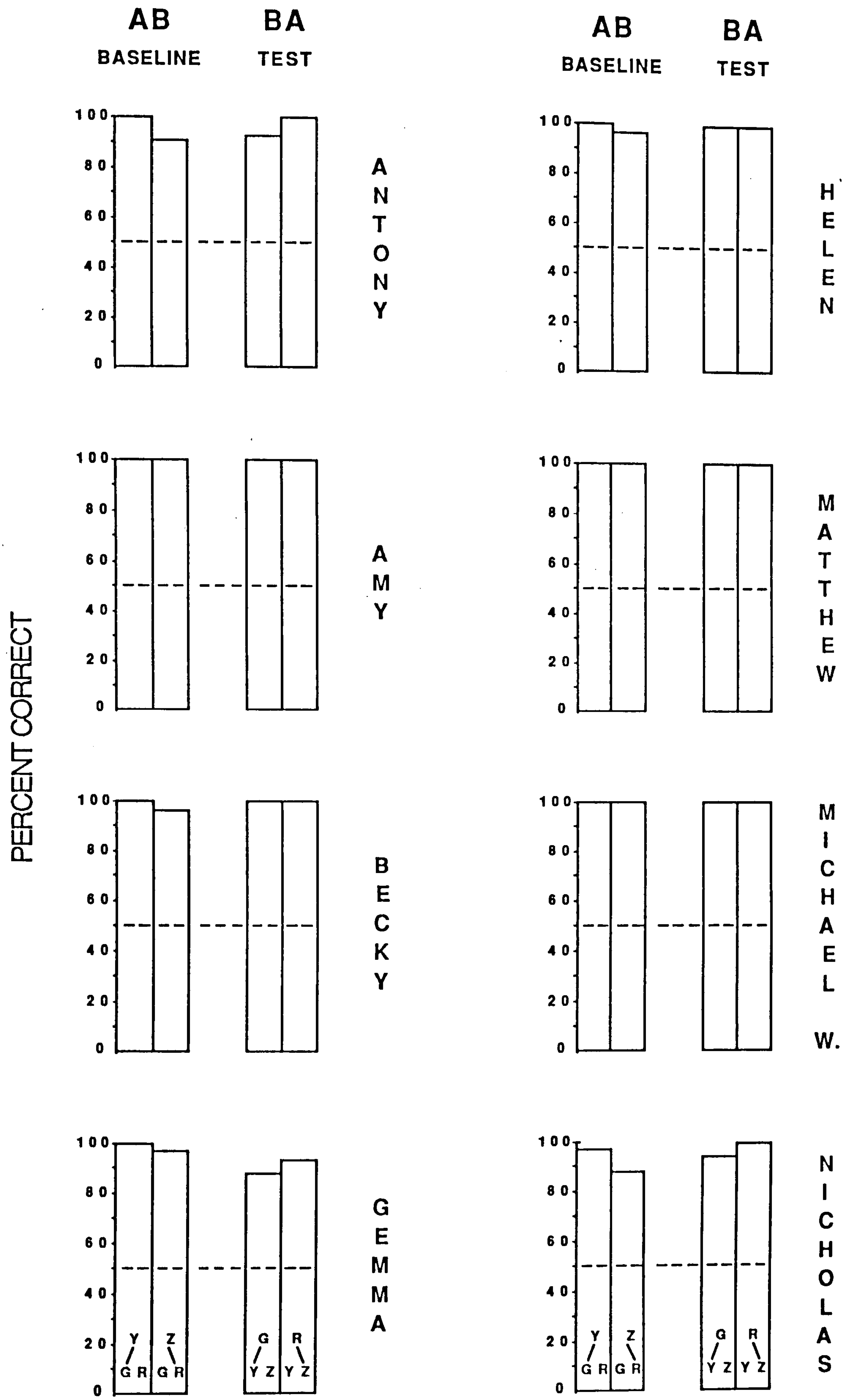
Train CA

Following their success on symmetry tests, the subjects were taught CA matching. All eight children learned CA extremely quickly; Becky required two 48-trial sessions, but all the other subjects learned CA matching within one session. Clearly, the number of trials needed to learn this second arbitrary matching problem was far less than the number required to learn the first.

Combine AB and CA

All subjects continued to respond correctly when the AB and CA trial-types were combined within the same session. Procedures were then implemented to reduce the reinforcement probability to zero, and one more combined session followed. The absence of differential reinforcement did not affect any of the subjects' AB and CA

Figure 4.5 Overall results (percentage of correct responses) on AB baseline and BA symmetry trial-types for the eight subjects in Experiment 3 (each baseline bar represents 32 trials, and each test bar represents 16 trials).


AB
BASELINE
BA
TEST
AB
BASELINE
BA
TEST
PERCENT CORRECT
100
80
60
40
20
0
ANTONY
HELEN
AMY
MATTHEW
BECKY
MICHAEL W.
GEMMA
NICHOLAS
Y
G R
Z
G R
G
Y Z
R
Y Z

performances, and so they all proceeded to the final test phases. Table 4.4 shows the number of trials each subject received in each phase leading to final testing.

Test BC (equivalence), CB (transitivity) and AC (symmetry)

Figures 4.6 and 4.7 depict the subjects' performances on BC (equivalence), CB (transitivity) and AC (symmetry) tests. Each subject's scores on unreinforced baseline and test trials were excellent throughout. These results may be translated into the number of errors each subject made. Out of a total of 96 test trials, Michael W. made only 2 errors, Amy and Becky each made only 3, Nicholas made 4, Matthew 6, Antony 10, Gemma 11 and Helen 17.

The results clearly show that two classes of equivalent stimuli had formed for each child tested; one class consisted of the Y-shape, green, and triangle stimuli and the other of the zig-zag, red, and cross.

Analysis of the children's verbal behaviour

Measures were taken of each child's verbal behaviour, both when it occurred spontaneously on matching trials and when it was prompted during a post-experimental naming test. The children all proved capable of labelling the stimuli at some point during the proceedings. However, the children's success on equivalence tests did not correlate with any single 'global' labelling strategy. Rather, a variety of labelling patterns were produced, some of which were perhaps more 'informative' than others.

Spontaneous labelling of the stimuli was virtually absent from some of the children's repetoires. Table 4.5 shows that Becky, for

Table 4.4 The number of trials presented to each subject in Experiment 3, for each phase prior to final testing. (Red.Rft. = reduce reinforcement probability)

SUBJECT	NUMBER OF TRIALS PER PHASE						
	Train AB	Red Rft	Test BA	Train CA	Combine AB & CA	Red Rft	Total
MATTHEW	103	48	96	24	48	48	367
ANTONY	126	48	96	24	48	48	390
MICHAEL W.	203	48	96	36	48	48	479
HELEN	218	48	96	24	48	48	482
GEMMA	242	48	96	28	48	48	510
BECKY	194	48	96	96	48	48	530
AMY	255	48	96	48	48	48	543
NICHOLAS	585	48	96	24	48	48	849

Figure 4.6 Overall results from BC (equivalence), CB (transitivity) and AC (symmetry) tests for four of the subjects from Experiment 3. Each test bar represents 16 trials.

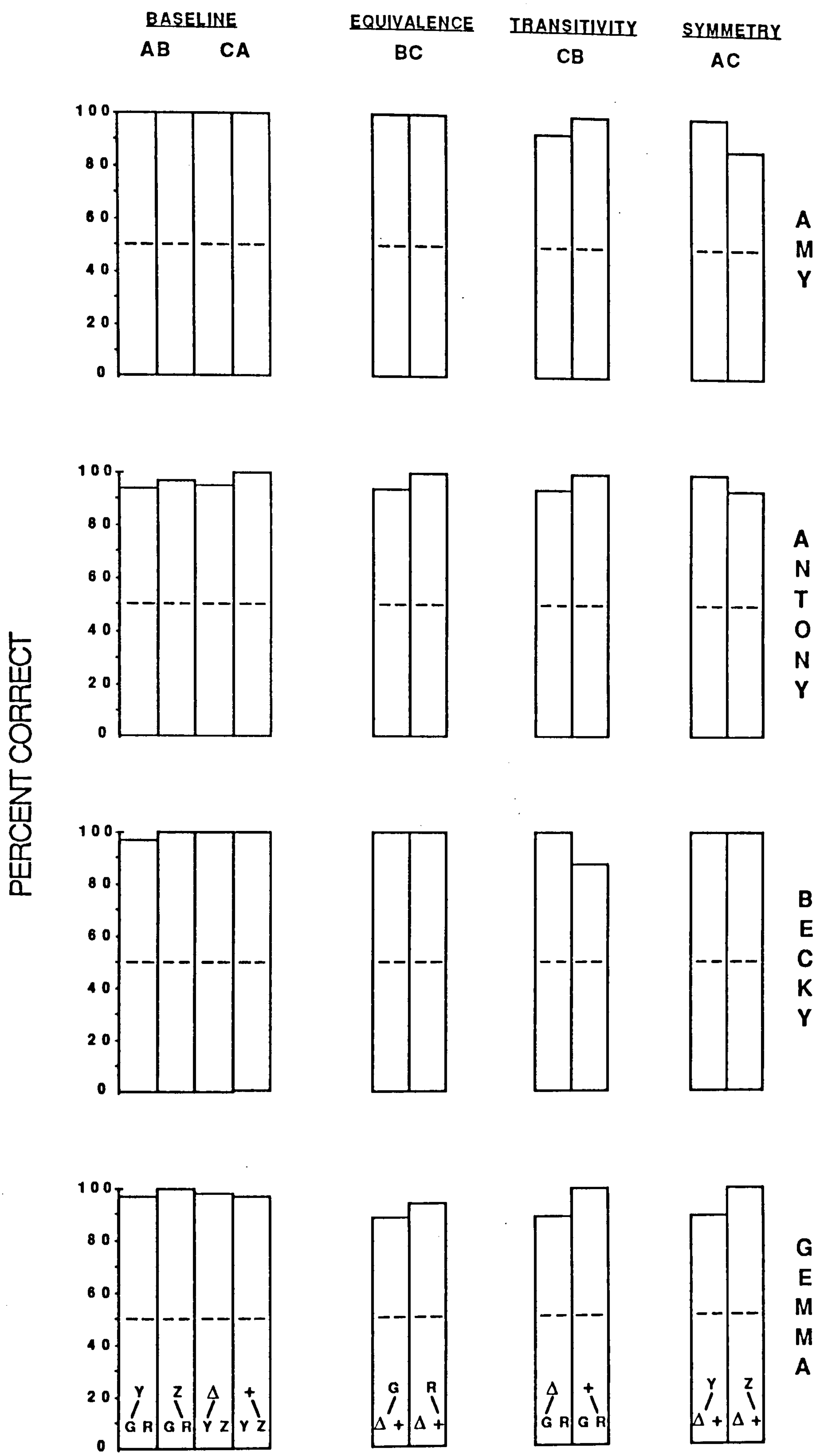

BASELINE
AB CA
EQUIVALENCE
BC
TRANSITIVITY
CB
SYMMETRY
AC
PERCENT CORRECT
100
80
60
40
20
0
AMY
ANTONY
BECKY
GEMMA
Y / G R
Z \ G R
Δ / Y Z
+ \ Y Z
G / Δ +
R \ Δ +
Δ / G R
+ \ G R
Y / Δ +
Z \ Δ +

Figure 4.7 Overall results from BC (equivalence), CB (transitivity) and AC (symmetry) tests for four of the subjects from Experiment 3. Each test bar represents 16 trials.

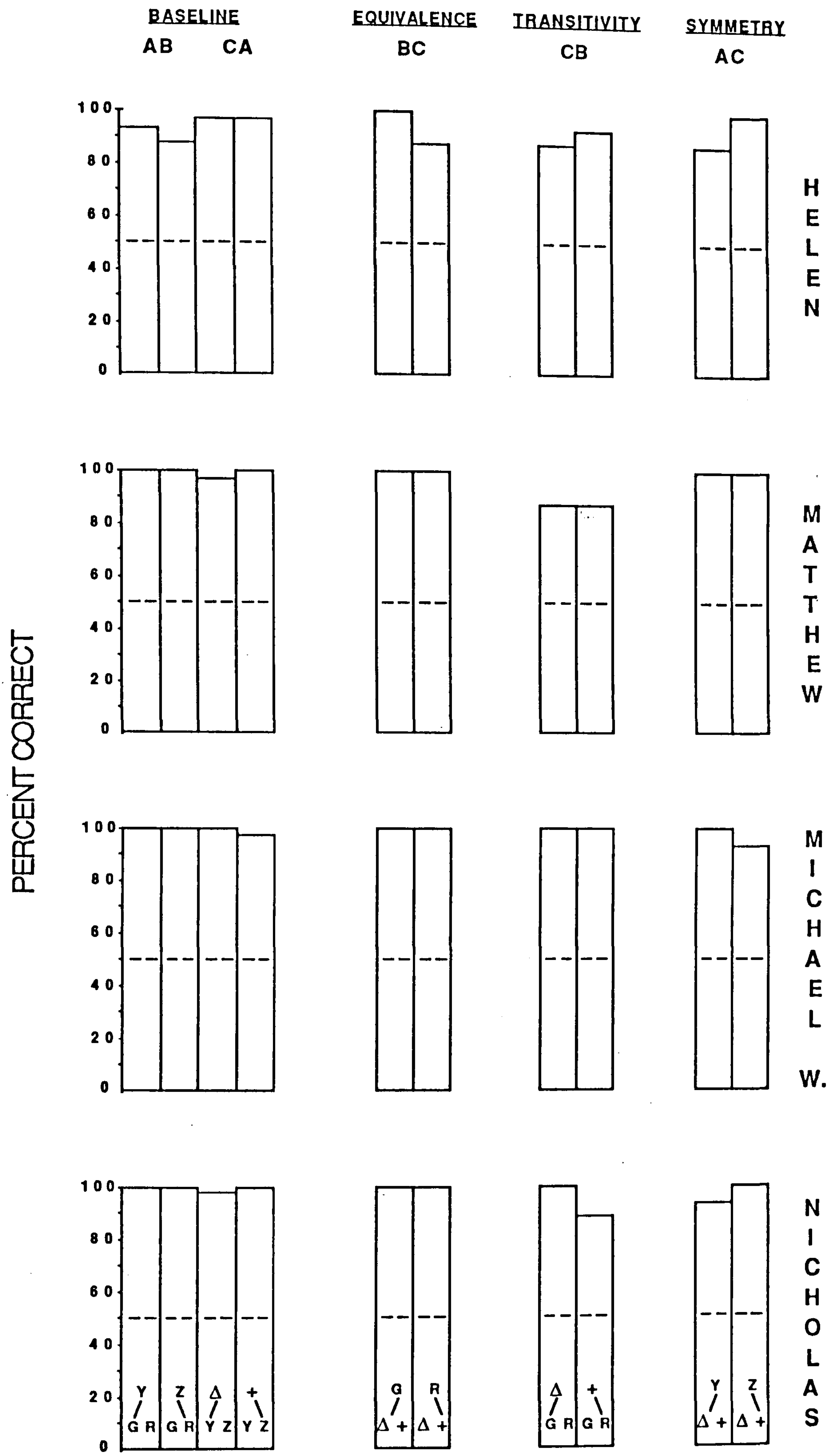

BASELINE
AB
CA
EQUIVALENCE
BC
TRANSITIVITY
CB
SYMMETRY
AC
PERCENT CORRECT
100
80
60
40
20
0
HELEN
MATTHEW
MICHAEL W.
NICHOLAS
Y / G R
Z \ G R
Δ / Y Z
+ \ Y Z
G / Δ +
R \ Δ +
Δ / G R
+ \ G R
Y / Δ +
Z \ Δ +

Table 4.5 Becky's verbal responses to the stimuli presented during matching trials and during her post-experimental naming test (Y=Y-shape, Z=zig-zag, G=green, R=red, Δ=triangle, +=cross; S=sample, CC=correct comparison, IC=incorrect comparison).

SPONTANEOUS VERBALISATIONS DURING MATCHING TASK

PHASE	S CC IC	VERBAL BEHAVIOUR	CONTEXT
CA	+-Z Y	'Cross'	When + appeared (no other stimuli present).

VERBALISATIONS DURING NAMING TEST

STIMULUS	RESPONSE
Y	No response (x2)
	'Lines' (x2)
Δ	'Triangle' (x4)
G	'A square... garage door... green'
	'Green' (x3)
Z	'Snake' (x4)
+	'Noughts'
	'Cross' (x3)
R	'Square garage....red'
	'Red' (x3)

instance, only produced one spontaneous label (she said 'Cross' in the presence of the corresponding shape). Becky did, however, give distinct labels to each of the stimuli during her naming test. Some of these prompted responses were, however, somewhat idiosyncratic. When the cross was first presented Becky said 'Noughts', but on subsequent presentations she gave it the same conventional label ('Cross') which she had previously produced spontaneously. The triangle, green and red stimuli were also given conventional names, although the colours were both initially labelled as square garages! Becky also labelled the zig-zag as 'Snake' and the Y-shape as 'Lines', although she was silent on the first two trials with the Y-shape.

Matthew did not overtly label any of the stimuli on matching trials. On the naming test, however, he consistently labelled the triangle, green and red stimuli with their conventional names (see Table 4.6). He also consistently named the zig-zag as 'Muh' (for Matthew or me). But his responses to the Y-shape and cross were perhaps the most significant. Although Matthew always gave the conventional name for the cross, on one occasion he additionally called it 'Red'. He also said 'Green' to the Y-shape after initially attempting to describe what it looked like. On these occasions, then, Matthew apparently labelled the cross and the Y-shape with the names of their corresponding colours.

Helen labelled the triangle, the cross, and the colours with their conventional names both spontaneously during matching trials (see Table 4.7) and when prompted on the naming test (see Table 4.8). She also gave the conventional name for the zig-zag, but only during the naming test. Helen never once overtly labelled the Y-shape spontaneously, but on the naming test her response to it was similar to Matthew's. She began by attempting to describe what the Y-shape looked like and then

Table 4.6 Matthew's verbal responses from his post-experimental naming test.

STIMULUS	RESPONSE
Y	'A straight stick with some things coming out' (x2).
	'It's green... I don't know the names of them, though'.
	'A stick, and it's got things coming out, and it's green'.
Δ	'Triangle' (x4)
G	'Green' (x4)
Z	'I don't know what that one is.. ....Muh. Muh for Matthew'
	'Muh'
	'Muh for Matthew'
	'Muh for me'
+	'I don't know what its name is... Cross'
	'Cross' (x2)
	'That one's red and it's cross'
R	'Red' (x4)

[illegible] spontaneous verbal responses to the stimuli presented during matching trials.

PHASE	TRIAL TYPE	S CC IC	VERBAL BEHAVIOUR	CONTEXT
Test BA	BA	R-Z Y	'It's red. Red in the middle. Oh, somethings happened'	When R appears in sample position for first time.
	BA	G-Y Z	'The green's in the middle'	When G appears in sample position.
Train CA	CA	Δ-Y Z	'What's that one?.. A triangle one'	When Δ appears for first time.
	CA	+-Z Y	'A cross in the middle'	When + appears for first time.

Table 4.6 Helen's verbal responses from her post-experimental naming test.

STIMULUS	RESPONSE	
Y	'It's a thing what goes down..... (tries to draw it in the air with her finger)... It's one of the greens'	
	'It's a thing that goes up and another thing that goes up. I don't know what this shape is, but it's one of the greens'	
	'Green'	(x2)
Δ	'Triangle'	(x4)
G	'Green'	(x4)
Z	'Zig-zag'	(x4)
+	'Cross'	(x4)
R	'Red'	(x4)

said 'It's one of the greens', a phrase which she repeated when the shape next appeared. Helen then went on to label the Y-shape as 'Green' on its final two presentations.

This ability to call a shape by its corresponding colourname was also exhibited by Michael W. Michael never once overtly labelled the stimuli on matching trials, but during his naming test he labelled the Y-shape, triangle and green stimuli as 'Green' and the zig-zag, cross and red stimuli as 'Red' (see Table 4.9). This, then, was a clear example of common naming. Equivalence tests had earlier revealed the presence of two classes of equivalent stimuli (Class 1 = Y-shape, triangle and green; Class 2 = zig-zag, cross and red). Michael's naming test responses were consistent with these stimulus classes; he applied his label for one of the class members (the colour) to each other class member (the shapes). The question was did Michael apply common labels to the stimuli prior to the equivalence test? Although he did not do so overtly, he may have done so covertly. If so, then common naming may have mediated the formation of stimulus equivalence i.e. the stimuli may have become equivalent precisely because they were given the same name ('green' or 'red').

Although Michael only produced common labels overtly during the naming test, others did so spontaneously, during the matching trials. One such subject was Antony (see Table 4.10). During AB training, Antony labelled the Y-shape as 'Green', even though no comparison colours were present at the time. In the absence of the colours, Antony also labelled the zig-zag as 'red-green' which was the same label he occasionally applied to red objects in his classroom (it had been noted prior to the experiment that Antony often used the terms 'red' and 'red-green' interchangeably to denote the colour red, whereas he always appeared to use the word 'green' to label green things). Antony also applied his colour labels spontaneously on CA trials, in

Table 4.9 Michael W.'s verbal responses from his post-experimental naming test.

STIMULUS	RESPONSE	
Y	'Can't think of the name'	
	'Green'	(x3)
Δ	'Green'	(x4)
G	'Green'	(x4)
Z	'Red'	(x4)
+	'Red'	(x4)
R	'Red'	(x4)

Table 4.10 Antony's spontaneous verbal responses to the stimuli presented during matching trials.

PHASE	TRIAL TYPE	S CC IC	VERBAL BEHAVIOUR	CONTEXT
Train AB	AB	Y-G R	'I know..this one's green'	When Y appears, (no other stimuli are present)
	AB	Z-R G	'I bet it's going to be red-green this time....' '...see, I was right, it is red-green'	Before Z appears...
Train AB	CA	+-Z Y	'Don't want green'	In between pressing + and Z.
	CA	Δ-Y Z	'Green' 'Green'	When Δ appears. Before pressing Y.
	CA	Δ-Y Z	'Gree-ee-een' 'It is green'	When Δ appears. Before pressing Y.
AB/CA	CA	+-Z Y	'I nearly pressed the green'	After nearly pressing the Y.
	CA	Δ-Y Z	'I nearly pressed the red'	After nearly pressing the Z.
Test BC	CA	Δ-Y Z	'I nearly touched the red-green'	After nearly pressing the Z.
	AB	Y-G R	'Green that one'	When Y appears, and no other stimuli are present.
	BC	G-Δ +	'Whoops-a-daisy, I did that wrong..'	During intertrial interval after having just pressed the incorrect comparison (+)
			'..When this green, yeh (points to blank centre window), I pressed the red-green (points to blank comparison window which had just held the +).	

Table 4.10 (cont'd)

PHASE	TRIAL TYPE	S CC IC	VERBAL BEHAVIOUR	CONTEXT
Test CB	CA	Δ-Y Z	'No not red'	As he goes to press the Z, which he subsequently avoids.
	CB	Δ-G R	'The red up there, the green down there, it was the green'	During intertrial interval after having just pressed the correct comparison (G).
Test AC	CA	Δ-Y Z	'What is it now... green'	Before pressing Y.
	AC	Y-Δ +	'Oooh, I know this is green'	In between pressing Y and Δ.
	AC	Y-Δ +	'Got green, green, green' (sings)	After pressing Δ.
	CA	Δ-Y Z	'Press the green ones'	In between pressing Y and Δ.
	CA	+-Z Y	'That one's red-green'	Just before pressing Z.
	CA	Δ-Y Z	'Green'	During intertrial interval, after he just pressed the correct comparison (Y).

which shapes alone were presented. For example, on one CA trial he said 'Green' to the triangle sample and then, after pressing it to produce the comparisons, he said 'Green' again just before pressing the Y-shape. On a later CA trial (one presented during an equivalence test session) he nearly pressed the zig-zag comparison by mistake (the zig-zag was the incorrect comparison) and afterwards said 'I nearly touched the red-green'. During the same session he appeared to use the same label ('red-green') to denote the cross. This occurred on a BC test trial, in which the green sample was presented with the triangle (correct comparison) and the cross (incorrect comparison). Antony made a rare error by choosing the cross, but then, when all the stimuli had disappeared from the screen, he said:

'Whoops-a-daisy, I did that wrong...... When this green, yeh (pointing to the blank centre window which had just held the green stimulus) I pressed the red-green' (pointing to the blank comparison window which had just held the cross).

The examples given above were by no means the only ones Antony produced. Inspection of Table 4.10 leaves little doubt that Antony was spontaneously labelling the Y-shape, triangle and green stimuli as 'green' and the zig-zag, cross and red stimuli as 'red-green' (or occasionally 'red'). In addition, most of his naming test responses were consistent with the labels he had earlier produced spontaneously. During the naming test he said 'green' to the two shapes which had become equivalent to the green stimulus, and 'red-green' to the other two shapes which had become equivalent to red (see Table 4.11).

Gemma also seemed to give common names spontaneously to each of the stimuli (see Table 4.12). After learning AB matching, she labelled

Table 4.11 Antony's verbal responses from his post-experimental naming test.

STIMULUS	RESPONSE	
Y	'I don't know what its name is'.	
	'Green'	(x3)
Δ	'Green'	(x4)
G	'Green'	(x4)
Z	'I don't know what the names are. I think it is a picture'.	
	'We did that one! I think it is.. I don't know what it is..I don't know what its name is, but I think it's red-green'.	
	'Red-green'	(x2)
+	'I don't know what its name is'.	
	'Red-green'	(x3)
R	'Red'	(x4)

Table 4.12 Gemma's spontaneous verbal responses to the stimuli presented during matching trials.

PHASE	TRIAL TYPE	S CC IC	VERBAL BEHAVIOUR	CONTEXT
Train AB	AB	Z-R G	'I think it's going to be red'	When Z appears (no other stimuli are present).
	AB	Z-R G	'Red again'	When Z appears (no other stimuli are present).
	AB	Y-G R	'Green!'	When Y appears (no other stimuli are present).
CA/AB	CA	Δ-Y Z	'I know which colour to do'	Before pressing Y.
Test BC	CA	Δ-Y Z	'Green again'	When Δ appears.
	CA	+-Z Y	'What colour now? Press red'	Before pressing Z.
	BC	R-+	'That's red' 'Reds!'	When R appears Before pressing Z.
	BC	R-+	'That's red'	Before pressing +.
	CA	+-Z Y	'It's the red things'	After pressing Z.

the zig-zag as 'red' and the Y-shape as 'green' even though the colours were not present at the time. Later on, during a CA trial (in which shapes alone were presented) she said 'I know which colour to do', and then proceeded to select the correct comparison shape. Later still, during equivalence testing, she was presented with a CA baseline trial and she said 'It's the red things' after pressing the cross sample and the zig-zag comparison.

Although Gemma clearly applied common labels to the stimuli during matching trials, she did not continue to do so when she was later given a naming test. Gemma in fact spent most of the naming test giggling to herself and, whilst she labelled the Set-B colours appropriately, she labelled the shapes somewhat indiscriminately with what appeared to be novel 'nonsense' words (see Table 4.13).

Common labelling appeared to be absent from Amy's verbal repetoire (see Table 4.14), but her verbal behaviour suggested another way in which language may promote stimulus equivalence. Amy not only labelled each of the stimuli with distinct names but, in addition, she also overtly labelled the relations between the stimuli. After learning AB matching, and before her last AB session prior to symmetry testing, she said:

'The squiggly line is for the red and the other one's for the green' (the other one, presumably, was a reference to the Y-shape).

Amy's statement was consistent with the possibility that the stimuli had already become equivalent prior to any formal equivalence test (cf. Sidman et al, 1985). Furthermore, at one stage Amy seemed to be trying to relate some of the stimuli verbally, even before she had been given an opportunity to match them on the screen. When presented with a CA trial for the first time, and prior to selecting a

Table 4.13 Gemma's verbal responses from her post-experimental naming test.

STIMULUS	RESPONSE	
Y	'Stripey'	(x2)
	'Liney'	(x2)
Δ	'Stairs'	
	'Stairsey'	
	(Giggling)	(x2)
G	'Green'	(x4)
Z	'Stairs'	
	'Stripey'	
	(Giggling)	(x2)
+	'Stripey'	(x2)
	(Giggling)	(x2)
R	'Red'	(x4)

Table 4.14 Amy's spontaneous verbal responses to the stimuli presented during matching trials.

PHASE	TRIAL TYPE	S CC IC	VERBAL BEHAVIOUR	CONTEXT
Train AB			'The squiggly line is for the red and the other one's for the green'	Comment made before session began (the last AB session prior to BA symmetry test).
Train CA	CA	+-Z Y	'There is a cross. I wonder what does the cross mean?'	First time + appears.
			'That seems funny to my mind...what does the cross mean?'	Before pressing Z.
	CA	Δ-Y Z	'There's a triangle in the middle with two shapes at the side as well'	Before pressing Z.

comparison, she said:

There is a cross' (referring to the sample). 'I wonder what does the cross mean?'

and then, after pressing the cross sample to produce the Y and zig-zag comparisons, she added:

'That seems funny to my mind.....what does the cross mean?'

Presumably she then found out what the cross meant, because she pressed the zig-zag comparison, which happened to be correct, and then proceeded to make only three errors in the next 48 CA trials.

On the naming test, Amy continued to give distinct names to each of the stimuli, but not always in accordance with the names she had previously produced spontaneously. Table 4.15 shows that she called the zig-zag a 'ziggy line' and a 'ziggy-zaggy line' during the naming test (whereas she had previously labelled it spontaneously as a 'squiggly line'), and despite referring to the Y-shape as 'the other one' during matching trials, she consistently called it a 'Y' during the naming test.

The final subject to consider is Nicholas. Nicholas's spontaneous labelling was particularly interesting because it appeared to consist of a combination of both common labelling and relational labelling. Examples are given in Table 4.16.

During the course of learning CA matching, the triangle sample appeared and Nicholas said:

'Is it green now? That one says green, the triangle, yeh?'

This was a remarkable statement because not only was Nicholas apparently attempting to relate the stimuli verbally (through the relational term, 'says') but he was also producing the label for a

Table 4.15 Amy's verbal responses from her post-experimental naming test.

STIMULUS	RESPONSE	
Y	'That looks a bit like a letter Y'	
	'Y'	(x3)
Δ	'Triangle'	(x4)
G	'A green square'	(x4)
Z	'It looks a bit... I don't know what it is...I can just call it a ziggy line'	
	'A ziggy-zaggy line	(x3)
+	'Cross'	(x4)
R	'A red square'	(x4)

Table 4.16 Nicholas's spontaneous verbal behaviour to the stimuli presented during matching trials.

PHASE	TRIAL TYPE	S CC IC	VERBAL BEHAVIOUR	CONTEXT
Train AB	AB	Z-R G	'There's two reds'	Points to Z and R before pressing R.
	AB	Y-G R	'These are both green'	Before pressing G.
Train CA	CA	+-Z Y	'What's that doing there for?' 'Is it green?'	+ appears for the first time. Before pressing Y.
	CA	+-Z Y	'Red, yeh'	In between pressing + and Z.
	CA	Δ-Y Z	'Is it green now? That one says green, the triangle, yeh?'	When Δ appears.
	CA	+-Z Y	'That looks like Jesus's cross yeh?'	When + appears.
	CA	+-Z Y	'I thought it was the green, then I found red'	After nearly pressing Y.
	CA	+-Z Y	'I get the reds, yeh'	After pressing + and Z.
AB/CA			E: 'They're now coming on. Let's see if you can remember? S: 'Yeh, greens to greens and reds to reds'	Verbal behaviour recorded prior to Nicholas's first session following a one month break.
	CA	+-Z Y	'Green'	Inbetween pressing + and Y (incorrect).

Table 4.16 (cont'd)

PHASE	TRIAL TYPE	S CC IC	VERBAL BEHAVIOUR	CONTEXT
	CA	+-Z Y	'If that was green (points to blank comparison window which had just held Y), that was red' (points to blank sample window which had just held +)	During intertrial interval, after having just pressed the incorrect comparison (Y).
AB/CA	CA	Δ-Y Z	'When I see the triangle that means green'	During intertrial interval, after having just pressed the correct comparison (Y).

colour which was not present at the time, and to which the triangle had never been paired directly via matching. This kind of relational labelling was recorded on one other occasion, on one of the CA trials which appeared during a combined AB and CA session (see the final entry in Table 4.16). After correctly matching the triangle sample to the Y comparison, and during the intertrial interval which followed, he said:

'When I see the triangle, that means green'.

(Once again, no colours were present at the time).

This relational labelling appeared to co-exist with common labelling, the latter being the more predominant of the two patterns. For example, on the last AB training session prior to symmetry testing, Nicholas pointed to the zig-zag sample and corresponding red comparison and said 'There's two reds'. He also said 'These are both green' in the presence of the Y-shape sample and corresponding green comparison. The impression given was that Nicholas, too, had somehow learned to label each Set-A shape with the name of its corresponding Set-B colour. This notion is consistent with his spontaneous labelling during subsequent CA trials. It also seemed that in the process of learning CA matching, the colournames which were initially applied to the Set-A shapes were somehow passed on to the Set-C shapes. For example, after pressing the cross sample and corresponding zig-zag comparison, Nicholas said 'I get the reds yeh?'. This in fact occurred on Nicholas's last session before a one-month break (the holiday between school terms). Upon returning, and just before presenting the first stimulus, the experimenter said 'They're now coming on. Let's see if you can remember', to which Nicholas replied, 'Yeh, greens to greens and reds to reds'. Nicholas produced other examples of common labelling, all of which appear in Table 4.16.

Furthermore, the labels which Nicholas produced spontaneously corresponded well with those he produced during the naming test (although a few of his naming test responses were somewhat idiosyncratic; see Table 4.17).

DISCUSSION

In this experiment, eight 4-5 year old children passed standard tests of stimulus equivalence. If each subject's non-verbal matching performance had been the only behaviour of interest, then one would have perhaps been none the wiser about why equivalence emerged. But this study did not just confine itself to monitoring the subject's non-verbal behaviour on matching trials; provision was also made for the detailed monitoring of each subject's verbal behaviour. To the author's knowledge, only one other study has given equal weight to the verbal and non-verbal behaviour of children during trials leading up to, and including, equivalence testing. Beasty (1988) found that the way in which his subjects labelled the stimuli had a direct bearing on their subsequent equivalence test performance. Only those children who labelled the sample and corresponding comparison in sequence then went on to pass the tests.

Although extremely informative, Beasty's study (like all good experiments) raises many interesting questions. Perhaps the foremost of these is what is it about sequential naming which makes for the effective formation of stimulus equivalence?

None of the subjects in the present experiment were heard to produce sequential naming of the sort described by Beasty. However, some (e.g. Becky, Matthew and Helen) did overtly label each stimulus with a distinct name, so one cannot rule out the possibility of

Table 4.17 Nicholas's verbal responses from his post-experimental naming test.

STIMULUS	RESPONSE
Y	'That hasn't got any names'
	'Don't know....green'
	'Green' (x2)
Δ	'Bunk bed' (Nicholas told the experimenter about his new bunk bed just before the test began)
	'Triangle' (x2)
	'It's a triangle and it's green'
G	'Green' (x3)
	'Another square and it's green'
Z	'Like in a hospital'
	'I don't know...Red'
	'Red' (x2)
+	'Cross'
	'Cross...red again'
	'Red' (x2)
R	'It's a square and it's red'
	'Red' (x3)

sequential naming at the covert level in these children. Others, while not producing straightforward sequential naming, did produce other patterns which may be related to (and indeed may give some insights into the effectiveness of) the sequential naming which Beasty reported.

Amy, for example, not only gave distinct names to each individual stimulus but she also labelled the relation between the stimuli. Nicholas also gave some evidence of relational labelling. Relational labelling is significant precisely because it makes explicit one way in which stimuli may become equivalent. If, for example, a subject matches A to B and C to A according to an arbitrary relation such as 'is for', then one should not be too surprised if that subject proceeds to pass equivalence tests. If A 'is for' B and C 'is for' A then the subject may well conclude that B 'is for' C and C 'is for' B.

The interest in arbitrary relational responding has recently been championed by Steven Hayes (see Hayes, in press

). Hayes's theory of 'relational frames' is quite complex, and judging from successive papers on the topic it still appears to be under development, so a detailed examination of the theory will not be presented here. But briefly, when Hayes talks of relational frames he refers to arbitrary relations like those mentioned above. He also suggests that an arbitrary relation may itself emerge after reinforcing a subject's appropriate responses to many exemplars of the relation (but see Experiment 2 of this thesis). Furthermore, this 'training history' is said to occur in a particular context, and in all of the examples which Hayes provides, this context is a linguistic one. Hayes appears to draw short of saying that arbitrary relations are necessarily linguistic in origin, but this would not seem an unwarranted claim given the current weight of empirical evidence in support of a relationship between language proficiency and success on equivalence tests.

Although beyond the scope of the present thesis, future research may well focus on the origins of relational verbal behaviour and the conditions under which it may promote stimulus equivalence. Indeed, such research already appears to be underway (see Sofroniou, 1988). Certainly, verbal behaviour from subjects like Amy and Nicholas would seem to indicate the importance of such an undertaking. Furthermore, the sequential naming produced by Beasty's subjects may have involved an element of arbitrary relational control. Certainly the phrase 'Up-Green', for example, could conceivably operate as an abbreviated version of 'Up is the same as Green'. Indeed, some of Beasty's subjects did produce relational labels spontaneously during their matching trials.

However, the results from the present experiment suggested other ways in which naming may promote stimulus equivalence. Four of the subjects produced common names for the stimuli. Three of the subjects, namely Antony, Gemma and Nicholas, produced common labels spontaneously during the matching trials. One other, Michael W., only produced common labels overtly during his naming test, although he may have been producing them covertly prior to this. It was possible, then, that all four subjects passed the equivalence tests as a consequence of applying the same name to each prospective equivalent stimulus.

If common naming mediated equivalence formation, then where did the common naming come from? How did the subjects learn to label each shape with the name of its corresponding colour? During AB training, subjects were reinforced for pressing the green comparison, but only in the presence of the Y sample. If the subject also named the colours as he chose them, then a reinforcer would only follow saying 'Green' in the combined presence of the Y-shape and green stimuli. The Y-shape may have therefore gained some control not only over the subject's choice of green comparison but also over his label for that colour. Then, on CA training trials, the word 'Green', produced by the subject in the presence of the Y-shape comparison, could by similar means come under the control of the triangle sample. The same mechanism could bring about naming of the zig-zag, cross and red stimuli with the common label, 'Red'. Once established in this way, common naming could mediate correct matching responses during unreinforced tests of stimulus equivalence.

But this need not be the only way in which common labelling could develop. There is an alternative possibility which involves turning all of the previous argument completely on its head. Common labelling may have produced stimulus equivalence. But, alternatively, stimulus equivalence may have produced common labelling. Prior to the experiment, each subject could name the Set-B colours. If the names are denoted as Set-X responses then the colour naming relation can be denoted as BX. During the experiment, each subject was taught AB. Now, given AB and BX it is possible that AX could emerge through transitivity (AB and BX producing AX). Then, when the subject subsequently learns CA, CX too could emerge through transitivity (CA, and AX producing CX). The end result would be common labelling; Set-A, Set-B and Set-C stimuli would all be labelled appropriately with the Set-X colour words, 'Green' and 'Red'.

So, did common labelling produce equivalence, or was common labelling merely a by-product of extant equivalence processes? There seems to be nothing in the present data which could answer this question. Nevertheless, the present experiment had served a purpose; it confirmed that common labelling was worthy of further investigation, and thus it suggested a format for further experiments, reports of which now follow.

EXPERIMENT 4(a)

In the previous experiment, some subjects spontaneously applied the same label to unidentical stimuli prior to matching those stimuli on equivalence test trials. Although suggestive, this finding does not clearly define the role of common labelling in the formation of equivalence. The common labelling may have been responsible for the emergence of equivalence, but it is also possible that equivalence emerged independently of common labelling, the latter in turn being a product of extant equivalence processes.

Subsequent attempts to teach AB matching met with little success. Although this posed difficulties for the experimental programme, it also presented an opportunity to circumvent the problems of interpretation mentioned above. The subjects could be taught to apply common labels to the very stimuli they had failed to relate via the AB matching task. If the AB relation emerged, together with its symmetrical counterpart BA, without differential reinforcement and immediately after common labels had been established, then one would have clear evidence for the facilitative effect of common labels on the formation of stimulus equivalence.

METHOD

The general plan was:

(1) Identify a group of subjects who fail to learn AB matching.

(2) Teach the subjects to label the Set-A stimuli with the Set-X words 'Omni' and 'Delta'.

(3) Return the subjects to the AB baseline to establish whether their new labelling skills would help them acquire the matching task. If

not, then:

(4) Teach the subjects to label the Set-B stimuli with the Set-X words.

(5) Test the effects of common labelling by the subjects on both AB and BA matching (for further details see Procedure).

One subject, however, was not taught to produce common labels but rather to select Set-A and Set-B comparisons conditional upon common labels dictated by the experimenter. This subject's initial test performance was such that further test sessions were required, and at that stage the subject was prompted to produce common labels. Other than this, all subjects followed the general plan, although exact routes to testing were determined by each subject's performance at particular stages of the experiment (see below).

SUBJECTS

Four male and four female children took part. Table 4.18 lists subjects by their chronological age and by their equivalent language age according to the Reynell Developmental Language Scale.

APPARATUS

Details of apparatus may be found in the General Method section.

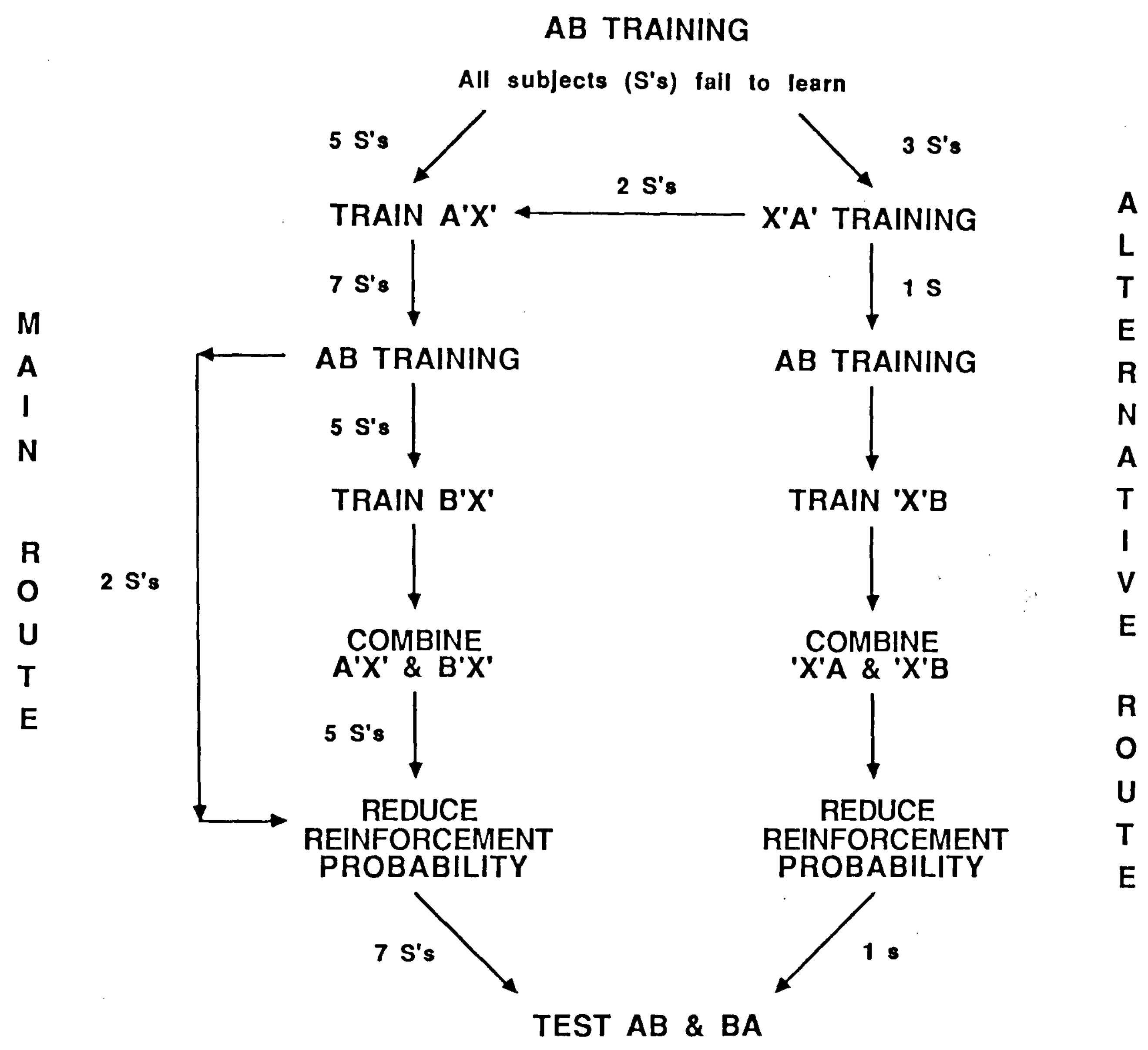
PROCEDURE

The general sequence of training and testing is schematically represented in Figure 4.8. Each stage is explained more fully below. The tasks presented in each stage were judged to have been learned when

Table 4.18 Chronological age (in years-months) and equivalent language age (from the Reynell Developmental Language Scale) for the eight subjects in Experiment 4(a).

SUBJECT	CHRONOLOGICAL AGE	EQUIVALENT LANGUAGE AGE	
		VERBAL COMPREHENSION	EXPRESSIVE LANGUAGE
ALEX	4 - 11	4 - 10	5 - 0
FRANCIS	4 - 7	4 - 6	5 - 3
JESSICA	4 - 8	7 - 0	7 - 0
LINDA	5 - 0	4 - 3	7 - 0
MICHAEL P.	4 - 11	4 - 9	7 - 0
NICHOLA	4 - 3	5 - 7	7 - 0
SARA R.	4 - 9	7 - 0	7 - 0
STEPHEN	4 - 8	5 - 7	7 - 0

Figure 4.8 Sequence of training and testing in Experiment 4(a) (see text).


AB TRAINING
All subjects (S's) fail to learn
5 S's
3 S's
2 S's
TRAIN A'X'
X'A' TRAINING
7 S's
1 S
MAIN ROUTE
AB TRAINING
AB TRAINING
5 S's
TRAIN B'X'
TRAIN 'X'B
2 S's
COMBINE A'X' & B'X'
COMBINE 'X'A & 'X'B
5 S's
REDUCE REINFORCEMENT PROBABILITY
REDUCE REINFORCEMENT PROBABILITY
7 S's
1 s
TEST AB & BA
ALTERNATIVE ROUTE

the subject had satisfied a criterion of at least six consecutive correct responses per trial-type.

Training Procedure

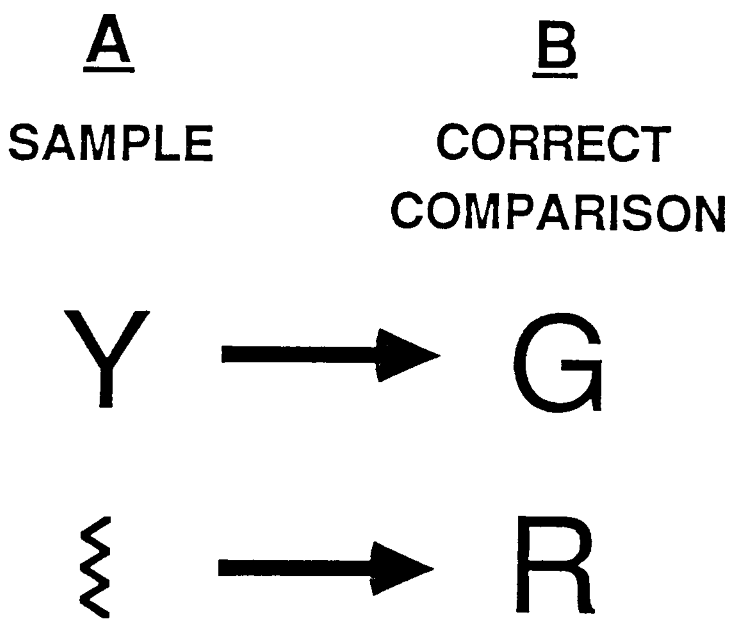
AB Training All subjects began with AB arbitrary matching trials. On each trial the sample was either the Y-shape or the zig-zag, and the comparisons were the green hue and the red hue. Reinforcers were contingent upon choosing the green comparison in the presence of the Y sample, and the red comparison in the presence of the zig-zag sample (see Figure 4.9).

In an attempt to promote learning of the AB relation, several intervention procedures were superimposed upon the basic AB matching trials. Details of some of these interventions can be found elsewhere (see Experiment 3). Other interventions, namely sample schedules and reward reduction, were introduced for the first time during this experiment. When sample schedules were in effect, the subjects were required to press the sample five times to produce the comparisons. It was hoped that this would increase the likelihood that the subject would attend to the sample. When reward reduction was in effect, incorrect comparison choices resulted in the removal of a token. This was implemented because some subjects seemed to be satisfied with the number of tokens obtained from chance level performance. When sessions incorporated reward reduction, the subject had to score better than chance in order to earn any tokens.

Having failed to learn the AB relation, all subjects proceeded to testing via one of two possible routes, depicted in Figure 4.8. The route represented down the left side of the figure was eventually taken by all but one subject (Jessica), and is therefore called the main route. The alternative route depicted on the right

Figure 4.9 Stimulus relations presented during AB training trials.

TRAIN AB



side of the figure was initially taken by three subjects but, due to a lack of progress, two of these (Linda and Sara R.) were eventually transferred to the main route, leaving only one subject (Jessica) to continue along the alternative path. There were important theoretical reasons for the inclusion of this alternative route to testing, even if it was followed by only one subject, and these reasons will become clear in the Discussion section. Both routes are explained below.

Main Route

(1) Train A'X' On each trial either the Y-shape or the zig-zag appeared in the centre window. Subjects were required to say 'Omni' in the presence of the Y-shape and 'Delta' in the presence of the zig-zag (see Figure 4.10). Pressing the sample then produced reinforcers. If the subject said the wrong word then sample presses produced the incorrect tone followed by the intertrial interval. A sample press produced a consequence only if preceded by the spoken word 'Omni' or 'Delta'. All other presses had no scheduled consequence. Just before the first session of A'X' training, the subjects were given the following instructions:

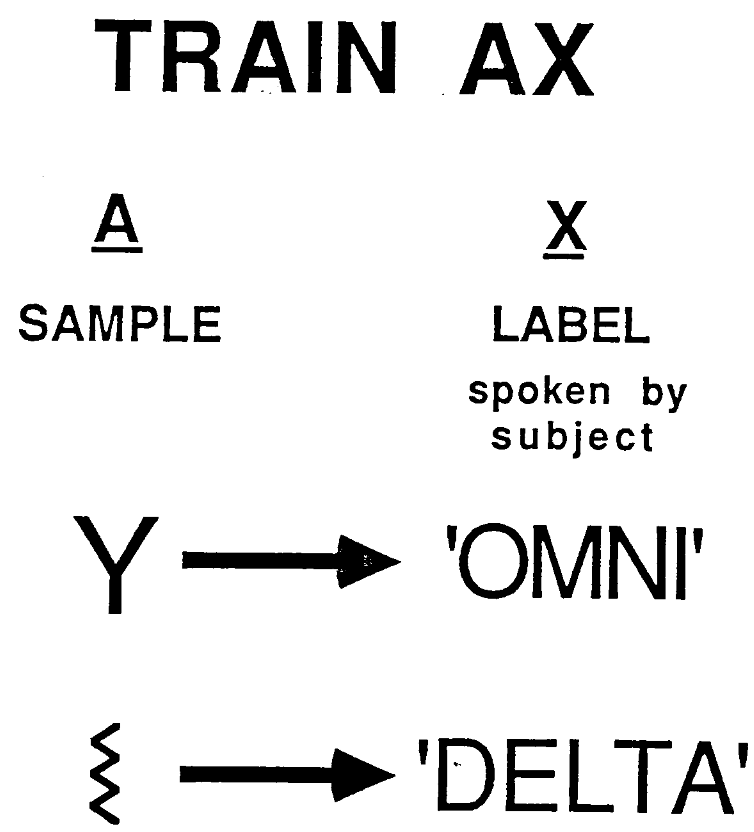
'In a while you will see something come on here' (experimenter points to the centre window). 'When it comes on I want you to tell me what it is.'

Then, when the first sample appeared, e.g. the Y-shape, the experimenter said:

'That is an Omni. You say it and then press it.'

If the subject hesitated, the latter instruction was repeated

Figure 4.10 Relations taught during A'X' training. Arrows point from sample shapes to corresponding labels spoken by the subjects.


TRAIN AX
A
SAMPLE
X
LABEL
spoken by
subject
Y
'OMNI'
'DELTA'

until the subject said the word.

The experimenter was not concerned if the subject's pronunciation of 'Omni' or 'Delta' was less than perfect. Any approximation to the desired sound was accepted, provided it was easy to distinguish and remained consistent across trials.

If the children had initial difficulty remembering the unfamiliar words they were asked 'Is it Omni or Delta?'

Occasionally, subjects were reminded to press the centre window after saying the word. As their labelling improved prompting was gradually reduced until it was no longer necessary. By the end of this phase, all subjects were able to label the Set-A shapes appropriately without any intervention from the experimenter.

(2) AB Training (resumed) All subjects were returned to the AB baseline for a maximum of two 48-trial sessions, to see if they would continue to label the Set-A shapes on matching-to-sample trials and, if so, whether this would help them acquire the AB relation. If the sample shapes were not labelled spontaneously, then prompting was given during additional sessions; the subject was asked 'Is it Omni or Delta?' when the sample appeared, and the comparisons were witheld until the subject said either word and then pressed the sample.

If the AB relation was learned then the child proceeded to stage 5 (see below). All other subjects proceeded to the next stage.

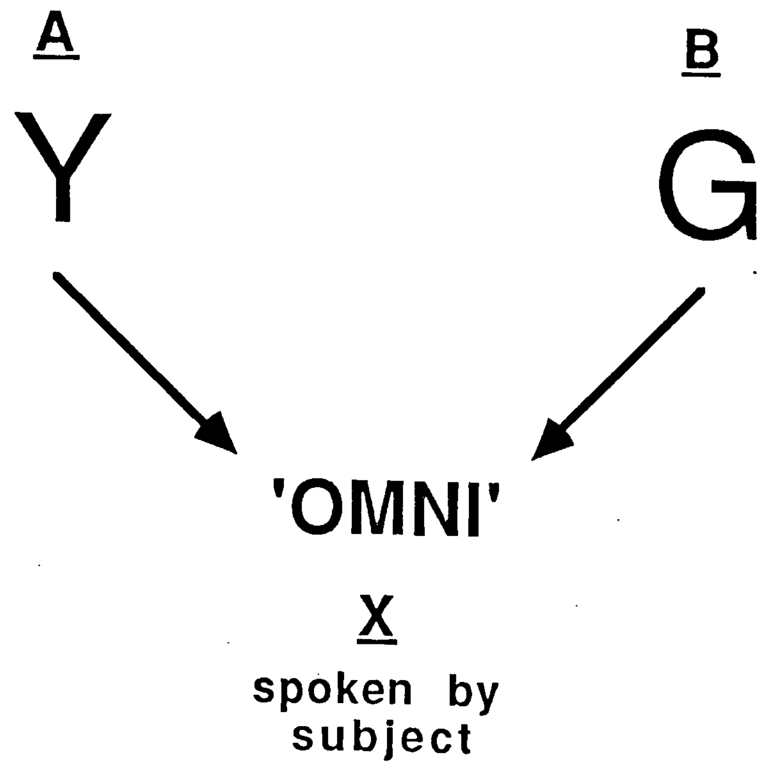
(3) Train B'X' The procedure here was identical to A'X' training (see above), except the green hue replaced the Y-shape and red replaced the zig-zag. So on each trial the green or red hue appeared in the centre window and the subjects were required to say 'Omni' in the presence of green and 'Delta' in the presence of red (see Figure 4.11).

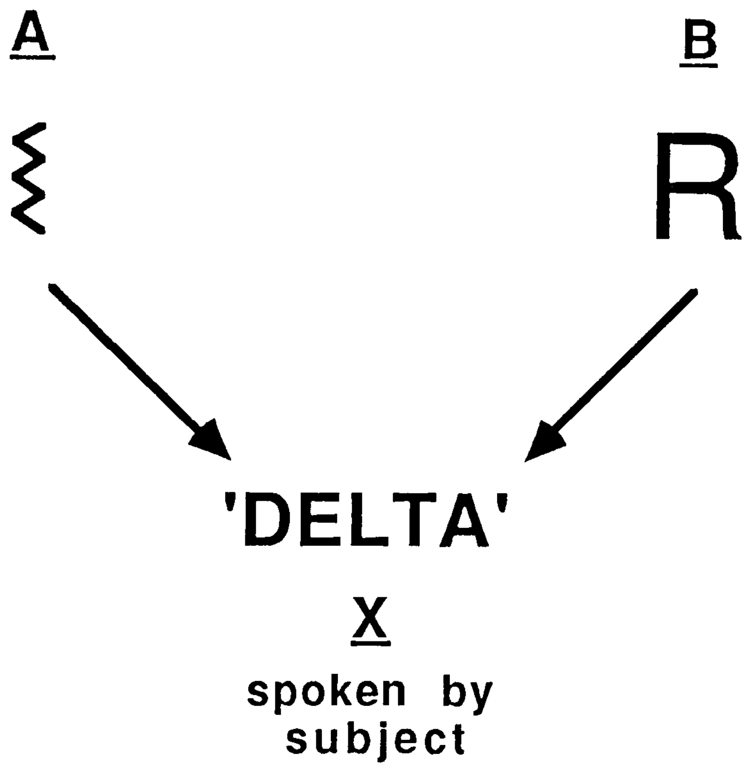
Figure 4.11 Relations taught during B'X' training. Arrows point from sample colours to corresponding labels spoken by the subjects.

TRAIN BX

B	X
SAMPLE	LABEL spoken by subject
G →	'OMNI'
R →	'DELTA'

Figure 4.12 Relations maintained in sessions comprising of combined A'X' and B'X' trials.


A
Y
B
G
'OMNI'
X
spoken by
subject


A
B
R
'DELTA'
X
spoken by
subject

(4) Combine A'X' and B'X' During this stage the Y-shape, zig-zag, green and red appeared equally often in the centre window and the subject was required to label the stimuli appropriately by saying 'Omni' or 'Delta' before pressing the window (see Figure 4.12). No prompting was given. If the subjects continued to label both Set-A shapes and Set-B colours correctly within the same session then they advanced to stage 5.

(5) Reduce reinforcement probability Reinforcement probability was reduced in accordance with the procedures outlined in the General Method section. Subjects were required to maintain criterion performance before progressing further.

Alternative Route

(1) 'X'A Training On 'X'A trials, subjects were required to select Set-A shapes conditional upon Set-X words spoken by the experimenter (see Figure 4.13). When the experimenter said 'Omni' reinforcers were contingent upon choosing the Y-shape, and when the experimenter said 'Delta' reinforcers were contingent upon choosing the zig-zag. Subjects were given the following instructions prior to their first session of 'X'A training:

'This time you will not see anything come on here' (experimenter points to centre window). 'Instead, I will tell you which one to choose. When you hear the beep, get ready and listen carefully. I will say a word. When I have said the word I want you to press this window (experimenter points to centre window) and then choose the right one.'

Figure 4.13 Relations taught during 'X'A training. Arrows point from sample words, dictated by the experimenter, to corresponding comparison shapes.

TRAIN XA

<u>X</u>	<u>A</u>
SAMPLE spoken by experimenter	CORRECT COMPARISON
'OMNI' →	Y
'DELTA' →	

Shortly after the beep indicated the start of the trial, the experimenter said either 'Omni' or 'Delta' according to a pre-determined random sequence recorded in advance on a printed sheet. Nothing happened if the subject pressed the centre window before the experimenter said the word. If the subject failed to respond after the word was dictated then the word was repeated. If a response was still not forthcoming then the experimenter repeated the word again and then told the subject to press the centre window. A single press to the blank sample produced the set-A comparisons. If the subject failed to choose a comparison the experimenter said, 'Now choose the right one.'

From time to time during the intertrial interval, the experimenter said:

'Remember to listen to what I say. I will tell you which one to choose.'

As the subjects became familiar with the procedure, prompting was reduced until it was no longer necessary. Eventually, all three subjects waited for the experimenter to say 'Omni' or 'Delta' before first pressing the blank sample and then choosing the comparison. Jessica quickly learned the 'X'A relation, but Linda and Sara R. did not. The latter pair were therefore transferred to the main route (see above) while Jessica continued to the next stage.

(2) AB Training (resumed) Jessica was returned to the AB baseline for two sessions, but her performance did not improve as a function of learning the 'X'A relation.

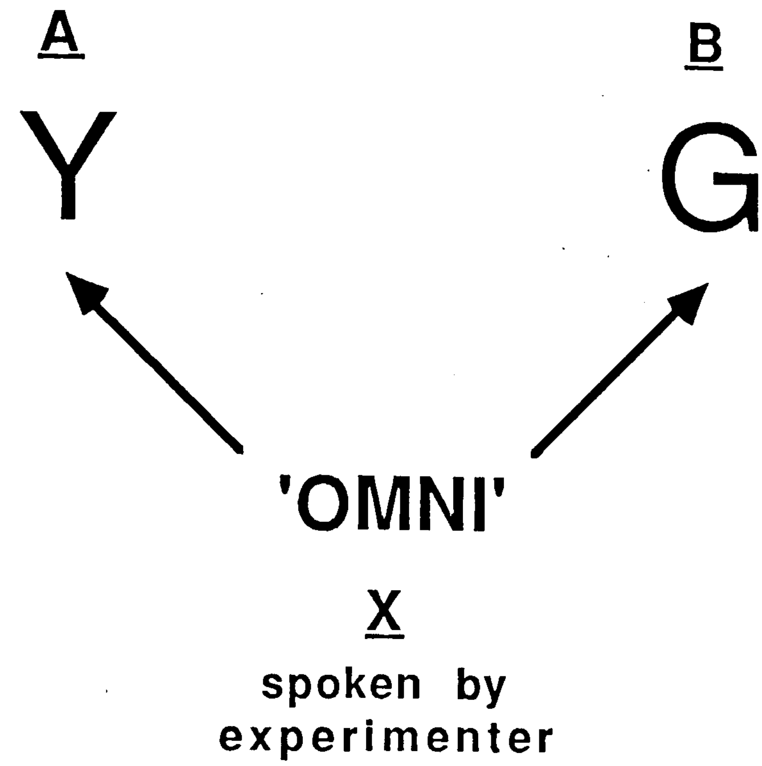
(3) Train 'X'B This was identical to 'X'A training except the Y-shape was replaced with a green comparison and the zig-zag was replaced with red. So, the subject was required to select green when the

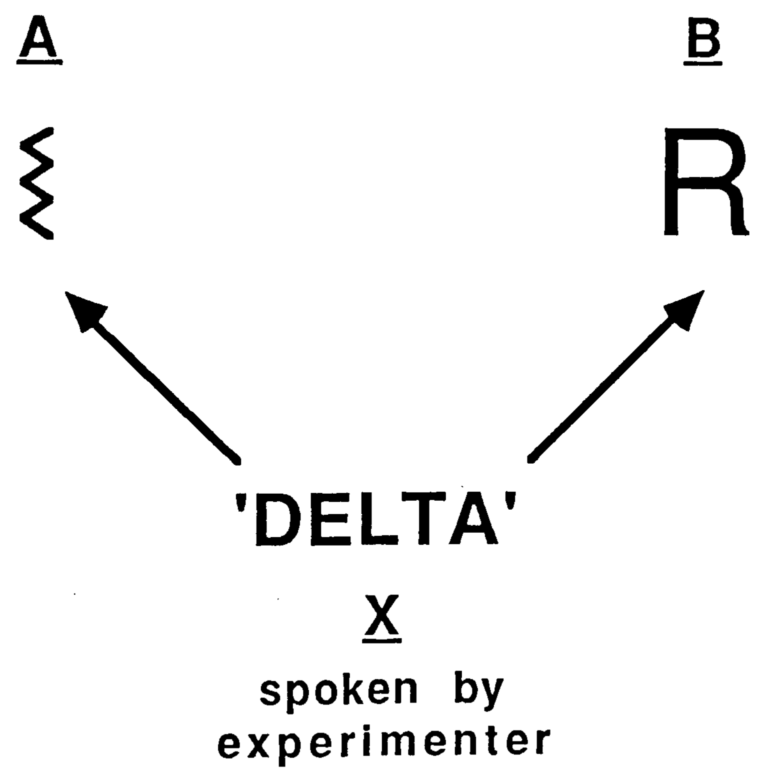
Figure 4.14 Relations taught during 'X'B training. Arrows point from sample words, dictated by the experimenter, to corresponding comparison colours.

TRAIN XB

<u>X</u>	<u>B</u>
SAMPLE spoken by experimenter	CORRECT COMPARISON
'OMNI' →	G
'DELTA' →	R

Figure 4.15 Relations maintained in sessions comprising of combined 'X'A and 'X'B trials.


A
B
Y
G
'OMNI'
X
spoken by
experimenter


A
B
R
'DELTA'
X
spoken by
experimenter

experimenter said 'Omni' and red when he said 'Delta' (see Figure 4.14).

(4) Combine 'X'A and 'X'B Either Set-A shapes or Set-B colours appeared equally often as comparisons, and the subject was required to select the comparison which corresponded to the word 'Omni' or 'Delta' dictated by the experimenter (see Figure 4.15). If criterion performance was maintained then the subject advanced to the next stage.

(5) Reduce reinforcement probability (see above).

Test Procedure

AB and BA Tests After completing training, all subjects received both AB and BA test trials. Testing was conducted in extinction; correct and incorrect responses were followed by the intertrial interval only, and neither tokens nor tones were delivered (see General Method section).

The BA trials, which had never been presented prior to testing, were the symmetrical counterparts of the AB trials. On BA trials, red or green appeared as samples and the Y-shape and zig-zag were comparisons.

In general, the subjects were given equal numbers of AB and BA trials. However, in some cases the subject's performance warranted more BA than AB trials (see Results).

In addition, subjects were prompted to label the sample if they did not do so spontaneously. The initial prompts were general; when the sample appeared the experimenter asked the subject to 'Say what it is'. If this failed to induce common labelling of the samples then a more specific prompt was given; in further test sessions the subject was asked, 'Is it Omni or Delta?' when the sample appeared.

Testing was continued until the subject was correct on at least six consecutive trials per trial-type (sample - comparison combination).

Naming Tests Each subject received at least one naming test during the course of the experiment. Set-A shapes and Set-B colours were presented one at a time in the centre window and verbal responses were elicited from the child in accordance with the procedures outlined in the General Method section.

RESULTS

TRAINING RESULTS

AB Training All subjects failed to learn the AB relation. Table 4.19 lists the number of trials each subject received for each intervention during AB training, and, where necessary, further detail appears below.

(a) Fading

The fading programme was administered to all subjects but with no effect. Five of the subjects (Francis, Linda, Michael P., Nicola and Stephen) never even learned the presumably simple skill of colourmatching, upon which the success of the fading programme depended. In contrast, both Jessica and Sara R. matched the colours from the start of their first fading session, and Alex began to do so in his third session. All three children continued to match correctly up to and including the penultimate fading step, on which the sample

Table 4.19 Number of trials per AB intervention for each subject in Experiment 4(a).

	AB INTERVENTION								
	I	II	III	IV	V	VI	VII	TOTAL	CP
ALEX	245	33	27	48	-	-	-	353	133
FRANCIS	96	-	-	-	-	-	-	96	0
JESSICA	421	96	-	-	-	101	120	738	260
LINDA	254	-	-	240	-	96	96	686	151
MICHAEL P.	252	-	48	48	48	-	-	396	75
NICOLA	207	-	-	-	-	-	-	207	159
SARA R.	258	48	-	-	-	99	168	573	187
STEPHEN	283	-	-	93	144	-	134	654	303

KEY
I = Fading
II = Increased intertrial interval
III = Sample schedules
IV = Delayed cueing
V = Reward reduction
VI = Instructions
VII = Standard AB trials
CP = Number of trials from total which were subject to correction procedures.

colour cue was at its smallest. However, on the final step of fading no sample colour cue appeared (i.e. the trial was a normal AB matching trial), and the subjects' responding fell to chance level. The breakdown of performance at the final step of a fading program is a ubiquitous problem which has been noted elsewhere (see Tennant, Cullen and Hattersley (1981) for a review).

(b) Delayed cueing

The delayed cueing trials were thwarted by three of the four subjects who received them. On the majority of trials Alex, Michael P. and Stephen managed to secure reinforcers simply by waiting for the incorrect comparison to disappear before selecting the one remaining (correct) comparison. Michael P. did this on all 48 trials in which the cueing interval (i.e. the time between comparison presentation and removal of the incorrect choice) was 1 second. Over the course of 48 trials with cueing intervals of 1 and 2 seconds, Alex had only three trials on which he responded prior to the removal of the incorrect comparison, and he got all three wrong! Stephen had 93 delayed cueing trials with intervals ranging from 0 to 10 seconds. He waited for the incorrect comparison to disappear on all but three of the trials, and got two of these wrong.

In contrast to the others, Linda received 240 delayed cueing trials with cueing intervals from 0 to 1.2 seconds, and on 95 of these she responded before the incorrect comparison was scheduled to disappear. Unfortunately, her performance on these normal AB matching trials remained at or around chance level.

(c) Reward reduction

Removing tokens for incorrect responses did not affect either Michael's or Stephen's matching performance, although they did become generally reluctant to participate further. This problem disappeared when reward reduction was subsequently dropped.

(d) Sample schedules

Both Alex and Michael were required to press the sample five times to produce the comparisons, but this tended to decrease rather than increase the degree to which they looked at the sample. Both subjects tended to start pressing and then continued to press while looking at other things in the room. As a consequence, they often pressed the sample longer than was necessary before realising that the comparisons had appeared.

(e) Correction

All subjects except Francis received correction at some stage of AB training. During correction, incorrect responses produced repeat trials until the subject was correct. Correction was introduced in an attempt to disrupt strong preferences for one or other comparison. For example, Linda began her first AB matching session by choosing the red comparison 16 times in the first 18 trials. Correction was then introduced, which forced her to eventually select the green comparison in order to secure a reinforcer. Although effective in disrupting comparison preferences, correction did not promote acquisition of the AB relation. The problem was that whilst weakening straightforward comparison preferences, correction simultaneously reinforced other equally undesirable comparison-based response tendencies. For example, Linda shifted from exclusively choosing the red comparison to adopting a win-stay/lose-shift strategy i.e. she selected one comparison until incorrect and then shifted to the other comparison. Linda also spent distinct periods choosing red on one trial and then green on the next, alternating thereafter. Both win-stay/lose-shift and alternation are actually encouraged by the reinforcement contingencies imposed during correction.

To recap, all subjects failed to learn AB matching within the number of trials allotted (Francis received 96 trials, Nicola 207, Alex 353, Michael P. 396, Sara R. 573, Stephen 654, Linda 686, and Jessica 738). Subsequently, all subjects except Jessica followed the main route to testing (see Procedure). Linda and Sara also began the alternative route along with Jessica, but were later transferred to the main route. For ease of exposition, the results from each route are presented separately.

Main Training Route

(1) Train A'X' In general, the A'X' relation was learned extremely quickly. Four subjects needed only 48 trials or less to learn to label the shapes to criterion (Sara required 27 trials, Stephen 36, Linda 44, and Alex 48). Two others needed just over one session to learn (Nicola needed 56 trials and Michael 76). In contrast, Francis required 210 trials to learn this relation. He appeared to have particular difficulty remembering what words to say, and he therefore required a great deal of prompting from the experimenter. When Francis was not forthcoming with a label for the shape, the experimenter assisted by asking 'Is it Omni or Delta?' By the final session of A'X' training Francis no longer needed prompting.

Despite Francis's difficulty, it seems clear that these children find some conditional relations harder to learn than others. None of the children learned the AB relation involving purely visual stimuli, but all of them learned A'X', which related visual stimuli to verbal responses.

(2) AB Training (resumed) Having learned to label the Set-A shapes,

the subjects were returned to the AB matching task. This raised two questions. Firstly, would they continue to label the Set-A shapes when presented as samples on matching trials? Secondly, if so, would this help them to learn the AB relation? Several experiments have shown that responding differentially to the samples can facilitate learning of matching tasks (e.g. Urcuioli, 1985).

All of the subjects except Linda and Sara spontaneously labelled the sample shapes with the Set-X words 'Omni' and 'Delta'. However, only Nicola apparently benefited from this differential labelling; she reached criterion after only 48 AB trials with sample labelling. The four others were each given 96 AB trials, but their scores did not improve as a function of spontaneously labelling the samples.

Only Sara and Linda failed to label the samples upon return to the AB matching trials. This absence of labelling was not caused by a failure to remember the words, because both subjects labelled the shapes correctly when their A'X' performance was subsequently reviewed (Sara's A'X' review came immediately after the first of two AB sessions, whereas Linda's was given after the second AB session). Since both were capable of labelling the samples, but were not doing so spontaneously, they were given further AB trials during which they were prompted to label the samples (see Procedure). The prompting proved successful; both children correctly labelled the samples from this point on. The labelling produced no change in Linda's AB performance, but Sara's behaviour changed dramatically. Her AB score rose to 77% after the first short 26 trial session with prompting, and to above 80% on the second session with prompting removed. Sara's verbal behaviour during the prompted session was revealing, and the session itself illustrated some peculiar difficulties with respect to the reinforcement contingencies. These points are best seen in relation to the transcript of the prompted session, which is presented in Table 4.20. Sara began the session by making no errors on the first 8 trials, and she also labelled the sample correctly before choosing the correct comparison. Furthermore, on trial 7 she expressed a rule linking colour to shape by saying, 'The Red's the Delta, isn't it?'

However, one trial later, on trial 9, Sara showed signs of becoming confused. The Y-shape appeared and she incorrectly called it 'Delta'; she then proceeded to choose the red comparison in accordance with the rule she had expressed on the preceding trial. Her choice was incorrect, and at the end of the trial she said she was getting confused. In retrospect, her confusion should have come as no surprise because on incorrect trials there were two ways in which she could interpret the outcome. Firstly, she could think she was wrong because she had perhaps labelled the sample incorrectly. Alternatively, she could attribute her error not to an incorrect sample label but to an incorrect choice of comparison. Sara apparently believed her error on trial 9 was caused by an incorrect choice of comparison. On trial 9 she adopted her rule - the Red is the Delta - and got the trial wrong, so on the next two trials she chose green after calling the sample 'Delta' and also said, 'The Green is the Delta' and 'The Red is the Omni'. Failure on trial 9 had apparently led Sara to modify her verbal rule.

Since Sara's confusion appeared to stem from not knowing the source of her error on any given trial, then it seemed likely that one could resolve her confusion by indicating the source. So, from trial 12 onwards, the procedure was deliberately modified so that if the sample was labelled incorrectly then sample presses had no effect. Sara had to label the sample correctly in order to produce the comparisons. This change apparently helped, because for the next seven trials, and despite several labelling errors, Sara stuck to the correct

Table 4.20 Sara R's responses from the AB training session in which she was prompted to label the sample shapes with the words 'Omni' or 'Delta'. Trials are listed consecutively, and for each trial the sample is listed along with the verbalisations of the subject (S) and the experimenter (E). In addition, sample and comparison presses are indicated in relation to the subject's verbalisations, along with the outcome of each trial (correct or incorrect comparison choice).

Table 4.20

Trial	Sample	Responses	Outcome
1	Y	E: 'Is it Omni or Delta?' S: 'That's the Omni, isn't it? (presses sample and green comparison)	Corr
2	Z	E: 'Is it Omni or Delta?' S: 'Delta' (presses sample and red comparison)	Corr
3	Z	S: 'This is the Delta, isn't it?' (presses sample and red comparison)	Corr
4	Y	E: 'Is it Omni or Delta?' S: 'Omni' (presses sample and green comparison) S: 'I nearly said Delta!'	Corr
5	Y	S: 'Omni' (presses sample) S: 'Which is the Omni?' (presses green comparison)	Corr
6	Z	S: 'Delta' (presses sample and red comparison)	Corr
7	Z	S: 'Omni isn't it? - No it's not! Delta' (presses sample and red comparison S: 'The red's the Delta, isn't it?'	Corr
8	Y	S: 'Omni' (presses sample and green comparison)	Corr
9	Y	S: 'Delta' (presses sample and red comparison) S: 'I keep getting confused'	Inc
10	Z	S: 'Delta' (presses sample and green comparison)	Inc

Table 4.20 (cont'd)

Trial	Sample	Response	Outcome
11	Z	S: 'This one's the Delta. The other one's the Omni. (presses sample) S: 'The green is the Delta, isn't it?' (presses green comparison) S: 'The red is the Omni, isn't it?'	Inc
		NB: Procedure modified so that sample press produces the comparisons only if sample is labelled correctly.	
12	Y	S: 'Delta, this is, isn't it?' (presses sample, but no effect) S: 'Oh, it's the Omni' (presses sample, comparisons appear) S: 'And that's red, isn't it?' (presses red comparison) S: 'It's the green, isn't it?'	Inc
13	Y	S: 'Omni' (presses sample, comparisons appear) S: 'and the Omni is the Green, isn't it?' (presses green comparison)	Corr
14	Z	S: 'Delta, this is. Delta is the Red, isn't it?' (presses sample, comparisons appear) S: 'Delta' (presses red comparison)	Corr
15	Z	S: 'Omni this is, isn't it, because the Green is the Omni, isn't it?' (presses sample, no effect) S: 'Delta' (presses sample, comparisons appear, presses red comparison)	Corr
16	Y	S: 'This is the Green isn't it? It's the Omni' (presses sample, comparisons appear, presses green comparison)	Corr

Table 4.20 (cont'd)

Trial	Sample	Response	Outcome
17	Y	S: 'This is the same word, isn't it? Omni' (presses sample, comparisons appear, presses green comparison)	Corr
18	Z	S: 'This is the Red, isn't it? Delta.' (presses sample, comparisons appear, presses red comparison)	Corr
19	Y	S: 'The other one is the Delta isn't it. That one's the Green isn't it. What name shall I give to this? .....Omni' (presses sample, comparisons appear, presses green comparison)	Corr
		NB: Revert to original contingency. All sample presses produce comparisons.	
20	Z	S: 'This is the Omni one' (presses sample) S: 'Green' (presses green comparison)	Inc
		S: 'The other one's the Green, isn't it?'	
21	Z	S: 'Omni, this is isn't it?' (presses sample) S: 'Red, isn't it?' (presses red comparison)	Corr
22	Y	S: 'Omni' (presses sample) S: 'You press Red for that, don't you?' (presses red comparison)	Inc
23	Y	S: 'You press the Green for this, don't you? Omni, isn't it?' (presses sample and green comparison)	Corr

Table 4.20 (cont'd)

Trial	Sample	Response	Outcome
24	Z	S: 'Delta' (presses sample) S: 'You press the Red this time, don't you?' (presses red comparison)	Corr
25	Y	S: 'This is the Omni' (presses sample) S: 'and it's the green isn't it?' (presses green comparison)	Corr
26	Z	S: 'Delta' (presses sample) S: 'The Delta's the Red' (presses red comparison)	Corr

form of the verbal rule, which she expressed on several occasions (i.e. she said 'Omni is Green' and 'Green is the Omni' on trials 13 and 15 respectively, and on trial 14 she said 'Delta is the Red'). The effect was finally confirmed by reverting to the original contingency on trial 20, when once again the comparisons were produced even when Sara labelled the sample incorrectly. On trial 20, Sara incorrectly said 'Omni' to the zig-zag sample and then proceeded to choose the green comparison in accordance with her verbal rule. Now Sara probably assumed that she had labelled the sample correctly, rather than incorrectly, because her subsequent sample press produced the comparisons, and incorrect labels had prevented the sample from appearing on previous trials. She therefore had good reason to assume that she was wrong on trial 20 because of an incorrect choice of comparison. If her incorrect comparison choice was determined by the verbal rule 'Omni is Green' then one would predict she would change the rule yet again to 'Omni is Red', but continue to label the samples incorrectly. This is exactly what happened on the next trial (trial 21). As predicted, Sara again labelled the zig-zag sample incorrectly with the word 'Omni' but this time chose the red comparison, and so, by default, she got the trial right! On the next trial, after saying 'Omni' in the presence of the sample, and <u>before</u> selecting a comparison she said, 'You press Red for that don't you?', and then chose red. Once again she was incorrect, and once again this prompted her to change the rule. On trial 23 a Y sample was presented and <u>before pressing the sample</u> she said:

'You press the Green for this, don't you? Omni, isn't it?'

and then, after correctly choosing green she said:

'The other one is Red, so the other one is Delta.'

On the last three trials, Sara labelled the sample correctly and chose the correct comparison, and on the first two of these she labelled the correct comparison prior to selecting it. She ended the session on trial 26 by announcing the correct verbal rule 'Delta is Red' prior to choosing the red comparison.

Reading the transcript leaves one in little doubt that Sara's verbal behaviour, in the form of a rule linking sample to corresponding comparison, was guiding her choice of comparison stimulus and that, in addition, the rule was extremely sensitive to both the changes in, and the effects of, the reinforcement contingencies.

In the next session, Sara's score on AB matching trials rose to criterion levels, but she did not overtly label the samples or verbalise any rules. No further prompting was given, because she was responding correctly anyway, and further prompting at that stage may have constrained future experimental manipulations (e.g. if Sara failed her subsequent symmetry test and also failed to label the samples, one could then prompt sample labelling to see what effect, if any, this might have on her test performance).

To recap, both Nicola and Sara learned the AB relation at this stage, and so proceeded to stage 5 to prepare for testing. The five other children failed to learn AB and therefore went on to the next stage.

(3) Train B'X' Linda required 96 trials (two sessions) to learn to label green with the word 'Omni' and red with the word 'Delta'. All the other children learned the B'X' relation within one 48 trial session.

(4) Combine A'X' and B'X' Combining the A'X' and B'X' trials

randomly within the same session did not affect the subjects' labelling performances. By the end of this period, all five remaining subjects were labelling the Y-shape and green stimuli with the common label 'Omni', and zig-zag and red with the common label 'Delta'.

(5) Reduce reinforcement probability Only one subject, Sara, was affected by the removal of reinforcement from training trials. Her AB score dropped slightly below criterion but recovered one session later to 100% correct.

Summary Seven subjects followed the main training route to testing. Five of these (Alex, Francis, Linda, Michael P. and Stephen) had failed to learn the AB relation, but had succeeded in learning to produce common labels for corresponding Set-A and Set-B stimuli. The other two (Nicola and Sara) learned AB after learning to label the Set-A samples.

Alternative Training Route

(1) 'X'A Training Both Sara and Linda failed to learn the 'X'A relation despite receiving 178 and 144 trials respectively, and were subsequently transferred to the main training route (see above). Jessica, however, needed only 80 trials to learn to select Set-A shapes conditional upon the words 'Omni' and 'Delta' dictated by the experimenter. The rapid acquisition of this auditory-visual relation contrasts sharply with her failure to learn the AB visual-visual relation even after more than 700 trials.

(2) AB Training (resumed) After learning 'X'A, Jessica was given a

further 74 AB trials, but her AB score remained at or around chance level.

(3) Train 'X'B   Jessica required only 15 trials to learn to select the green comparison when the experimenter said 'Omni', and to choose the red comparison when the experimenter said 'Delta'. She made no errors.

(4) Combine 'X'A and 'X'B   Jessica continued to respond appropriately when the 'X'A and 'X'B trials were combined within the same session. After 25 trials she was ready for testing, having made no errors in the last 12 unreinforced trials.

Overall Summary of Training

Table 4.21 shows that each subject, prior to testing, had received a different number of AB training trials and a different number of training trials in total. Trials were deliberately staggered in order to produce a multiple baseline design across subjects. At the end of training, two of the eight subjects, Sara and Nicola, had learned the AB relation after learning to label the Set-A samples. For these two subjects the AB trials formed a baseline which could be assessed for symmetry in subsequent test sessions. The other six subjects had failed to learn the AB relation during training. Of these six, five had learned to produce common Set-X labels to the Set-A and Set-B stimuli, as depicted in Figure 4.12. The remaining subject, Jessica, had not learned to produce common labels but rather to select Set-A and Set-B comparisons conditional upon common labels dictated by the experimenter, as depicted in Figure 4.15.

Table 4.21 Number of trials per training task for each subject in Experiment 4(a).

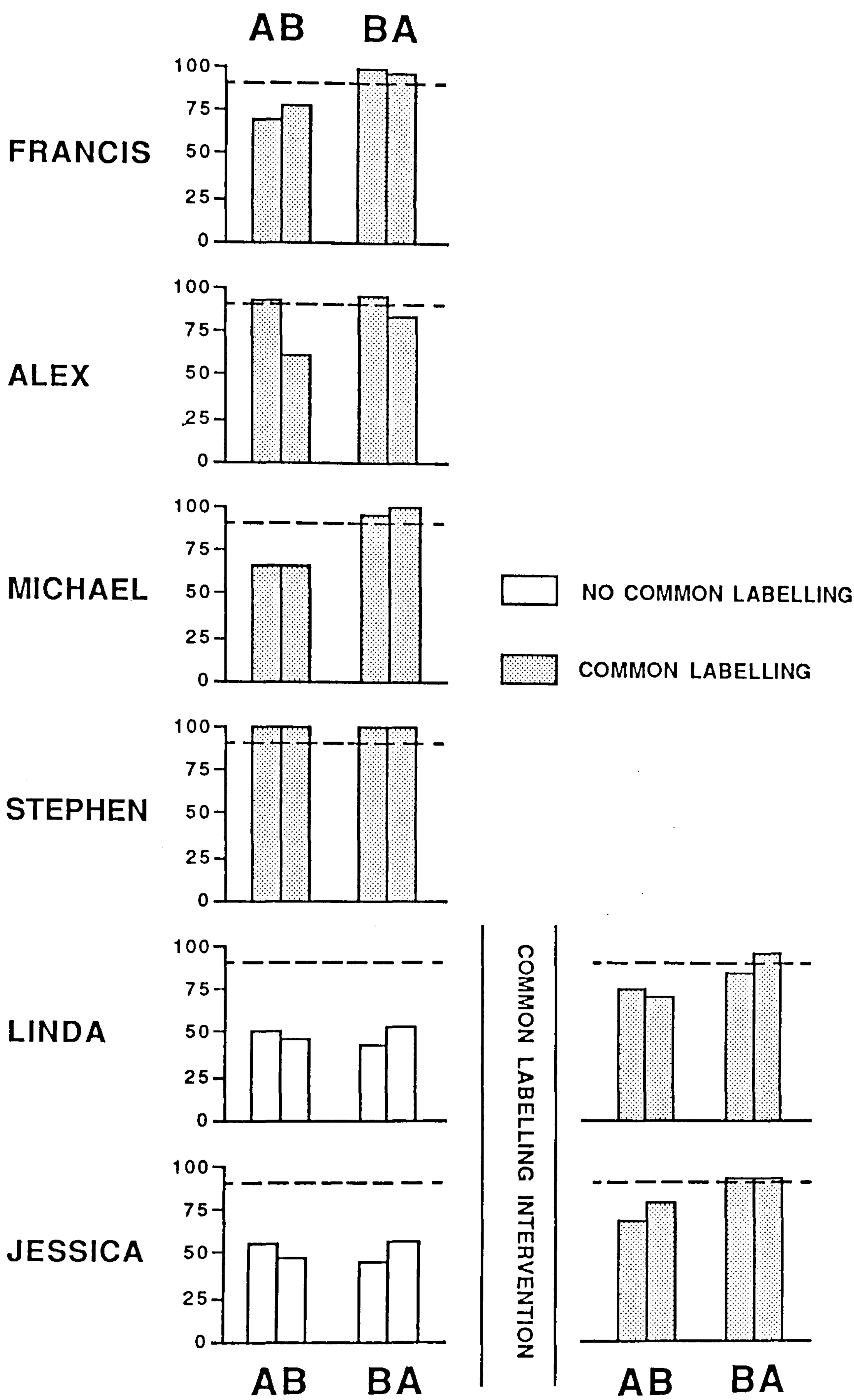
Subject	Training Task					
	AB	A'X'	B'X'	'X'A	'X'B	TOTAL
ALEX	449	76	52	-	-	577
FRANCIS	192	220	34	-	-	446
JESSICA	812	-	-	92	28	932
LINDA	782	92	120	144	-	1138
MICHAEL P.	492	124	96	-	-	712
NICOLA	255	56	-	-	-	311
SARA R.	781	72	-	178	-	1031
STEPHEN	750	60	60	-	-	870

TEST RESULTS

The test results for the six subjects who had failed to learn the AB relation are examined first, and are depicted in Figure 4.16. Each bar represents the overall test score for a particular trial-type. AB trial-types appear to the left of BA. The dotted line represents an overall criterion level of 90% correct responses per trial-type. The shading depicts those matching-to-sample trials on which common Set-X labels were correctly applied to the Set-A and Set-B samples. So the four subjects at the top of Figure 4.16 gave common labels to the samples right from the start of testing. On AB trials they labelled the Y - shape with the word 'Omni' and the zig-zag with the word 'Delta', and on BA trials they labelled the green sample with 'Omni' and the red sample with 'Delta'. However, Linda and Jessica, the two subjects at the bottom of Figure 4.16 did not begin by applying common labels to the samples, as depicted by the absence of shading. Initially, common labelling was absent and their performance on both AB and BA was at or around chance level or 50% correct. So, both Linda and Jessica, in the absence of common labelling, had failed the tests. Jessica had failed even though at this stage she could select Set-A shapes and Set-B colours conditional upon common labels spoken by the experimenter. For Jessica then, the mere experience of hearing common labels, distinct from producing them herself, was apparently insufficient for mediating the emergence of the AB and BA relations.

Both Linda and Jessica did eventually join the others in producing common labels, when in subsequent tests the two girls were prompted with the question, 'Is it Omni or Delta?' when the samples appeared. This prompting is represented in Figure 4.16 by the vertical column

Figure 4.16 Overall test scores (percentage of correct responses) on AB and BA trial-types for the six subjects in Experiment 4(a) who failed to learn AB matching prior to testing.

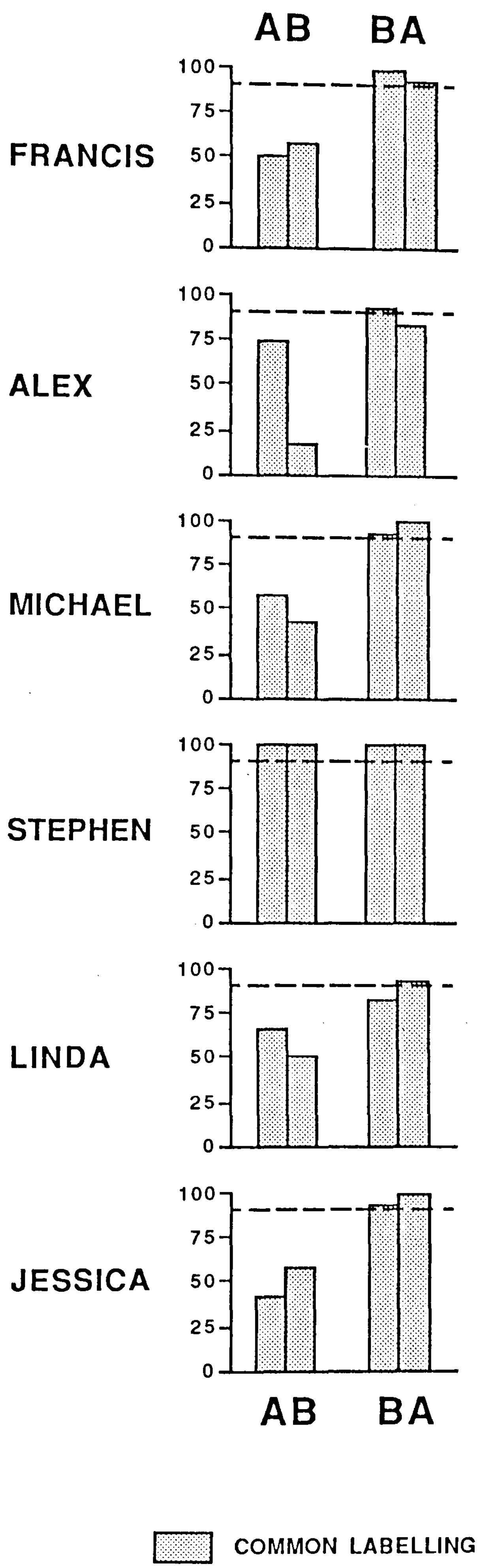

AB
BA
FRANCIS
ALEX
MICHAEL
STEPHEN
LINDA
JESSICA
100
75
50
25
0
NO COMMON LABELLING
COMMON LABELLING
COMMON LABELLING INTERVENTION

enclosing the text 'common labelling intervention'. Now all six subjects were applying common labels to the samples. The critical question is what happened to the test performance when the subjects applied common labels to the samples? First of all, and what one will not notice from Figure 4.16 is that both the AB and BA relations eventually emerged when all six subjects applied common labels. By the end of this phase of testing, subjects were making no errors on any of the test trial-types; in fact testing under common labelling was continued until each subject was correct on six consecutive trials per trial-type. So, each shape and its corresponding colour had become equivalent through the subjects' common Set-X labels. Common labelling of the samples by the subjects was sufficient for mediating the emergence of the AB and BA relations.

Although both the AB and BA relations had completely emerged by the end of this test phase, Figure 4.16 actually represents each subject's performance throughout and not just at the end of each test phase. The figure shows that, taken overall, the BA scores were better than the AB scores for five of the six subjects. The subjects overall scores on AB trial-types tended to fall short of the 90% criterion line, whereas their overall scores on BA trial-types tended to be above criterion. Figure 4.17 shows each subject's performance over the first 48 trials of testing (each bar represents the first 12 trials per trial-type). Inspection of Figure 4.17 confirms that the difference between AB and BA test performance was most pronounced particularly in the initial stages of the test phase.

So there were two main findings. Firstly, common labels produced by the subjects resulted in the emergence of both the AB and BA relation. Secondly, although both relations emerged as a function of common labelling, BA emerged prior to AB for all subjects except Stephen, who scored 100% correct on both.

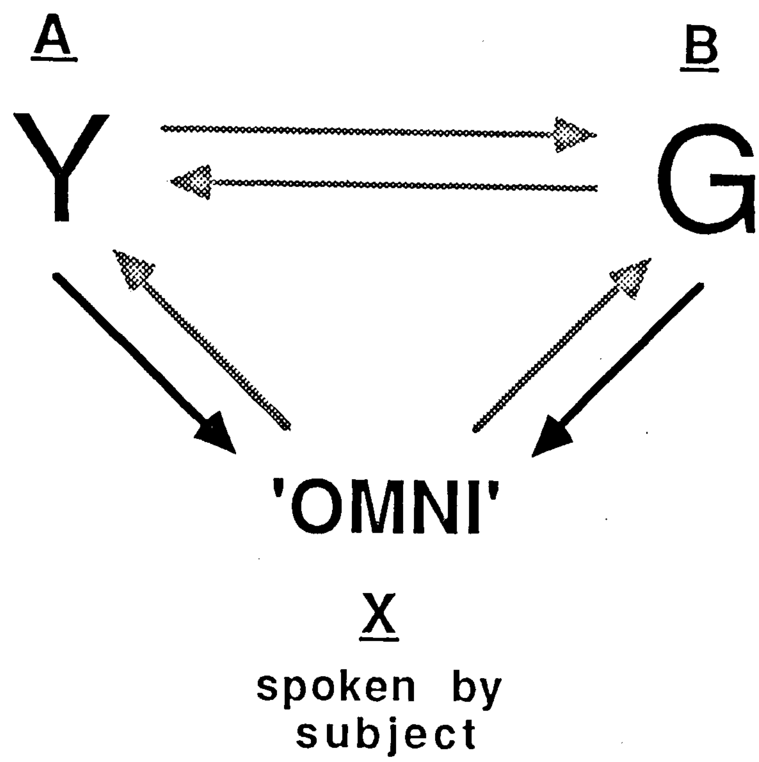
Figure 4.17 Initial AB and BA scores from the first 48 trials of the common labelling phase of testing, for each of the six subjects in Experiment 4(a) who failed to learn AB prior to testing.

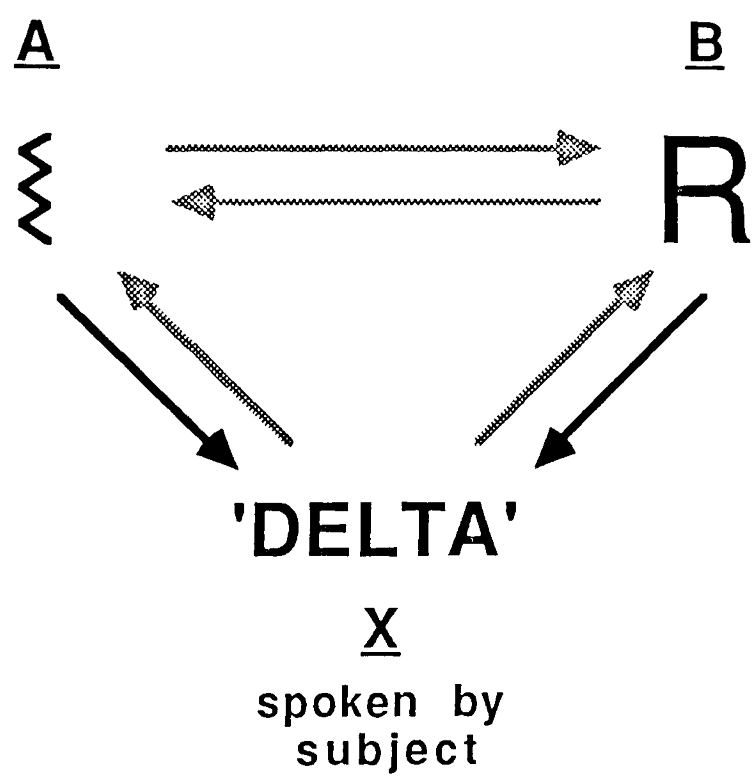

AB BA
FRANCIS
ALEX
MICHAEL
STEPHEN
LINDA
JESSICA
100
75
50
25
0
AB BA
COMMON LABELLING

Why did BA emerge prior to AB? To find out one must examine how the subjects' labels helped them to match the stimuli. Figure 4.18 depicts the stimuli in question, and the relations between them. To simplify matters a little it might be best to concentrate on one pair of stimuli, for example, the Y-shape and green stimuli depicted at the top of Figure 4.18 (of course, all of the following explanation also applies to the other stimulus pair).

Let us consider the BA relation first since this emerged virtually straight away when common labelling was introduced. On BA trials, subjects matched the green sample to the Y-shape comparison. How did their labels help them to do this? Well, two things must happen. First, when the green sample appears the subject must say 'Omni' (i.e. B'X'). Secondly, when the subject has said 'Omni' he must choose the Y comparison (i.e. 'X'A). Matching green to Y then becomes a two-stage process; B'X' and 'X'A, therefore BA. During testing, all the subjects proved capable of B'X'; they all, for example, overtly labelled green with the word 'Omni' - their earlier training had established this skill. But none of them had been taught the second stage, 'X'A; none of them (not even Jessica) had been taught to say 'Omni' and then choose the Y-shape. So, from where did this critical second component emerge? The answer is from the A'X' relation, which had been established by the subjects' earlier training. Somehow, when the skill of labelling the Y-shape with the word 'Omni' was established, so too was the potential for its symmetrical counterpart, saying 'Omni' and then choosing the Y-shape. The subjects' earlier training had resulted in the formation of symmetry between the Set-A shapes and the subjects' spoken Set-X words. Had it not done, then 'X'A would have been absent during testing and the BA relation could not have emerged as a function of the subjects' Set-X labels.

Figure 4.18 Relations between Set-A shapes, Set-B colours and Set-X words (spoken by the subjects). Black arrows represent relations established prior to testing. Shaded arrows represent relations assessed during tests (see text).


A
B
Y
G
'OMNI'
X
spoken by subject


A
B
R
'DELTA'
X
spoken by subject

The same analysis may be applied to the AB relation. Although the AB relation eventually emerged, it did not emerge straight away when the subjects produced common labels. The AB relation was at first not fully present. On AB trials, when the Y-shape appeared the subjects overtly said 'Omni' but then having said 'Omni' they did not consistently choose green. In other words the 'X'B relation was not fully present at first. Although the subjects had learned, from their earlier training, to label green by overtly saying 'Omni' (B'X'), they were not initially capable of doing the reverse - saying 'Omni' and then choosing Green ('X'B). For some reason, the earlier establishment of B'X' had not resulted in the immediate formation of symmetry between the Set-B colours and the subjects' spoken Set-X words.

So, the AB and BA matching trials therefore provided a convenient framework within which one could determine whether the A'X' and B'X' relations were themselves symmetrical. The test results confirmed that for five of the subjects, and for the most part of testing, the A'X' relation was symmetrical but the B'X' relation was <u>not</u>. And there are good reasons for this disparity. But in order to make sense of the data, one needs to examine the subjects' labelling skills <u>prior</u> to establishing the A'X' and B'X' relations. There were two sources of information; firstly, the subjects' spontaneous naming monitored throughout the experiment and secondly, their responses during naming tests.

All five subjects who acquired BA prior to AB also spontaneously labelled the colours with their conventional names, 'Green' and 'Red', prior to the establishment of A'X' and B'X'. None of the subjects, however, consistently labelled the shapes with any other than the Set-X words, either spontaneously or during naming tests. During naming tests, Jessica was the only one of these children to label the shapes at all, but even then she labelled them inconsistently. She responded to the Y-shape by first saying 'I don't know', and then on subsequent trials she called it 'a stick with spikes' and 'a palm tree'. On successive trials she called the zig-zag a 'worm', a 'snake' and a 'caterpillar'. Jessica never gave a consistent name to the shapes. In fact one could argue that she was not naming the shapes at all; rather she appeared to be saying <u>what they looked like</u> (and it was interesting to note that this descriptive tendency also extended to the colours which she called 'blocks', despite spontaneously naming them 'Green' and 'Red' during AB matching trials).

A knowledge of the subjects' pre-existing naming skills may be crucial for understanding why labelling of the Set-A shapes was governed by a symmetrical relation while at the same time labelling of the Set-B colours was not. All five subjects already had names for the Set-B colours; they used the conventional labels, 'Green' and 'Red'. It therefore seemed likely that these conventional labels were somehow interfering with the formation of symmetry between the colours and the Set-X labels 'Omni' and 'Delta', and that this in turn interfered with the emergence of AB during testing. In contrast, BA emerged straight away, perhaps because the subjects had no other names for the Set-A shapes prior to labelling them with the Set-X words. The A'X' training, free of any interference from pre-existing names, could then result in the formation of symmetry between each shape and its corresponding Set-X word.

Further evidence of the interfering effect of the conventional colour names comes from a more detailed examination of each subject's behaviour during testing. But before presenting this data it might be worthwile reflecting upon the results presented so far.

PRELIMINARY DISCUSSION

Although both the AB and BA relations emerged after all six subjects applied common labels, BA emerged prior to AB in all but one case. This in turn showed that the Set-X words 'Omni' and 'Delta', spoken by the subjects, were symmetrically related to the Set-A shapes but not to the Set-B colours (at least not until the last few trials of testing). A symmetrical relation is bi-directional, and bi-directionality, as was outlined in the introductory chapters, has often been proposed as a defining property of symbolic behaviour (see Bates, 1976; Devany, Hayes and Nelson, 1986). This notion may be applied to the results of the present experiment. Given that the relation between the Set-X words and Set-A shapes was bi-directional, one has grounds for claiming that the subjects were behaving symbolically when they labelled the shapes with the words 'Omni' and 'Delta'. However, although the subjects applied the same words to the Set-B colours it seemed that, in so doing, they were not behaving symbolically, because the B'X' relation was not initially bi-directional.

The argument may be taken one step further by proposing that naming is itself a symbolic skill and as such may be defined in terms of bi-directionality. A defining characteristic of a naming response may be that it is a response which is symmetrically related to its controlling stimulus. Proof of naming would require the formation of two symmetrically related components; not only must a particular stimulus control the subject's verbal response but also the subject's verbal response must exert control over his choice of that particular stimulus.

Traditionally, psychologists have struggled to define the essential characteristics of naming. Attempts have often ended up not by defining what naming is, but rather by defining what it is not (see, for example, Lock, 1980). There appears to be widespread agreement over what does not qualify as naming. Terms like paired associate, pure performative and conditional discrimination all refer to discriminative responding which bears a formal resemblance to naming but lacks the necessary symbolic relevance normally reserved for the term. Most of us recognise that, for example, a pigeon is not necessarily naming a stimulus to which it is responding differentially. Something other than this is required for naming, but when it comes to saying what this other property might be, talk tends to become vague or circular, and often naming becomes defined in terms of other equally elusive conceptual terms. What is needed is a definition of naming in terms of behaviourally specifiable events. Given that naming is a symbolic skill, and given that symbolic behaviour has been defined in terms of behaviourally specifiable properties such as symmetry, it is perhaps surprising that, to date, no one has explicitly defined naming as a kind of stimulus-response symmetry.

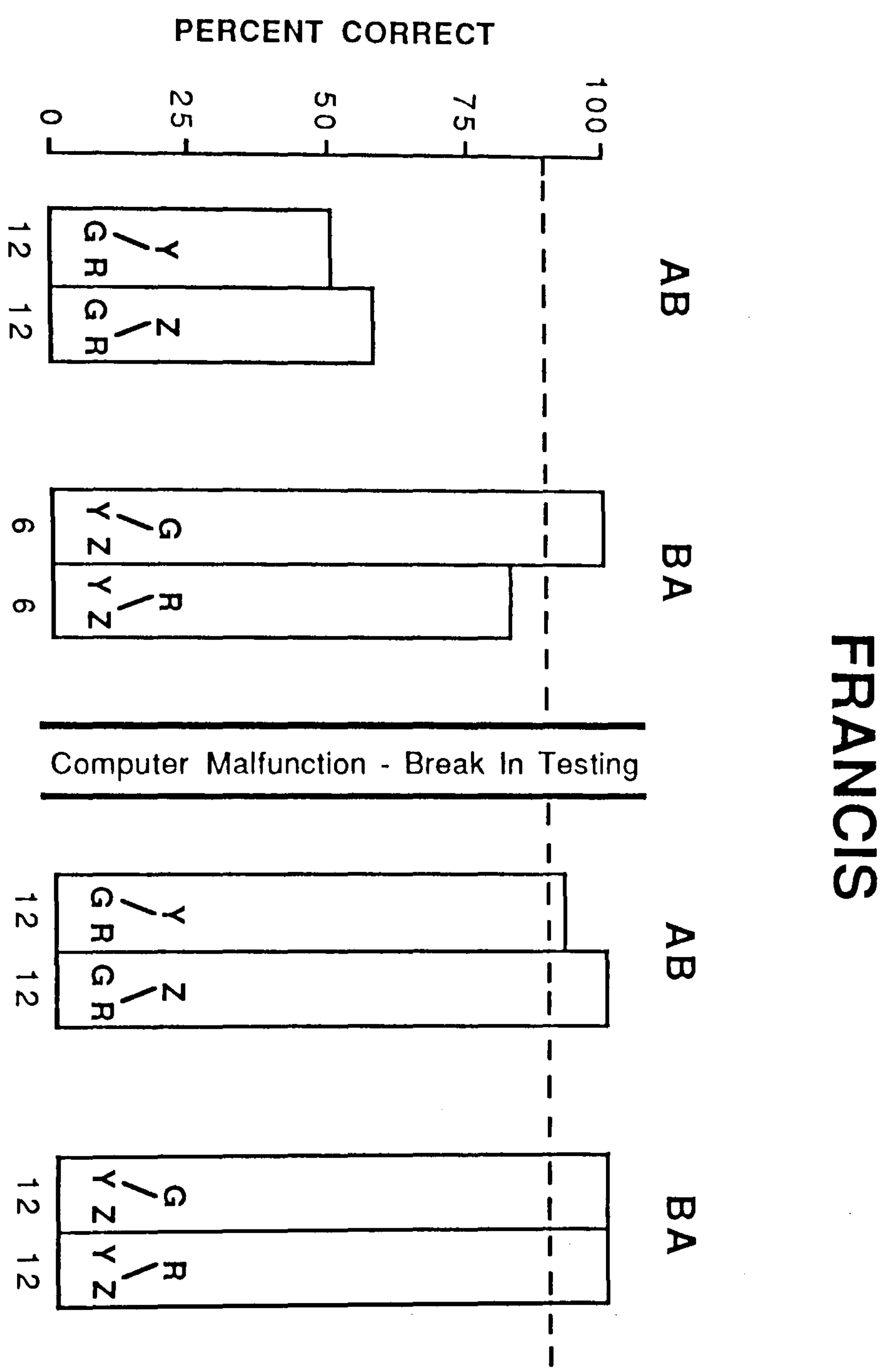
There is a certain face-validity in adopting this definition with respect to the current data. The words 'Omni' and 'Delta', spoken by the subjects, were found to be bi-directionally related to the Set-A shapes. According to the above definition, the subjects weren't just saying 'Omni' in the presence of the Y - shape and 'Delta' in the presence of the zig-zag; in so doing they were also naming the shapes. The same words, however, for five of the six subjects and for the most part of testing, were not symmetrically related to the Set-B colours. In other words, these five subjects were not initially naming the colours with the words 'Omni' and 'Delta'; they were merely saying those words in the presence of the colours (presumably because they already had other 'conventional' names for them). However, by the end of testing the subjects were, according to the definition, naming the colours as well as the shapes with the words 'Omni' and 'Delta', thus enabling the stimuli to become fully equivalent.

Other evidence of an independent nature supports the view that the subjects were naming the shapes but not the colours with the Set-X words 'Omni' and 'Delta'. This evidence will now be examined in the context of a detailed analysis of each individual subject's test results (including those from Sara and Nicola which have yet to be examined).

Francis

Francis produced what was perhaps the most revealing test data. The left side of Figure 4.19 shows his performance on the first 36 test trials, during which he scored above criterion on BA trials and at chance level on AB, despite having received twice as many AB than BA trials. This, then, indicated that he was naming the shapes but not the colours as 'Omni' and 'Delta' (where naming is defined as symmetry between stimuli and corresponding labels). This claim is supported by the manner in which he labelled the stimuli. Francis, unlike any other subject in this group, spontaneously labelled the comparisons as well as the samples. He began each trial by labelling the sample and pressing it to produce the comparisons. He then moved his finger over to one of the comparisons and labelled that too, again prior to pressing it. Francis always labelled the Set-A shapes consistently, irrespective of whether they appeared as samples on AB trials or as comparisons on BA trials. In each case he called the Y - shape an 'Omni' and the zig-zag a 'Delta'. However, his labelling of the Set-B colours was far from consistent. On the BA trials, the colours appeared as samples and he said 'Omni' to the green and 'Delta' to the red, exactly as he had been taught to do prior to testing. But

Figure 4.19 Francis's test performance. The figure depicts the percentage of correct responses for each trial-type. The stimuli for each trial-type are placed at the bottom of the bars. Sample stimuli are placed above comparisons and a line connects each sample to it's corresponding comparison. The number directly below each bar depicts the number of trials which that bar represents. The group of bars to the left depict Francis's performance prior to a computer malfunction and those to the right depict his performance after testing was resumed (see text).


FRANCIS
PERCENT CORRECT
0
25
50
75
100
AB
BA
Computer Malfunction - Break In Testing
AB
BA

when green and red appeared as comparisons on AB trials, Francis no longer called them 'Omni' and 'Delta' but gave them their conventional names instead.

So, it seemed that on BA trials Francis was able to select the correct comparison because the label he gave to the sample just happened to correspond with his label for that comparison. This correspondence was absent on AB trials. On AB trials the sample label bore no relation to the conventional labels which Francis spontaneously assigned to the comparison colours. His performance on the AB trials remained at chance level perhaps because the sample label did not provide a basis for choosing any particular comparison.

Francis's behaviour therefore supports the view that although he was able to say 'Omni' to the green sample and 'Delta' to the red he did not consider these words as being names for the colours. He already had conventional names for the colours prior to B'X' training. The data suggests that he was simply saying 'Omni' to the green sample and 'Delta' to the red in order to satisfy the requirements imposed by the experimenter. No such constraints were imposed upon the labelling of the comparisons, so when green and red appeared as comparisons on AB test trials Francis was free to label them with their conventional names. Indeed, this appears to be one way in which the subjects' pre-existing colour names may interfere with the 'Omni' and 'Delta' labels in their intended roles as common mediators of stimulus equivalence. It also shows that although the procedures establish common labelling with respect to stimuli in the sample position, this does not necessarily guarantee common labelling of the stimuli when they later appear in different locations, as comparisons.

However, Francis's behaviour during testing was consistent with the notion that he considered the words 'Omni' and 'Delta' to be names for the shapes. On BA trials, when the shapes appeared as comparisons, he continued to label them 'Omni' and 'Delta' despite the fact that there was nothing implicit in the trials to constrain him to do so.

Unfortunately, a computer malfunction brought his initial test session to a premature halt. When the session was resumed, Francis labelled both the shapes and the colours consistently with the words 'Omni' and 'Delta', and his performance on the AB trials rose above the overall criterion in line with his BA score (see the right-hand side of Figure 4.19). Just why he began to respond correctly upon resumption of testing will, of course, remain a matter for speculation. One possibility is that the sudden halt in testing was somehow construed by Francis as a cue that he had been doing something wrong, which may have in turn provoked a search for alternative forms of responding.

Linda

Linda's testing went through three distinct phases, only the last of which was characterised by common labelling of the samples. In the first phase of 48 test trials, labelling of both the Set-A and Set-B samples was absent, and both AB and BA matching was at chance level. On the next 48 trials, when the sample appeared the experimenter asked her to say what it was. The effect of these prompts on sample labelling and subsequent comparison choice is shown in Table 4.22.

Let us first examine the sample labels which were elicited by the prompting (see the left-hand column of Table 4.22). On AB trials, the sample shapes were labelled entirely consistently; the Y - shape was always called 'Omni' and the zig-zag was always called 'Delta'. However, on BA trials, when the colours were samples, Linda did not always label them as 'Omni' and 'Delta'. Instead, in the majority of BA trials she labelled the sample colours with their conventional names, 'Green' and 'Red'. Clearly then, there was already some

Table 4.22 Sample labels produced by Linda during her second AB and BA test phase (the column on the far right shows to what degree each sample label controlled her subsequent comparison choice (see text).

Trial-Type	Sample	Label	Number of times produced	Number of correct comparison choices following sample label
Y - G	Y	Omni	12	7
Z - R	Z	Delta	12	6
G - Y	G	Green	9	3
		Omni	2	2
		Delta	1	0
R - Z	R	Red	9	4
		Delta	3	3

indication that the words 'Omni' and 'Delta' were Linda's names for the shapes but not the colours. These suspicions were supported by the degree to which her sample labels controlled her comparison choices (see the right-hand side of Table 4.22). On AB trials, Linda's 'Omni' or 'Delta' label did not control her subsequent choice of comparison colour. So, although she occasionally labelled the colours as 'Omni' or 'Delta' on BA trials (thus showing that the B'X' relation was to some extent present), she could not, at this stage, do the reverse - she could not choose the colours according to her 'Omni' and 'Delta' labels (i.e. 'X'B was absent). The words were not symmetrically related to the colours, so, according to the definition advanced earlier, Linda was not naming the colours with those words. A different picture emerges from the BA trials. In the majority of BA trials Linda labelled the sample colours with their conventional names, 'Green' and 'Red'. These conventional labels had no control over her subsequent choice of comparison shape. However, on a few occasions Linda labelled the colours with the words 'Omni' and 'Delta' as she had been trained to do prior to testing. On three occasions she correctly said 'Delta' in the presence of the red sample, and on two trials she correctly said 'Omni' in the presence of the green sample. Most important of all, when she said 'Omni' to the green sample and 'Delta' to the red she always proceeded to choose the correct comparison shape. Even at this stage then, the words 'Omni' and 'Delta' appeared to be symmetrically related to the Set-A shapes (i.e. those words were apparently acting as names for the shapes). On AB trials the Set-A shapes (Y and zig-zag) controlled the Set-X words ('Omni' and 'Delta'), and on BA trials the reverse applied; the Set-X words controlled her choice of the Set-A shapes. The only reason BA did not emerge at this point appears to be because the Set-B colours were not consistently labelled with the Set-X words.

To recap, common labelling of the samples was mostly absent during Linda's first two phases of testing and neither the AB or BA relation emerged. Her scores over the first two phases were combined to produce her left-hand group of bars in Figure 4.16.

In her final phase of testing Linda was prompted with the question 'Is it Omni or Delta?' when the sample appeared. This prompt successfully established common labelling of the samples; Linda said 'Omni' to both the Y - shape and green samples, and 'Delta' to both the zig-zag and red samples. Consistent labelling of the colour samples (B'X') coupled with her pre-existing ability to select the correct shape conditional upon the label ('X'A) resulted in above criterion performance on BA trials. Linda's performance on the AB trials, however, remained below criterion until the last few trials of testing. Although she was labelling the sample shapes correctly (A'X'), those labels did not immediately exert control over her choice of comparison colour (i.e. 'X'B was initially absent). The absence of 'X'B therefore prevented the immediate formation of the AB relation, and confirmed that she was still not naming the colours with the Set-X words. However, by the end of testing, the AB relation had fully emerged. Somehow, in the intervening period the Set-X words had acquired the function of names for the colours. Future investigations may focus on this critical transition period in order to gain a better understanding of what might be contributing to the change from mere stimulus labelling to actual stimulus naming.

Jessica

Jessica, like Linda, also failed to give common labels to samples in her initial series of test sessions. Her left-hand group of bars in

Figure 4.16 shows her overall test performance without common labelling. These unshaded bars actually represent the combined scores from several distinct phases, outlined below.

Jessica was the only subject not to receive A'X' and B'X' training prior to the test. She learned the 'X'A and 'X'B relations instead, as depicted in Figure 4.15 . When the experimenter said 'Omni' she chose the Y - shape or green comparison, and when he said 'Delta' she chose the zig-zag or red comparison. Jessica was never required by the experimenter to label the stimuli. Her training had provided another route via which the AB and BA relations could emerge. The question was had the 'X'A and 'X'B training allowed the words to function as names for the corresponding shapes and colours? i.e. had the training resulted in the formation of symmetry between the Set-X words and the Set-A and Set-B stimuli? Only appropriate testing could determine the answer.

During her first 48 test trials Jessica was presented with only the AB and BA trial-types. Her performance remained at chance level. Perhaps the pre-requisite 'X'A and 'X'B relations had been of insufficient strength to mediate AB and BA matching so, subsequently, the AB and BA trials were presented together with 'X'A and 'X'B. Jessica made no errors on 24 'X'A and 24 'X'B baseline trials, but despite this her performance on AB and BA remained at chance level (48 trials each). In addition, she hardly ever spontaneously labelled the stimuli, apart from a few occasions when she called the colours by their conventional names. She never once labelled the stimuli as 'Omni' or 'Delta', either spontaneously or during a subsequent naming test (see above). So although Jessica had learned to select shapes and colours conditional upon a common label spoken by the experimenter, there was no evidence that she used the same words to name the stimuli, and the mere experience of hearing the common labels, distinct from producing them herself, was apparently insufficient for mediating the emergence of the AB and BA relations.

Nevertheless, it was still possible that Jessica's 'X'A and 'X'B training had created the potential for stimulus naming, a potential that might be realised under appropriate environmental conditions. So Jessica was shown each shape in turn on the centre window and was asked 'Is it Omni or Delta?' No reinforcers were given during this stage. This then appeared to be a specific test of A'X', the symmetrical counterpart of the trained X'A' relation. Jessica received 21 trials and made no errors. When the Y - shape appeared she correctly said 'Omni' and when zig-zag appeared she correctly said 'Delta'. The A'X' relation had apparently emerged from her earlier 'X'A training. At this stage, then, the shapes appeared to be symmetrically related to the Set-X words and thus they (the words) were acting as names for the shapes. This symmetry was apparently not present before but had itself emerged within the highly contrived context of the A'X' test.

To find out the effect of this newly established naming skill on Jessica's AB and BA performance, a further test was given. Once again she scored 100% on 'X'A and 'X'B trials (4 trials each) but her AB and BA score remained at chance level (8 trials each). But then there was also a complete absence of labelling.

Because Jessica's 'X'A and 'X'B score had been perfect throughout, and in an attempt to increase the rate at which AB and BA trials were presented, a decision was taken to drop the 'X'A and 'X'B trials from the test. In addition, sample labelling was further promoted by prompting Jessica to 'say what it is' when the sample appeared. Once again this had little effect upon her matching performance; her AB and BA score fell below chance level over the 20 test trials. The prompt, however, did have the effect of establishing consistent sample labelling. On AB trials, the Y sample was called 'Omni' and the zig-zag was called 'Delta'. However, on BA trials the green and red samples were given their conventional names, 'Green' and 'Red'. Jessica never overtly labelled the colours as 'Omni' or 'Delta', which suggested that, thus far, those words were not functioning as names for the colours. By now Jessica had received a grand total of 92 AB and 88 BA test trials. Two factors had remained constant throughout the variations in testing: her AB and BA performance remained at chance level and common labelling of the samples <u>never occurred</u>.

To recap, the data so far suggested that Jessica was naming the shapes but not the colours with the Set-X words. However, those words had come to function as names for the shapes by virtue of very specific conditions, conditions which had not yet been applied to the colours. So, in the final phase of testing Jessica was prompted with the question 'Is it Omni or Delta?' when the shapes <u>and the colours</u> appeared as samples on matching trials. So, on the first part of any BA test trial, Jessica was required to label the Set-B colour sample with a Set-X word; this was itself a specific test of B'X', the symmetrical counterpart of the trained 'X'B relation. Under these conditions, Jessica never labelled the colours incorrectly; on each BA test trial she either labelled the green sample as 'Omni' or the red sample as 'Delta'. It now <u>seemed</u> as if Jessica was naming the colours with the Set-X words, because the emergence of B'X' itself indicated symmetry of the trained 'X'B relation, a symmetry which had probably emerged by virtue of the specific context of prompting.

Meanwhile, on AB trials, Jessica was also labelling the Set-A sample shapes with the appropriate Set-X word. So, in this final test phase (depicted by the shaded bars in Jessica's portion of Figure 4.16), prompting had elicited common labelling of the samples. Furthermore, the very fact that Set-X labels emerged at all suggested that those labels were functioning as names for both shapes and colours. However, all was not what it seemed. In this final phase of testing, the BA relation emerged straightaway whereas the AB relation did not (see Jessica's portion of Figure 4.16). This in turn indicated that although the Set-X words were acting as names for the shapes, the same words were not initially functioning as names for the colours (see the Preliminary Discussion).

Jessica's data may appear to be inconsistent, but, upon reflection, it need not be so. It seemed that Jessica was able to name the colours prior to the final test phase, because the B'X' relation had previously emerged. However, on B'X' trials the Set-B colours were confined exclusively to one position, namely the sample position. Perhaps, then, the process of naming was also confined to colours appearing in the sample position, and thus did not extend to the same colours when they appeared as comparisons. This account, although speculative, is entirely consistent with Jessica's initial failure to match on AB trials during the final phase of testing. On AB trials the Set-X word (spoken by Jessica in response to the Set-A sample) did not initially control her choice of comparison colour. Perhaps this occurred because the colours were still not being named as 'Omni' and 'Delta' when they appeared 'out of position', as comparisons. Perhaps the comparison colours were still being named (albeit covertly) as 'Green' and 'Red', thus preventing the 'Omni' and 'Delta' labels from assuming their intended roles as mediators on matching trials (c.f. Francis's data above).

Whatever the case, further investigation was precluded because by the end of this final test phase (i.e. after 48 AB and 48 BA trials) Jessica's AB matching performance had become completely error-free. At that stage the Set-X words were functioning fully as names for both the shapes and the colours, regardless of the positions in which these stimuli appeared.

Alex

Figure 4.16 shows that Alex produced common labels throughout testing and acquired the BA relation prior to AB. Further detail appears in Figure 4.20 . The BA advantage emerged almost immediately as shown on the left-hand side of the figure. Again, the failure of AB to emerge revealed that the words 'Omni' and 'Delta', spoken by Alex in response to the Set-A samples, did not exert control over his choice of Set-B colour (i.e. he was not naming the colours with the Set-X words).

This first test block was followed by what is commonly known as a test of verbal comprehension. The subject was given the following instructions prior to the test:

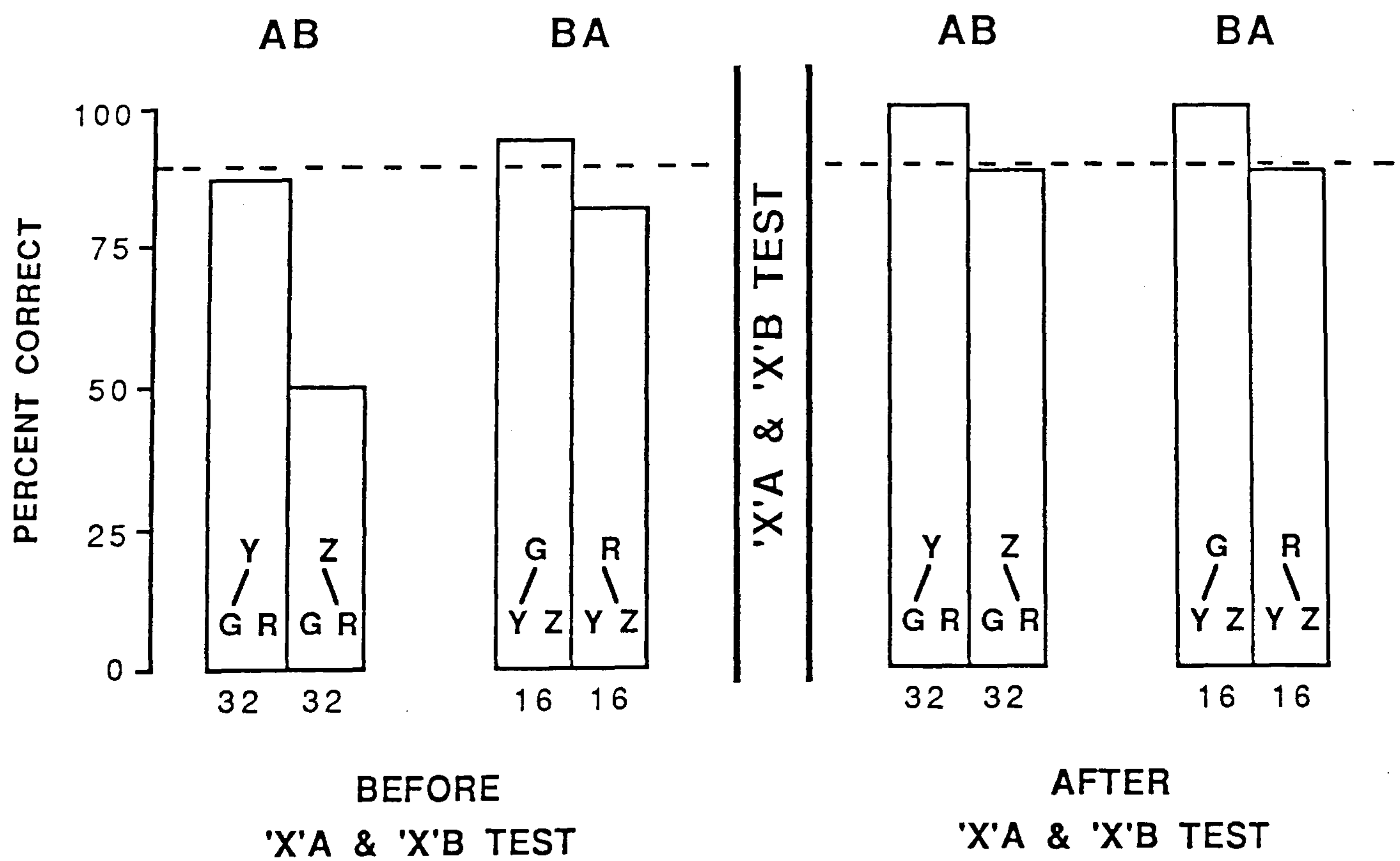
'This time you will not see anything come on here' (experimenter points to centre window). 'When you hear the beep, get ready and listen carefully. I will say a word. When I have said the word I want you to press this window (experimenter points to centre window) and then choose the right one.'

The Set-X words 'Omni' and 'Delta' were spoken by the experimenter and Alex had to choose between the Set-A shapes on 'X'A trials and the Set-B colours on 'X'B trials. Alex never made an error on any of the 24 unreinforced test trials.

The comprehension test had apparently set the occasion for the emergence of the very skill which had been missing from Alex's earlier AB and BA test performance, namely, choosing Set-B colours conditional upon the Set-X words. Apparently, this 'X'B relation had itself emerged from its symmetrical counterpart B'X', the labelling relation

Figure 4.20 Alex's test performance. The group of bars to the left depict his performance on AB and BA test trials prior to an 'X'A and 'X'B test, and those to the right depict his performance after the 'X'A and test (see text).

ALEX


AB
BA
AB
BA
PERCENT CORRECT
100
75
50
25
0
'X'A & 'X'B TEST
Y
G R
Z
G R
G
Y Z
R
Y Z
32 32
16 16
32 32
16 16
BEFORE
'X'A & 'X'B TEST
AFTER
'X'A & 'X'B TEST

established through Alex's earlier training. The presence of symmetry suggested that Alex was now capable of naming the colours with the Set-X words. The question was why should the requisite 'X'B relation emerge in the context of a comprehension test and not in the context of the preceding AB and BA test?

One possible answer is that under comprehension testing the context was more obviously 'instructional' than it was during AB and BA testing. Under comprehension testing, several factors may have made it more obvious to the subject that the comparison colours should be selected conditional upon the Set-X words i.e. that the words were meant to instruct the subject to choose a particular colour. During comprehension testing the experimenter was saying the Set-X words, whereas on AB and BA test trials the subject said the words himself. Since adults are an important source of instructional control over children's behaviour, one perhaps should not be too surprised if a child does not react to the instructional function of the Set-X words when he himself produces them, but does react accordingly when they are produced by the experimenter. Also, not only were the Set-X words spoken by the experimenter during comprehension testing but, because of this, they were not preceded by the presentation of any associated stimuli. However, on each AB and BA test trial the Set-X word <u>was</u> preceded by an associated stimulus, i.e. the sample. Furthermore, the word was produced as a <u>response</u> to the sample and so it (the word) was perhaps less likely to also act as a <u>stimulus</u>, instructing the subject to choose a particular comparison. On comprehension trials, however, the word was produced by the experimenter and so it would seem to have no function for the child other than as a stimulus to which one should respond accordingly. Finally, because the comprehension trials were of a different structure to those previously presented, the subject was given minimal instructions prior to testing (see above). Although minimal, these instructions may have alerted the subject to the instructional function of the Set-X words. In particular, the experimenter told the subject to 'listen carefully'. In contrast, the AB and BA test was not preceded by any experimental instructions but was simply presented as a 'normal' session.

Some or perhaps even all of these factors may have contributed to the emergence of 'X'B during Alex's comprehension test. Nevertheless, the data so far suggested that although the Set-B colours and Set-X words had not been symmetrically related on AB and BA test trials, Alex's former B'X' training had created the potential for symmetry, a potential that remained unrealised until he was placed within the unique context of the comprehension test.

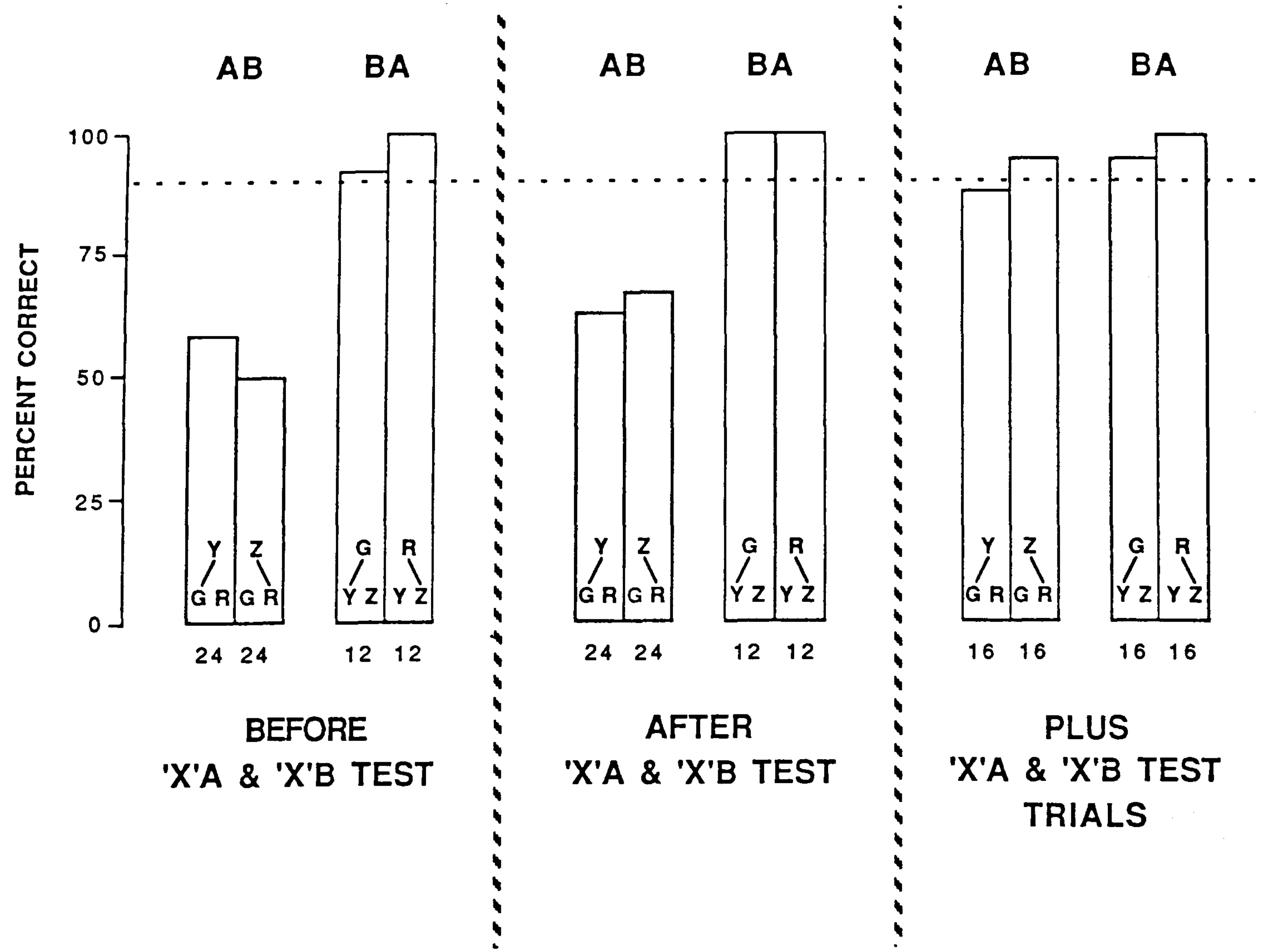
When Alex was finally returned to the AB and BA test his resulting behaviour was consistent with his new found skill. His AB performance immediately rose to criterion for the first time, as depicted in the right-hand group of bars in Figure 4.20.

Michael P.

Michael was also given an 'X'A and 'X'B comprehension test after AB (but not BA) failed to emerge in his initial block of test trials (see Figure 4.21). He, like Alex, made no errors on the comprehension test. However, unlike Alex, Michael's AB score did not immediately improve after the comprehension test. The AB relation did finally emerge in his last phase of testing when 'X'A and 'X'B trials were interspersed with the AB and BA trials. This suggests (but by no means proves) that the emergence of AB may have been a consequence of the interspersed comprehension trials.

Figure 4.21 Michael P.'s test performance during three phases of testing. The left-hand group of bars depict his AB and BA score on the first test phase, prior to an 'X'A and 'X'B test. The middle group of bars depict his AB and BA scores from the second test phase following an 'X'A and 'X'B test. The group of bars to the far right depict his AB and BA scores in the final test phase, when 'X'A and 'X'B test trials were interspersed amongst the AB and BA test trials (see text).

MICHAEL P.


AB
BA
AB
BA
AB
BA
PERCENT CORRECT
100
75
50
25
0
Y Z G R
G R G R Y Z Y Z
24 24 12 12
24 24 12 12
16 16 16 16
BEFORE
'X'A & 'X'B TEST
AFTER
'X'A & 'X'B TEST
PLUS
'X'A & 'X'B TEST
TRIALS

Stephen R.

Of the six subjects under consideration, Stephen was the only one to acquire AB and BA relations at the same rate. In fact, Figure 4.16 shows that he never once made an error during testing.

Stephen's behaviour was fundamentally different from the others on two additional counts. Firstly, he was the only subject to spontaneously and consistently label the shapes prior to the introduction of the words 'Omni' and 'Delta'. On several occasions he called the Y - shape a 'Y' and the zig-zag a 'Spring' as well as calling the colours 'Green' and 'Red'. Spontaneous labelling is of particular interest because it was central to understanding why the five other subjects initially failed AB but passed BA. Their data suggests that they failed AB because they already had names for the Set-B colours, and that these names interfered with the formation of symmetry between those colours and the Set-X words 'Omni' and 'Delta'. The subsequent absence of the 'X'B relation prevented AB from emerging via mediation of the 'Omni' and 'Delta' labels. Now, given that Stephen had his own names for both the colours and the shapes, and given that test failure was associated with the prior existence of names for the test stimuli, one would have perhaps predicted that Stephen might fail the AB and BA tests. Instead, he passed.

The fact that he passed both AB and BA may have been a consequence of the way in which he labelled the stimuli during A'X' and B'X' training. Stephen, unlike any other subject, spontaneously applied his own labels to the stimuli as well as those required by the experimenter. On A'X' trials he said 'Y - Omni' in the presence of the Y - shape, and 'Spring - Delta' in the presence of the zig-zag. On B'X' trials green was labelled 'Green - Omni' and red was called 'Red - Delta'. Stephen continued to label in this manner right the way through testing. It therefore seems that Stephen's data is not inconsistent with that of the other subjects. His own labels seemed to enhance rather than interfere with the mediating effects of the Set-X labels. On AB and BA test trials, he always gave his own name to the sample first, followed by the corresponding Set-X label. If his own labels played a key role in the production of the mediating Set-X labels then it is difficult to see how, at the same time, the former could interfere with the effects of the latter.

Sara and Nicola

Unlike the other children, both Sara and Nicola learned the AB relation prior to testing. The BA trials therefore tested for symmetry of the AB relation. Figure 4.22 shows that Sara passed the symmetry test. Her performance on the AB baseline trials and the BA symmetry test trials was uniformly excellent throughout. She never once spontaneously labelled any of the stimuli. However, in a subsequent naming test she gave common labels to the corresponding Set-A and Set-B stimuli; the Y - shape and the green were called 'Omni' and the zig-zag and red were called 'Delta'. This was rather surprising because, in sessions prior to testing, she had always called the colours by their conventional names, while apparently reserving the words 'Omni' and 'Delta' for the shapes. Furthermore, earlier evidence suggested that her AB performance prior to testing was governed by the verbal rules 'Omni is Green' and 'Delta is Red'.

There are two alternative explanations for Sara's success on the symmetry test, as illustrated in Figure 4.23 (N.B. this figure only depicts one stimulus pair, namely the Y-shape and green stimuli, although the following discussion applies equally to the other stimulus pair, zig-zag and red). One possibility is depicted on the left-hand

Figure 4.22 Sara R's test performance on AB baseline and BA test trials. (each bar represents 24 trials).

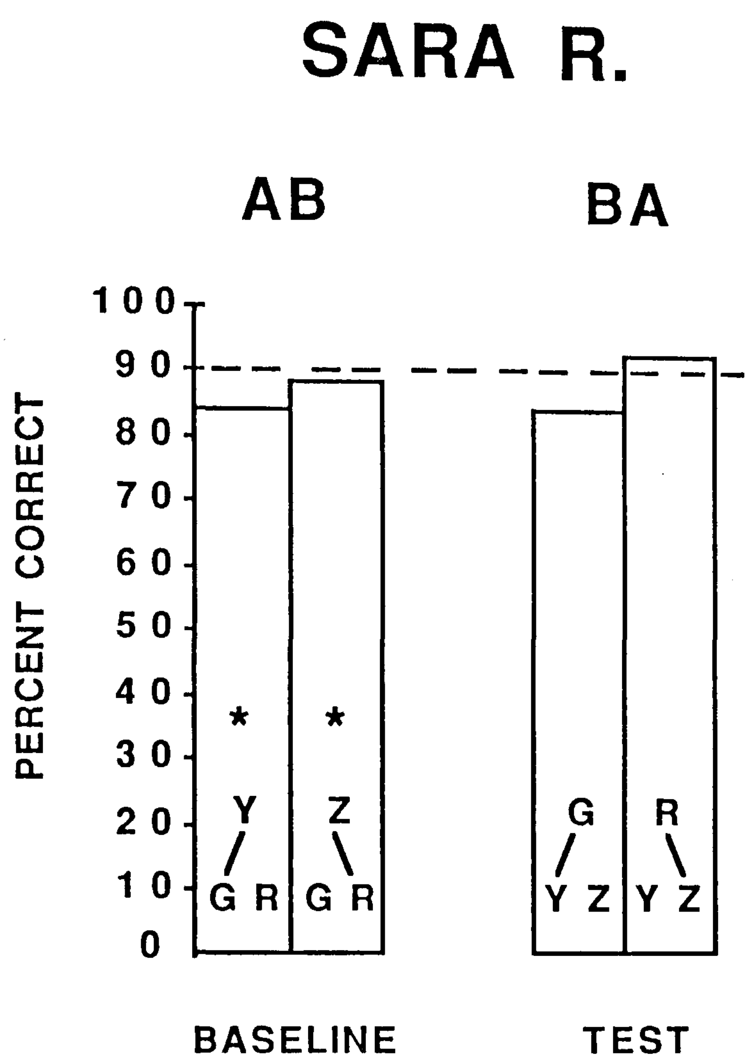
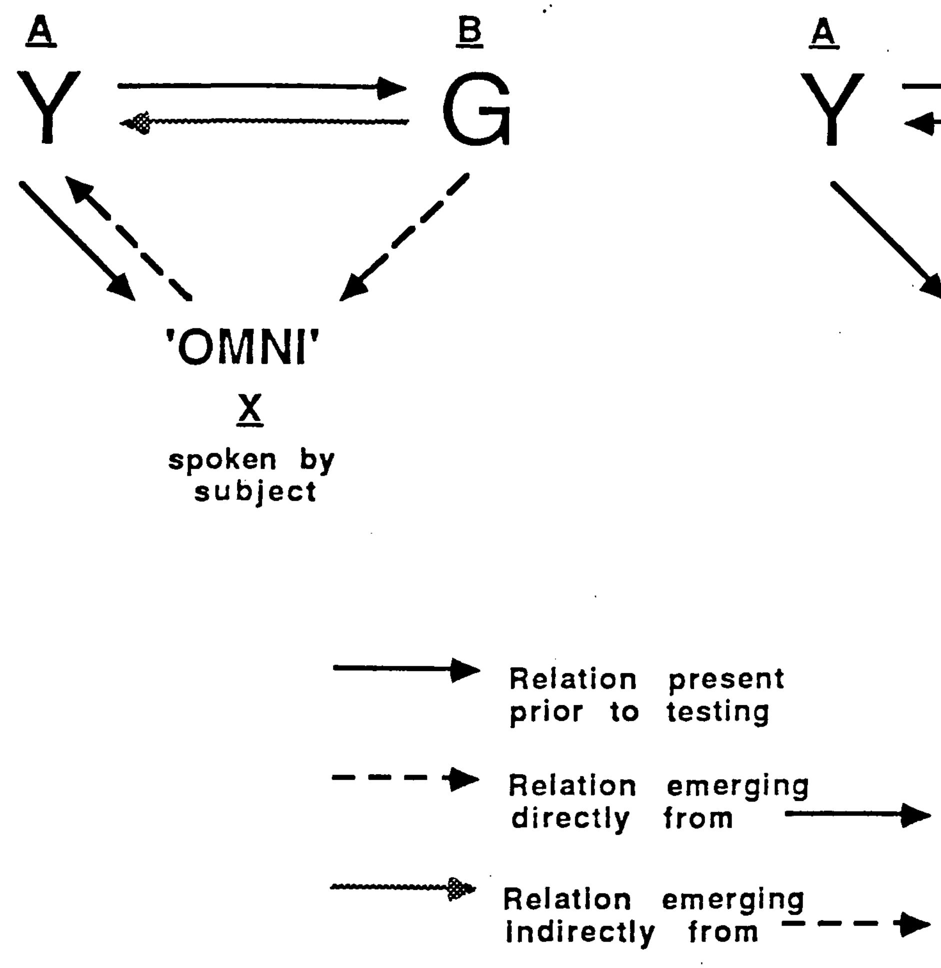

SARA R.
AB
BA
PERCENT CORRECT
100
90
80
70
60
50
40
30
20
10
0
*
*
Y
Z
G R
G R
G
R
Y Z
Y Z
BASELINE
TEST

Figure 4.23 Possible relations between Set-A shapes, Set-B colours and Set-X words during Sara R's test sessions (see text).


A
B
Y
G
'OMNI'
X
spoken by subject
A
B
Y
G
'OMNI'
X
spoken by subject
Relation present prior to testing
Relation emerging directly from
Relation emerging indirectly from

side of Figure 4.23 ; if Sara had covertly labelled the Set-B colours with the Set-X words then it would have been possible for common labelling to bring about the emergence of the BA relation (B'X' and 'X'A producing BA). The other possibility is depicted on the right-hand side of Figure 4.23. If the AB training had directly resulted in the formation of a symmetry relation, then BA would have emerged first. The BA relation could then bring about labelling of the Set-B colours with the Set-X words (BA and A'X' producing B'X'). So the question is, did common labelling bring about the emergence of symmetry or did symmetry bring about the emergence of common labelling? Sara's silence during the symmetry test prevents one from answering this question.

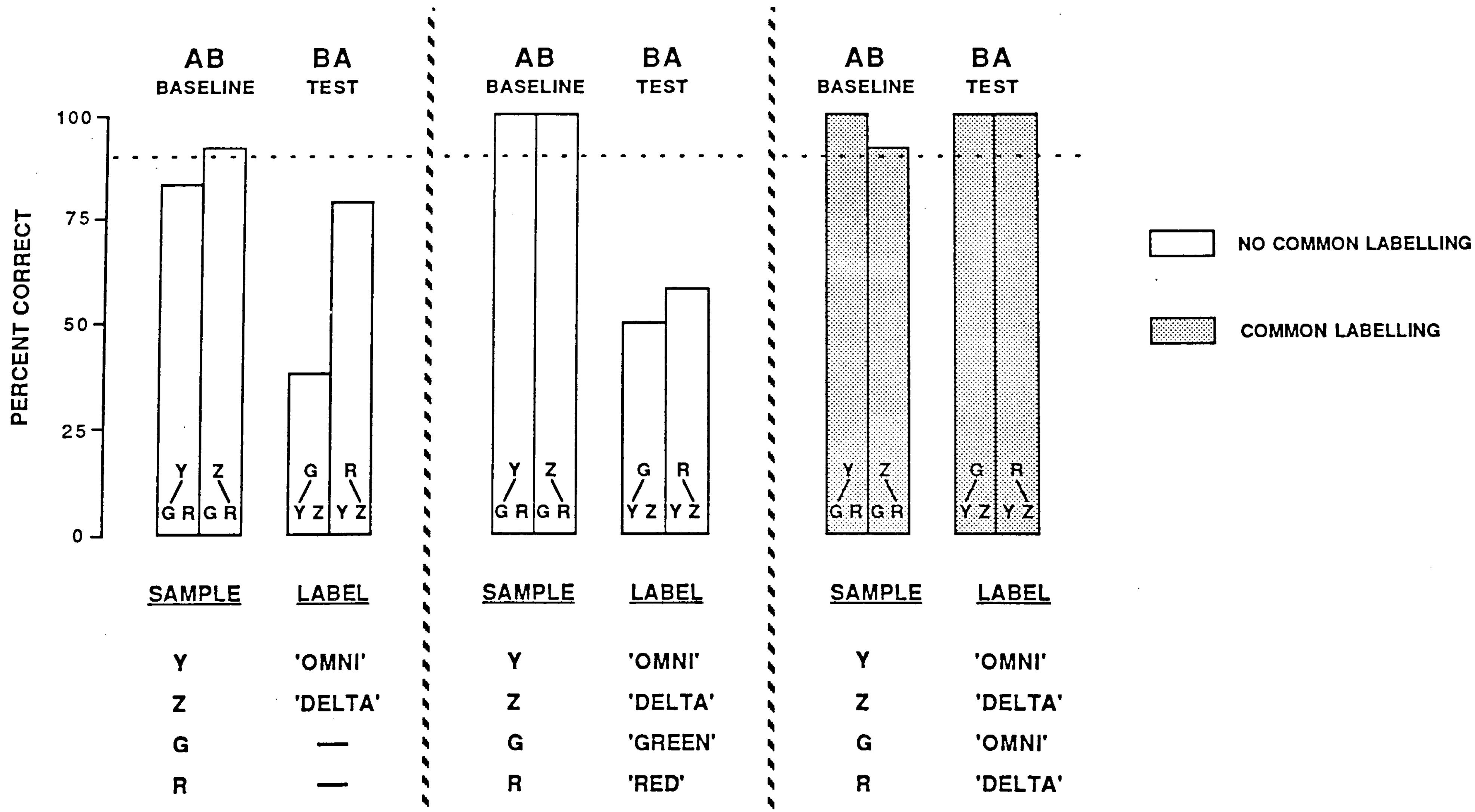
Fortunately, firmer conclusions may be drawn from Nicola's test data presented in Figure 4.24. In her first test series (depicted by the left-hand group of bars) Nicola's response to the AB baseline trials was exactly the same as it had been prior to testing. She continued to respond correctly on the AB baseline trials, and she also continued to label the Set-A samples with the words 'Omni' and 'Delta'. However Nicola failed the symmetry test; her BA score was poor. At this stage then, the corresponding Set-A and Set-B stimuli were not equivalent and the AB training alone was clearly not sufficient for symmetry.

In the following session (depicted by the middle group of bars in Figure 4.24) additional prompting was given on BA trials; when the red or green sample appeared, Nicola was asked to 'Say what it is'. She responded by labelling the colours by their conventional names. This additional labelling, however, had no effect upon her test performance; she still failed the BA test despite scoring 100% on the AB baseline trials.

Further testing was preceded by the establishment of common labelling. Nicola was first trained to label the Set-B colours with

Figure 4.24 Nicola's test performance on AB baseline and BA test trials during her three test phases (see text). Each bar represents 24 trials. The table below each set of bars depicts Nicola's overt sample labels from that phase of testing.

NICOLA



AB BASELINE
BA TEST
AB BASELINE
BA TEST
AB BASELINE
BA TEST
PERCENT CORRECT
100
75
50
25
0
NO COMMON LABELLING
COMMON LABELLING
SAMPLE
LABEL
Y 'OMNI'
Z 'DELTA'
G —
R —
SAMPLE
LABEL
Y 'OMNI'
Z 'DELTA'
G 'GREEN'
R 'RED'
SAMPLE
LABEL
Y 'OMNI'
Z 'DELTA'
G 'OMNI'
R 'DELTA'

the Set-X words, 'Omni' and 'Delta'. She required only 24 trials to learn to say 'Omni' in the presence of green and 'Delta' in the presence of red. Finally, Set-A and Set-B trials were combined, and Nicola continued to label them appropriately with the Set-X words.

In the final test session (depicted by the shaded bars to the far right of Figure 4.24), additional prompting was given; when the samples were presented, Nicola was asked 'Is it Omni or Delta?'. This prompt had the desired effect; Nicola continued to label the Set-A and Set-B samples with the corresponding Set-X labels. Furthermore, this common labelling had a dramatic effect on her test performance. Nicola made no errors whatsoever on the BA test trials.

These data provide additional evidence for the effects of common labelling on the formation of stimulus equivalence. Of course, one could argue that the BA relation emerged independently of common labelling, perhaps merely as a function of repeated testing. However, this possibility is highly unlikely. The BA relation never emerged in the absence of common labelling, despite the fact that the subject received a total of 96 BA test trials. Although relations have been known to emerge as a function of repeated testing, when this has occurred it has tended to be a gradual emergence over the course of several test sessions (Devany, Hayes and Nelson, 1986; Sidman, Kirk and Willson-Morris, 1985). In contrast, Nicola's BA performance immediately became error-free when common labelling was introduced.

DISCUSSION

This experiment produced some noteworthy results. Prior to testing, six subjects failed to learn AB matching. These subjects weren't just left to interact with the basic matching-to-sample contingency; every effort was made to get them to acquire the task through several standard intervention procedures, none of which proved effective. Placed against this background, the subjects' performances during testing were all the more astonishing. All six later proved capable of not only AB but also BA matching, in the complete absence of differential reinforcement or feedback of any kind. The AB and BA relations emerged, not spontaneously, but as a function of learning common labels for the corresponding Set-A and Set-B stimuli. Clearly, common labelling greatly facilitated the arbitrary matching of two sets of visual stimuli.

One should not lose sight of the practical significance of this finding. Others working in this area have ascribed a level of causal impotency to the role of common labelling in equivalence formation (Lazar, Davis-Lang and Sanchez, 1984; Sidman, Willson-Morris and Kirk, 1986). The adoption of such a stance provides a major justification for the teaching of conditional discriminations (such as those involved in reading) by automated methods, which involve only the presentation of stimuli, and do not require the subject to name the stimuli orally. The present data suggests that those who advocate the use of teaching machines through their theoretical formulations of equivalence may have been a little premature in overshadowing the role of common labelling. None of the subjects in this particular experiment benefited from direct contact with fully automated matching procedures, even when that translated into hundreds of reinforced trials. This initial teaching problem, though at times seemingly insoluble, simply ceased to exist once the experimenter took the relatively simple step of getting the subjects to apply common labels to the stimuli which they had earlier failed to match.

Jessica's performance in her initial stages of testing took on special significance, because she failed both the AB and BA tests despite making no errors on her 'X'A and 'X'B baseline trials. Previous experiments have shown that two sets of visual stimuli may become equivalent if the same word, spoken by the experimenter, controls the subject's choice of corresponding stimuli from each set (see Lazar et al, 1984, for a review). This has prompted Sidman and colleagues to claim that common naming may be successful in bringing about equivalence not through naming per se, but simply because the two stimuli have both become associated with the product of the name i.e. its sound. (Sidman, Willson-Morris and Kirk, 1986). Jessica's results, however, contradict this supposition. Although she had learned to select a shape and a colour conditional upon hearing a common sound (the Set-X word 'Omni' or 'Delta' spoken by the experimenter), this experience was not sufficient for mediating the emergence of the AB and BA relations. These relations emerged only in later tests, when Jessica was prompted to produce the sounds herself through common labelling.

But the most noteworthy feature of the present experiment was that although both AB and BA emerged when each of six subjects were taught common labelling, BA emerged prior to AB in all but one case. This finding seems to elude explanation in terms of present theoretical accounts of stimulus equivalence, especially those that deny the role of naming. Prior to this experiment there appeared to be two possible outcomes of an equivalence test. A subject either passed or failed the test with respect to a given set of stimuli. To these we must add a third possibility, for some of the subjects in the present experiment did both; they passed and failed a test with the same set of stimuli. This is precisely what happened when Alex, Francis, Linda, Michael P. and Jessica were initially tested under common labelling conditions. The point may be best illustrated by considering what may have been concluded had these subjects been tested on only a single matching task rather than two. If these subjects had been given a 24-trial test of AB only, then we may have concluded that the Set-A and Set-B stimuli had not become equivalent, because the AB relation had failed to emerge after this number of trials (see Figure 4.17). Conversely, if one had instead presented the same number of BA test trials then totally the opposite conclusion may have been drawn (i.e. that the stimuli were equivalent after all, a conclusion which is forced by the emergence of BA).

These data, then, appear to present something of a dilemma. But if one acknowledges naming, and particularly the definition of naming advanced in this thesis, then the 'dilemma' ceases to exist. Instead, the data begins to make sense, and what may otherwise have been troublesome variability ends up as an interesting phenomenon in its own right.

Finally, we must consider the significance of both Nicola's and Sara's results. These two differed from the others by learning the AB relation prior to testing. Their AB trials formed a baseline which was tested for symmetry by the inclusion of the BA trials.

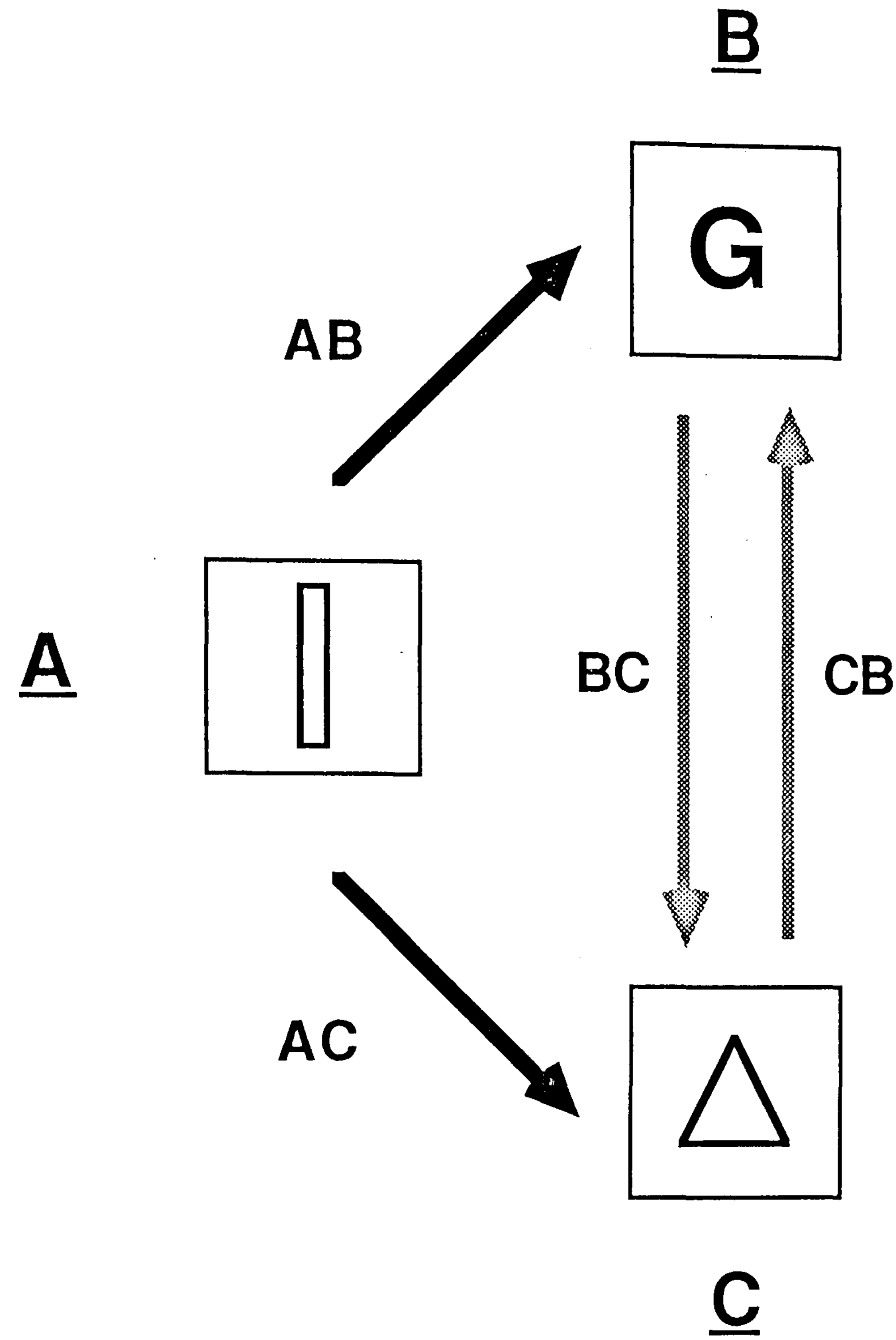
Nicola's test data is particularly significant because it extends the role of common labelling to beyond the mere facilitative. In her initial test phase, her performance on the AB baseline trials was good but she failed the symmetry test; at this stage, then, the corresponding Set-A and Set-B stimuli were not equivalent, and the AB training alone was clearly not sufficient for symmetry. Symmetry also failed to emerge when Nicola applied unique labels to each of the samples, thus demonstrating that labelling *per se* was not sufficient for equivalence. However, when common labelling of the samples was introduced the BA relation at last emerged. The Set-A and Set-B stimuli became equivalent not directly through the AB relation, but indirectly through the mediation of the Set-X labels.

This is not the first time a labelling intervention has been necessary for the formation of equivalence. Beasty and Lowe (1985) have shown that children (younger than four years of age) who initially failed equivalence tests later passed when they were taught to name the sample-comparison pairings whilst responding to baseline training trials (see Figure 4.25). For example, on some baseline trials subjects were required to match a vertical line sample to a green comparison (AB) and a vertical line sample to a triangle comparison (AC). Equivalence tests then assessed whether the subjects could match green to triangle (BC) and triangle to green (CB). After failing these tests the subjects were taught to say 'Up - Green' on AB baseline trials and 'Up - Triangle' on AC baseline trials. This intervention resulted in the immediate emergence of the BC and CB test relations.

Perhaps common naming was the active ingredient in the verbal intervention used in the Beasty and Lowe experiment. The word 'Up', which was spoken by the subjects in the presence of both the triangle and the green stimulus, may have functioned as a common name through which stimulus equivalence emerged.

But common naming need not be the only way in which language may promote stimulus equivalence. In Experiment 3 it was suggested that equivalence may emerge if the subject labels not only the stimuli but also the <u>relation</u> between the stimuli. This possibility is supported by Sara's data from the current experiment. Sara's AB matching prior to testing appeared to be governed by the rules 'Green is Omni' and 'Red is Delta', which she verbalised on several occasions. Sara had linked the stimuli verbally via the relational word 'is', and this alone may have been sufficient for her subsequent success on the symmetry test (although common naming too cannot be ruled out; see

Figure 4.25 Stimulus relations from the experiment by Beasty and Lowe (1985). Arrows point from samples to corresponding comparisons. Black arrows depict trained relations and shaded arrows depict relations assessed during testing.


B
G
AB
A
BC
CB
AC
C

Results). In similar manner, subjects from other experiments have apparently formed equivalence on the verbal plane prior to any formal test of equivalence. In studies by Hird and Lowe (1985), several mentally retarded subjects linked stimuli with relational terms such as 'is the same as' or 'goes with' after having learned to match those stimuli on baseline training trials (see also Beasty, 1987). This tactic appears to be a deal more sophisticated than common naming, but no less verbal in origin.

Nevertheless, common naming may represent the simplest means by which two or more stimuli may become equivalent, and as such it deserves further examination.

EXPERIMENT 4(b)

In Experiment 4(a) it was shown that two unidentical stimuli, a shape and a colour, could become equivalent, provided:

(a) The subject applied a common label to the shape and colour <u>and</u>

(b) That in so doing the subject was <u>naming</u> the shape and colour with the common label (where naming is defined as a bi-directional relation between the labelling response and its referent stimulus).

In the initial stages of testing, most of the subjects in Experiment 4(a) were, according to the definition above, naming the shapes but not the colours with the words 'Omni' and 'Delta'. The initial absence of bi-directionality (or symmetry) between the colours and the subjects' labels in turn led to an initial performance decrement on one of the test relations, AB. It was hypothesised that the colours and spoken words 'Omni' and 'Delta' were not initially bi-directionally related (ie the words 'Omni' and 'Delta' were not initially acting as names for the colours) because the subjects already had other names for these stimuli; they all called the colours by their conventional names, 'Green' and 'Red'. It seemed that these conventional names somehow interfered with the attempt to establish common naming with the Set-X labels, 'Omni' and 'Delta'.

If the interference occurred because the subjects were already capable of naming the colours with words other than those used as common Set-X labels, then one would anticipate no interference if the subjects have no other names for the colours prior to the experiment. Given that a subject is unable to consistently label both the Set-A shapes and the Set-B colours prior to the establishment of common labelling with Set-X words, then one would predict no interference; the Set-X words should be free to act as true names for the stimuli, thus enabling the AB and BA relations to emerge at the same rate during testing. Experiment 4(b) tested this prediction.

METHOD

SUBJECTS

Of the 4 - 5 year old children available, only two were unable to consistently name not only the Set-A shapes, but also the Set-B colours. These two subjects, Gareth and Peter, were otherwise judged to be of normal ability by their teacher, and this is supported by their above average scores on the expressive language component of the Reynell language test (see Table 4.23).

Assessment of naming skills

Both subjects were given extensive naming tests after their teacher had indicated that, in her opinion, they were unable to consistently name any colour. Both Set-A shapes and Set-B colours appeared one at a time in the centre window of the five key panel, and verbal responses were elicited from the child in accordance with the procedures outlined in the previous experiment. Table 4.24 lists the subjects' verbal responses and the number of times each response occurred. Neither child differentially labelled the shapes. The colours were also labelled indiscriminately. For example, Gareth said 'Red' and 'Blue' with equal frequency to the green stimulus. Although he occasionally labelled the red stimulus correctly, on the majority of trials he said 'Blue' when red was presented. On all but five of the 24 colournaming trials, Peter said 'Blue' to both the green and red

Table 4.23 Chronological age (in years-months) and equivalent language age (from the Reynell Developmental Language Scale) for the two subjects in Experiment 4(b).

SUBJECT	CHRONOLOGICAL AGE	EQUIVALENT LANGUAGE AGE	
		VERBAL COMPREHENSION	EXPRESSIVE LANGUAGE
GARETH	4 - 2	3 - 8	5 - 11
PETER	4 - 3	3 - 9	5 - 6

Table 4.24 Verbal responses elicited from Gareth and Peter during their naming tests.

	Stimulus	Verbal Response	Frequency
GARETH	Y - Shape	'Don't know'	6
		No response	6
	Zig-Zag	'Don't know'	7
		No response	5
	Green	'Red'	6
		'Blue'	6
	Red	'Blue'	7
		'Red'	5
PETER	Y - Shape	'Don't know'	9
		Shakes head	3
	Zig-Zag	'Don't know'	7
		Shakes head	5
	Green	'Blue'	9
		'Red'	3
	Red	'Blue'	10
		'Red'	2

hues. Both subjects' patterns of verbal responses therefore indicated that they had learned two colour words, 'Blue' and 'Red', but they had not yet learned which colours those words signified.

The subjects' difficulty with colour terms appeared to be present throughout the experiment. Both children, like the others, were given a Reynell language test at the end of the experiment. On comprehension items, the child was required to respond to a question or instruction from the experimenter. In nine of these items, a colour name constituted a critical part of the question or instruction put to the child. Both Gareth and Peter had particular difficulty with these items and thus their verbal comprehension scores were relatively poor. For example, one section involved coloured pencils - two long pencils (one red and one blue) and three short pencils (one red, one yellow, and one blue). When Peter was instructed to 'Find a yellow pencil', he pointed to a red pencil. When the experimenter asked 'Give me the longest red pencil', Peter selected the long blue one! After he had put the short pencils into a box, Gareth was asked 'Which red pencil has not been put away?'. He pointed to the long blue pencil. Of the nine commands involving colour names, Peter never once responded correctly to any of them, and Gareth only responded correctly to two of them. Two weeks after the experiment, Gareth's and Peter's teacher confirmed that they had still not learned any colour names.

APPARATUS (see General Method section).

PROCEDURE

The procedures for training and testing were identical to those used in Experiment 4(a); both subjects followed the main route to testing (see Figure 4.8).

RESULTS

TRAINING RESULTS

AB Training

Both subjects failed to learn the AB relation. Their scores on the AB task remained at or around chance levels throughout this phase. Gareth received 292 AB trials (158 of which were subjected to the fading programme described earlier), and Peter received 265 AB trials (all of which involved fading). Neither subject learned to colourmatch on fading trials. At no stage did either subject overtly label the Set-A or Set-B stimuli.

Train A'X'

The A'X' relation was learned relatively quickly. Gareth and Peter needed 80 and 120 trials respectively to learn to say 'Omni' in the presence of the Y-shape and 'Delta' in the presence of the zig-zag.

AB Training (resumed)

Both subjects were returned to the AB matching task for 96 trials. They continued to label each sample shape spontaneously with the appropriate Set-X word but their scores on the AB matching task remained at chance level.

Train B'X'

Both subjects quickly learned to say 'Omni' to the green hue and 'Delta' to the red. Gareth required only 28 trials, whilst Peter required 69 trials.

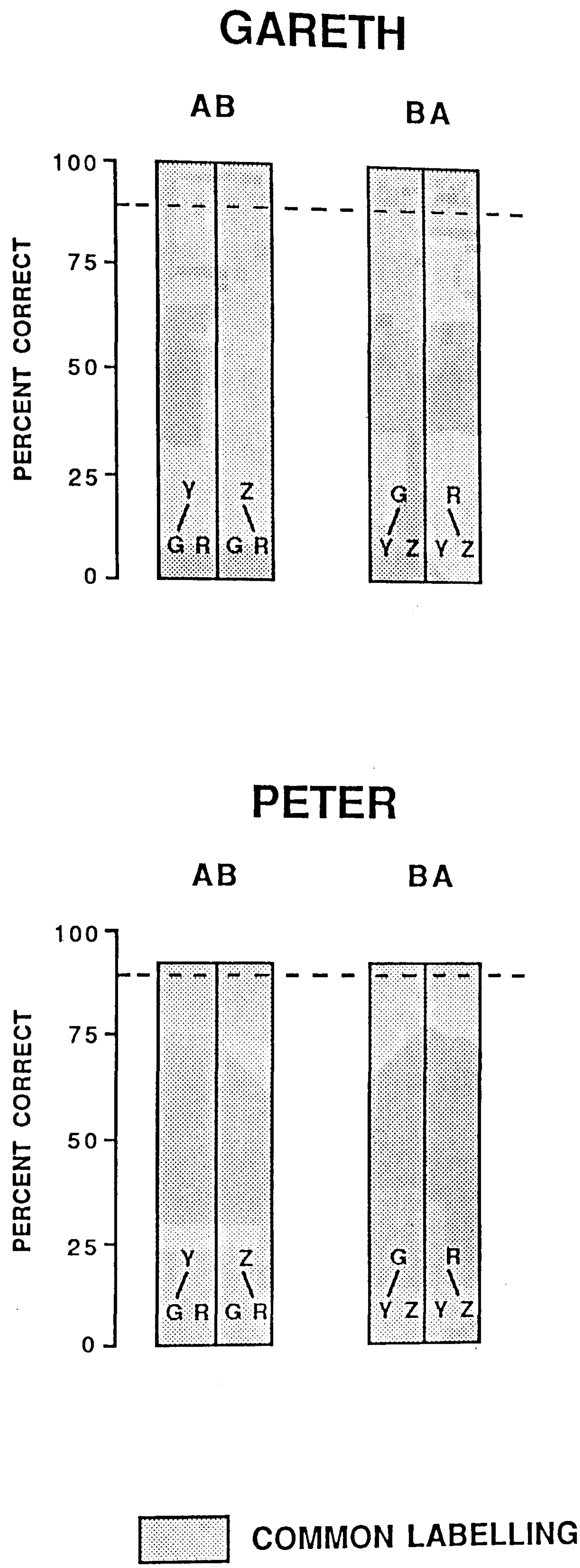
Combine A'X' and B'X'

Combining the A'X' and B'X' trials randomly within the same session did not affect either subject's labelling performance. After 36 trials, the reinforcement probability was reduced in accordance with the procedure described in the General Method section. At the end of this 48-trial session both Gareth and Peter were labelling the Y-shape and green stimuli with the common label 'Omni', and zig-zag and red with the common label 'Delta', without any differential reinforcement. Both subjects were now ready for testing. By the end of training, Gareth and Peter had received a total of 544 and 598 training trials respectively. Again, trials were staggered to produce a multiple baseline across subjects.

TEST RESULTS

Figure 4.26 depicts the two subjects' scores on AB and BA test trials. Both subjects gave common labels to the samples throughout the 96-trial test. Once again common labelling had a dramatic effect on matching performance; both AB and BA emerged straight away and at the same rate. Gareth scored 100% correct on all test trial-types, whereas Peter scored 92% correct.

Figure 4.26 Overall test scores (percentage of correct responses) on AB and BA trial-types for the two subjects in Experiment 4(b). Each bar represents 24 trials.



GARETH
AB
BA
100
75
50
25
0
PERCENT CORRECT
Y
G R
Z
G R
G
Y Z
R
Y Z
PETER
AB
BA
100
75
50
25
0
PERCENT CORRECT
Y
G R
Z
G R
G
Y Z
R
Y Z
COMMON LABELLING

DISCUSSION

The results from this experiment are entirely consistent with the interference hypothesis stated earlier. From the outset of testing, both subjects were not merely labelling the shapes and colours appropriately with the words 'Omni' and 'Delta' but, in so doing, they were also naming the shapes and the colours with those words. The words 'Omni' and 'Delta' were symmetrically related to the stimuli right from the start of testing, thus enabling both the AB and BA relations to immediately emerge. There was no evidence of any interference in equivalence formation like that found in Experiment 4(a). After the same amount of testing, most of Gareth's and Peter's counterparts in Experiment 4(a) had not yet acquired AB matching, which, as was shown earlier, indicated that unlike Gareth and Peter they were not initially naming the colours with the 'Omni' and 'Delta' words. But then, unlike Gareth and Peter, the subjects in Experiment 4(a) were already capable of naming the colours with other words, the conventional labels 'Green' and 'Red'. These conventional labels were a potential source of interference; it was possible that they initially prevented the 'Omni' and 'Delta' labels from acting as names for the colours and thus from acting as effective mediators for stimulus equivalence. This possibility is considerably strengthened by Gareth's and Peter's data; neither child was capable of consistently naming the colours prior to learning the Set-X labels. In addition, neither child spontaneously labelled the colours with any other than the Set-X words. According to the interference hypothesis, because Gareth and Peter had no other names for the colours there was nothing to interfere with or prevent their Set-X labels from acting as names for the colours, and therefore nothing to prevent AB from emerging straight away, alongside BA.

There is still plenty of scope for further confirmation of the interference hypothesis. The data so far suggests that the subjects in Experiment 4(a) did not perform AB matching in the initial stages of testing simply because the labels chosen as common mediators ('Omni' and 'Delta') were different from those the subjects normally applied to one of the stimulus sets ('Green' and 'Red'). In Experiment 4(b) interference was prevented by ensuring the subjects had no other names for the stimuli. Another way of preventing interference is to deliberately adopt as common labels those names which the subject already spontaneously applies to one of the stimulus sets. Using the same stimuli as before, one could teach the subjects to apply their pre-existing colour names to the shapes, for which they have no other names prior to the experiment. If, for example, a subject is taught to say 'Green' to the previously un-named Y-shape then one would expect no interference in the subsequent matching of green to Y and Y to green. On the contrary, this approach should be perhaps one of the most effective ways of making two stimuli become equivalent to each other. Experiment 5 assessed this possibility.

EXPERIMENT 5

Two changes were made in this experiment. Firstly, the Set-X labels were changed; 'Omni' was replaced with 'Green' and 'Delta' was replaced with 'Red'. The second change was a procedural one - this time the training involved one less step than before. In order to establish the potential for common labelling, the subjects only needed to be taught to label the Set-A shapes with the Set-X words, 'Green' and 'Red', because they had already learned, prior to the experiment, to assign those words to the Set-B colours. Consequently, all the subjects (except Donna; see below) were taken straight on to AB and BA testing immediately after learning the A'X' relation. Other than this, the procedure was identical to that of the previous experiment.

METHOD

SUBJECTS

The seven 4-5 year old children are listed in Table 4.25, which includes their chronological ages and their equivalent language ages from the Reynell language test. All subjects were given a 24-trial naming test prior to the experiment to ensure that:

(a) they could all assign the conventional labels 'Green' and 'Red' to the corresponding Set-B colours, and

(b) that they were unable to consistently label the Set-A shapes.

APPARATUS (See General Method Section)

PROCEDURE (see above)

Table 4.25 Chronological age (in years-months) and equivalent language age (from the Reynell Developmental Language Scale) for the seven subjects in Experiment 5.

SUBJECT	CHRONOLOGICAL AGE	EQUIVALENT LANGUAGE AGE	
		VERBAL COMPREHENSION	EXPRESSIVE LANGUAGE
BILLY JOE	4 - 11	4 - 6	5 - 6
DONNA	4 - 1	4 - 9	5 - 11
MELISSA	4 - 9	5 - 9	5 - 6
NICK	4 - 10	4 - 10	5 - 2
RICHARD	4 - 10	5 - 2	4 - 6
SARA L.	4 - 8	4 - 7	4 - 5
STEVE	4 - 9	4 - 8	7 - 0

RESULTS

TRAINING RESULTS

AB Training

The AB training again proved ineffective. Each subject's performance remained at or around chance level throughout. The number of AB trials each child received was as follows: Billy Joe, 96; Richard, 144; Nick, 240; Donna, 252; Steve, 312; Sara L., 360; and Melissa, 432. Billy Joe was given instructions (see Experiment 3 for details) but other than this the fading programme was the only intervention used, and this was administered to Donna, Sara L. and Steve, but without effect. Steve received 168 fading trials but never even began to colourmatch. All of Donna's AB trials involved fading; she matched the colours from the start of her first fading session. Sara L. was given 264 fading trials and she began colourmatching in her second fading session. Both Donna and Sara L. continued to match correctly up until the last fading step, at which point their responding fell to chance levels.

Train A'X'

The subjects very quickly learned to say 'Green' in the presence of the Y- shape and 'Red' in the presence of the zig-zag. All the subjects except Donna and Steve needed only 24 trials to learn the A'X' relation; Donna required 48 trials, and Steve required 72 trials before reaching criterion.

All the subjects were now capable of labelling the Y-shape and green stimuli with the common label 'Green', and zig-zag and red with the common label 'Red', and so they were ready for testing. Since testing was to be conducted in the absence of differential reinforcement, all the subjects (except Donna) were given 12 additional A'X' trials without reinforcement (the reinforcement probability was reduced to zero in accordance with the procedure described in the General Method section). None of these subjects made any errors on the 12 unreinforced A'X' trials. Meanwhile, Donna proceeded along one further training stage (see below).

AB Training (resumed)

Donna was returned to the AB matching task for two 48-trial sessions. When the Set-A samples appeared, she continued to label them with the appropriate Set-X words. Furthermore, her AB matching performance immediately rose to criterion; her overall AB score was 87.5% in the first session with labelling and 93.75% in the second. Although correct responses were reinforced during these sessions, Donna's AB score improved so rapidly that sample labelling alone may have been responsible for the improvement (see Test Results).

Prior to the last 12 trials of the second AB session, the probability of reinforcement was reduced to zero without affecting Donna's matching performance, and she therefore proceeded to the test phase, to join the other subjects.

Summary of Training

The total number of trials each subject received during training was as follows: Billy Joe, 132; Richard, 180; Nick, 276; Donna, 396; Steve, 396; Sara L., 396; and Melissa, 468.

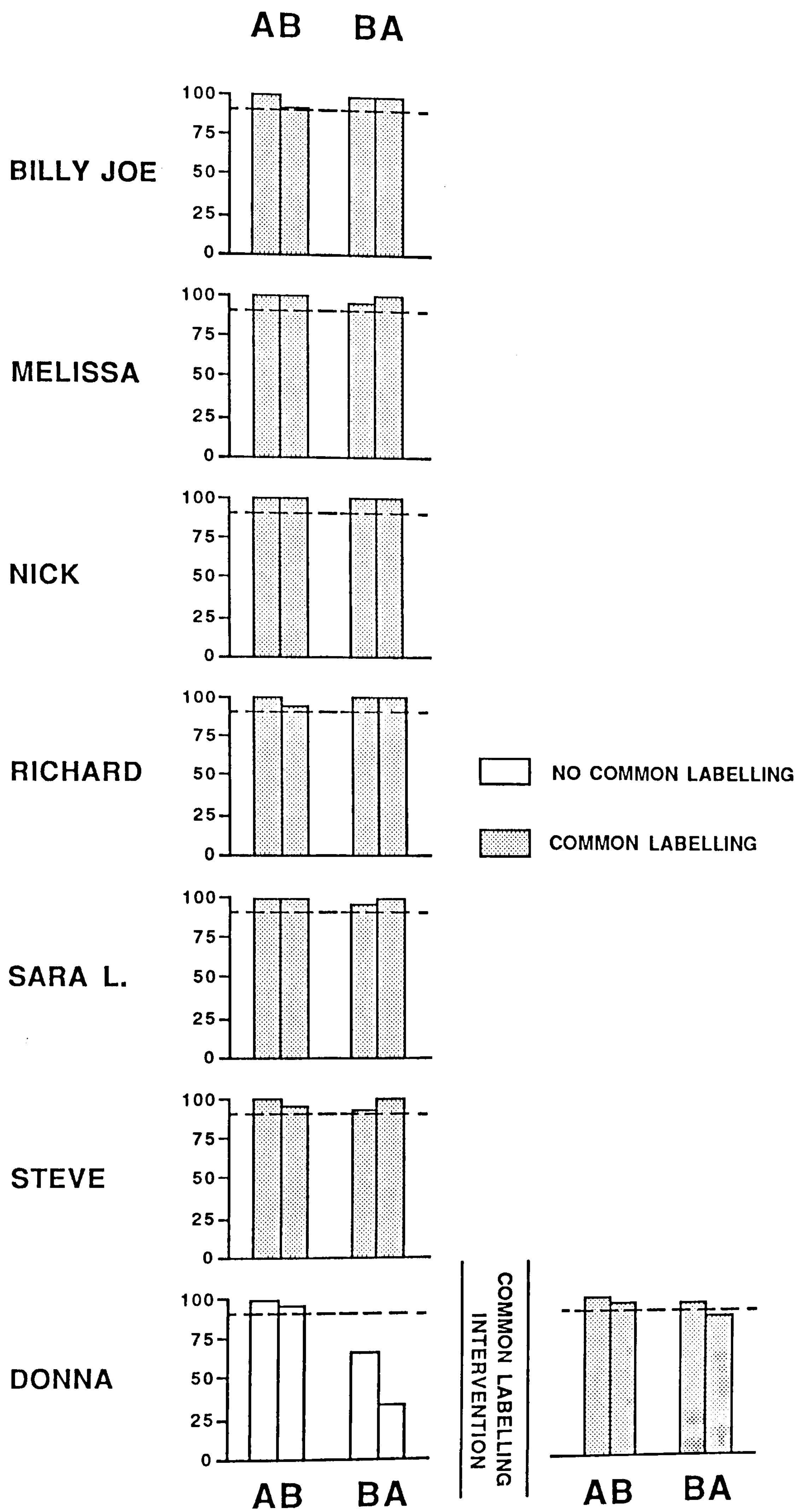
TEST RESULTS

The test results for all seven subjects are depicted in Figure 4.27. The shaded bars depict those matching-to-sample trials on which common Set-X labels ('Green' and 'Red') were correctly applied to the Set-A and Set-B samples. The figure shows that when common labelling occurred both the AB and BA relations emerged immediately for all seven subjects. Common labelling, then, was once again apparently sufficient for the formation of stimulus equivalence.

Donna, the subject at the bottom of the figure, merits particular attention. In her initial test phase she did not give common labels to the samples and she did not completely pass the tests (see the unshaded bars). In the absence of common labelling, Donna scored above criterion on AB trials and below criterion on BA. This pattern was exactly the opposite of that initially produced by many of the subjects in Experiment 4(a). Although Donna was capable of AB matching it was clear that the Set-A and Set-B stimuli had not become equivalent because she was unable to match the symmetrically related BA trials. At this stage, then, Donna had failed a symmetry test, which showed yet again that equivalence in humans is not a 'given', as Sidman has claimed (Sidman, 1988).

In order to understand why Donna had failed BA, one needs to refer once again to the labels she produced. Although common labelling was absent from Donna's initial test performance, labelling *per se* was not. Furthermore, the difference in her AB and BA performance appeared to be linked to the way in which she labelled the stimuli. On AB trials, Donna overtly labelled the sample shapes 'Green' and 'Red' in accordance with her earlier A'X' training. Then, when the green and red comparisons subsequently appeared, she selected the comparison

Figure 4.27 Overall test scores (percentage of correct responses) on AB and BA trial-types for the seven subjects in Experiment 5. Each bar represents 24 trials.


AB
BA
BILLY JOE
MELISSA
NICK
RICHARD
SARA L.
STEVE
DONNA
100
75
50
25
0
NO COMMON LABELLING
COMMON LABELLING
COMMON LABELLING INTERVENTION
AB
BA
AB
BA

which corresponded with the colour label she had just given to the sample shape. Donna's AB matching was apparently mediated by her Set-X labels, 'Green' and 'Red'. This labelling appeared to be sufficient for establishing the very relation which she had earlier failed to acquire through reinforced training trials, and in the apparent absence of any such labelling.

However, this labelling, which seemed to form a vital link in the mediated emergence of AB, was apparently absent on BA trials. On BA trials, Donna never once overtly labelled the Set-B colour samples. It was therefore possible that Donna had failed BA simply because she had not labelled the colour samples either overtly or covertly with the Set-X words 'Green' and 'Red', despite being capable of so doing. Additional testing confirmed this possibility. In subsequent tests, Donna was prompted to label the samples on BA trials; when the green or red sample appeared she was instructed to 'Say what it is'. This prompting was successful; Donna immediately began to label the colours appropriately with the words 'Green' and 'Red'. Since she was already applying the same labels to the Set-A sample shapes on AB trials, the prompting had actually established common labelling of the samples, and thus is represented in Figure 4.27 by the column entitled 'common labelling intervention'. Once common labelling was established, the BA relation emerged to join AB, confirming for the first time that the Set-A shapes and Set-B colours had become fully equivalent.

DISCUSSION

Prior to this experiment, the subjects were already capable of labelling the Set-B colours appropriately with the words 'Green' and 'Red'. These subjects were then taught to apply the same labels to corresponding Set-A shapes. This tactic proved extremely effective in promoting stimulus equivalence; as predicted earlier, there was no evidence of any interference in equivalence formation. Each shape and its corresponding colour became equivalent via mediation of the common colour labels, provided those labels were produced by the subject during equivalence testing. The immediate emergence of both AB and BA conditional upon production of the common labels 'Green' and 'Red', showed that those labels were also acting as names for the stimuli (i.e. that each label and its corresponding stimulus had become symmetrically related).

These results have obvious implications for anyone faced with the practical problem of establishing classes of equivalent stimuli within a subject's repetoire. Much time and effort could be saved by first identifying whether the subject can already name any of the stimuli in question. If the subject only has names for one of the stimulus sets then, provided they are 'acceptable' to the language community, those names could be incorporated into the experimental regime as common mediators for stimulus equivalence. The advantages would be two-fold. Firstly, less training would be required because the subject had already learned one of the component naming relations. Secondly, a positive outcome would be more likely because, when used in this way, the subject's pre-existing names for the stimuli cannot become a source of interference in equivalence formation. Compare this to what might happen if the teacher is ignorant of the subject's pre-existing stimulus names. If other labels are chosen as potential mediators for equivalence then they may be less than fully effective in their intended roles. The subject's pre-existing stimulus names may interfere with the process by which the chosen labels become common names for the stimuli. Furthermore, the problem may be more severe than the results of Experiment 4(a) suggest. In Experiment 4(a), the interference was manifested as an initial absence of the AB relation, but in two cases (i.e. with Linda and Jessica) the AB relation eventually emerged after a short period of uninterrupted testing. However, Experiment 4(a) involved only two sets of stimuli. If one is attempting to instill equivalence amongst a larger network (as would surely be the case in an applied setting) then any interference may be correspondingly magnified.

So far, interference has been avoided by making two changes to Experiment 4(a). Experiment 4(b) involved a change in the subject variables of Experiment 4(a). Unlike their counterparts in 4(a), the 4(b) subjects had no names for the stimuli prior to the experiment. Experiment 5, however, involved a change in the response variables of Experiment 4(a) (see above). For the sake of completeness, the final experiment of this series involves a change in the stimulus variables of Experiment 4(a). The Set-B colours are replaced with a set of shapes for which the subjects apparently have no names. Given that the subjects have no names for either the Set-A or the Set-B shapes prior to learning the common labels 'Omni' and 'Delta', then there should be nothing to prevent those labels from becoming common names for the stimuli; consequently, equivalence should emerge immediately, without any interference.

EXPERIMENT 6

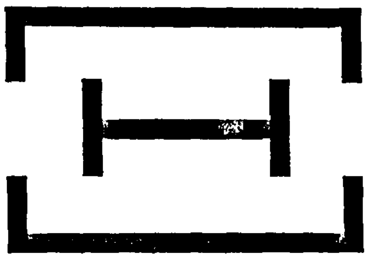
In this experiment, two new shapes were introduced as Set-B stimuli (see Figure 4.28). These shapes, hereafter referred to as Shape 1 and Shape 2, were deliberately designed to be obscure, so that the children would probably not have any consistent names for them. The procedure in this experiment was identical to the main route procedure of Experiment 4(a), except that Shape 1 replaced the green stimulus and Shape 2 replaced the red.

METHOD

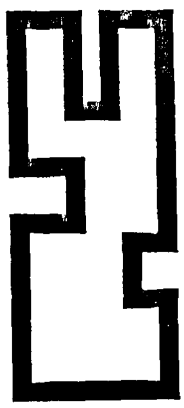
SUBJECTS

Five 4-5 year old children took part. Their chronological ages are listed in Table 4.26, along with their equivalent language ages from the Reynell language test. Each subject was given a 24-trial naming test prior to the experiment. This consisted of four trials each of the Y-shape, zig-zag, shape 1, shape 2, and the green and red hues from the previous experiments. The two hues were included as a control measure, to ensure that the children understood the questions posed by the experimenter. Although each child consistently assigned the conventional labels 'Green' and 'Red' to the corresponding colours, none of them consistently named the shapes. In fact, the subjects' reactions to the shapes fell into two broad categories; they either attempted to describe each shape in general terms or they remained silent. Further prompting only elicited negative responses such as 'Don't know'.

Figure 4.28 Shapes employed as Set-B stimuli in Experiment 6.



SHAPE 1



SHAPE 2

Table 4.26 Chronological age (in years-months) and equivalent language age (from the Reynell Developmental Language Scale) for the five subjects in Experiment 6.

SUBJECT	CHRONOLOGICAL AGE	EQUIVALENT LANGUAGE AGE	
		VERBAL COMPREHENSION	EXPRESSIVE LANGUAGE
CHERYL	4 - 3	4 - 10	4 - 8
LAURA	4 - 3	4 - 6	7 - 0
IOAN	4 - 7	4 - 10	6 - 5
SANDRA	4 - 4	5 - 1	4 - 0
WILLIAM	4 - 11	5 - 2	6 - 2

APPARATUS (see General Method)

PROCEDURE (see above)

RESULTS

TRAINING RESULTS

AB Training

On AB trials, reinforcers were available for matching the Y-shape sample to shape 1, and the zig-zag sample to shape 2. Four of the subjects, Cheryl, William, Laura and Ioan failed to learn this task. After 96, 150, 192 and 353 trials respectively, their scores were still at or around chance level. Furthermore, these children never overtly labelled the Set-A or Set-B stimuli. Neither did Sandra, but she reached criterion on AB after only two 48-trial sessions (her overall score in the second session was 91.7% correct). To prepare her for testing, the reinforcement probability was reduced to zero (see General Method). Sandra made no errors on 12 unreinforced AB trials, and therefore proceeded to testing. Meanwhile, the others were taken to the next stage.

Train A'X'

All the subjects (except Sandra) learned to label the Y-shape with the word 'Omni' and zig-zag with the word 'Delta'. No subject took more than one 48-trial session to reach the criterion of six consecutive correct responses per trial-type; Ioan required 36 trials, William 38; Laura, 42; and Cheryl, 44.

AB Training (resumed)

AB training was resumed for two 48-trial sessions but each subject's matching performance remained at or around chance level. All four subjects, however, continued to label the Y-shape and zig-zag samples with the words 'Omni' and 'Delta' respectively.

Train B'X'

Next, the four subjects were taught to say 'Omni' in the presence of shape 1 and 'Delta' in the presence of shape 2. Learning progressed extremely rapidly. William reached criterion after only 18 trials, whereas Cheryl required 26, Laura 33 and Ioan 48.

Combine A'X' and B'X'

When the A'X' and B'X' trial-types were presented randomly within the same session, Ioan, Laura and William continued to respond correctly, each making no errors on the 12 A'X' and 12 B'X' trials. In contrast, Cheryl began to label the shapes incorrectly. In her first combined session, and despite making no errors on B'X', her labelling performance fell to 66.7% correct for the Y-shape and 83.3% correct for the zig-zag. The following session consisted of A'X' trials only. Removing the B'X' trials apparently helped because Cheryl required only 34 trials to return to criterion on A'X'. Then, when A'X' and B'X' were once again combined she scored 100% correct on both.

By the end of this stage, all four children had learned to apply common Set-X labels to the Set-A and Set-B shapes. They were each able to say 'Omni' to both the Y-shape and shape 1, and 'Delta' to both the zig-zag and shape 2. When the probability of reinforcement was reduced to zero, each subject continued to label the shapes correctly (12 A'X' and 12 B'X' trials were presented without reinforcement). All four were ready for testing.

Summary

The total number of training trials for each subject was as follows: Sandra, 108; William, 360; Cheryl, 396; Laura, 417; and Ioan, 593.

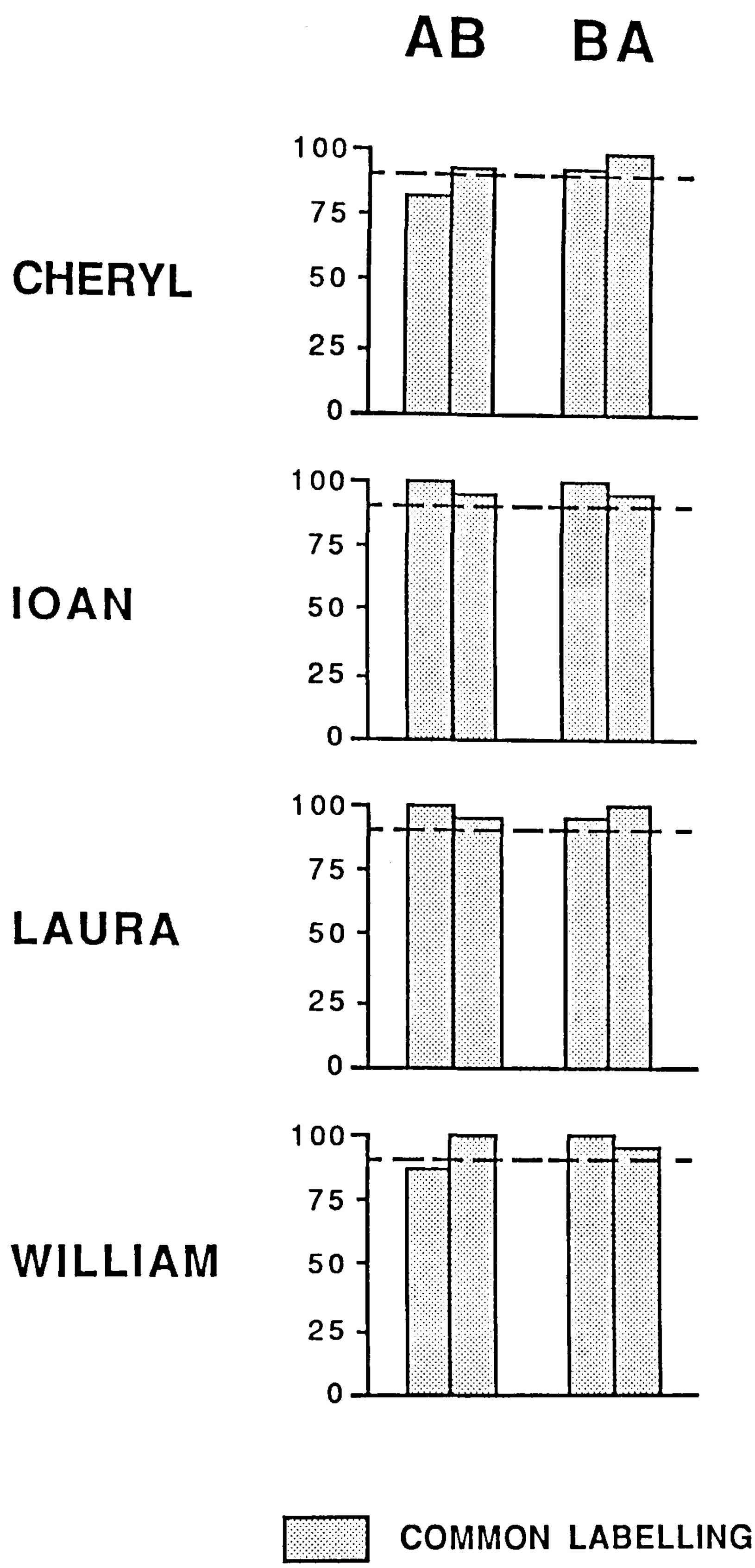
TEST RESULTS

Figure 4.29 shows the overall scores on AB and BA test trials for the four subjects who had learned common labels. Each bar represents 24 trials. All four children gave common Set-X labels to the samples throughout testing (as depicted by the shading), and both AB and BA emerged virtually straight away.

Sandra had learned AB prior to the introduction of label training procedures. Her symmetry test performance is depicted in Figure 4.30. Sandra failed the symmetry test during her first phase of testing (see the left-hand set of bars in Figure 4.30). Although her AB performance was excellent, she selected the zig-zag comparison on all but two of the BA symmetry test trials. The Set-B samples did not appear to be exerting any control over her choice of Set-A comparison. Furthermore, Sandra remained silent throughout this first test phase; she never once overtly labelled any of the stimuli.

Further testing was preceded by the introduction of A'X' training trials. Sandra required only 68 trials to learn to say 'Omni' to the Y-shape and ' Delta' to the zig-zag. However, this training alone did

Figure 4.29 Overall test scores (percentage of correct responses) on AB and BA trial-types for the four subjects in Experiment 6 who failed to learn AB matching prior to testing. Each bar represents 24 trials.


AB
BA
CHERYL
IOAN
LAURA
WILLIAM
100
75
50
25
0
COMMON LABELLING

not affect her subsequent symmetry test performance. When re-tested, Sandra made no errors on AB baseline trials, but her performance on BA remained the same as before i.e. she continued to select the zig-zag comparison on all but a few trials, as can be seen from the middle group of bars in Figure 4.30. During this second test phase, Sandra continued to overtly label the Set-A shapes correctly as 'Omni' or 'Delta', but only when those shapes appeared as samples on AB trials. Overt labelling of the Set-B shapes was completely absent.

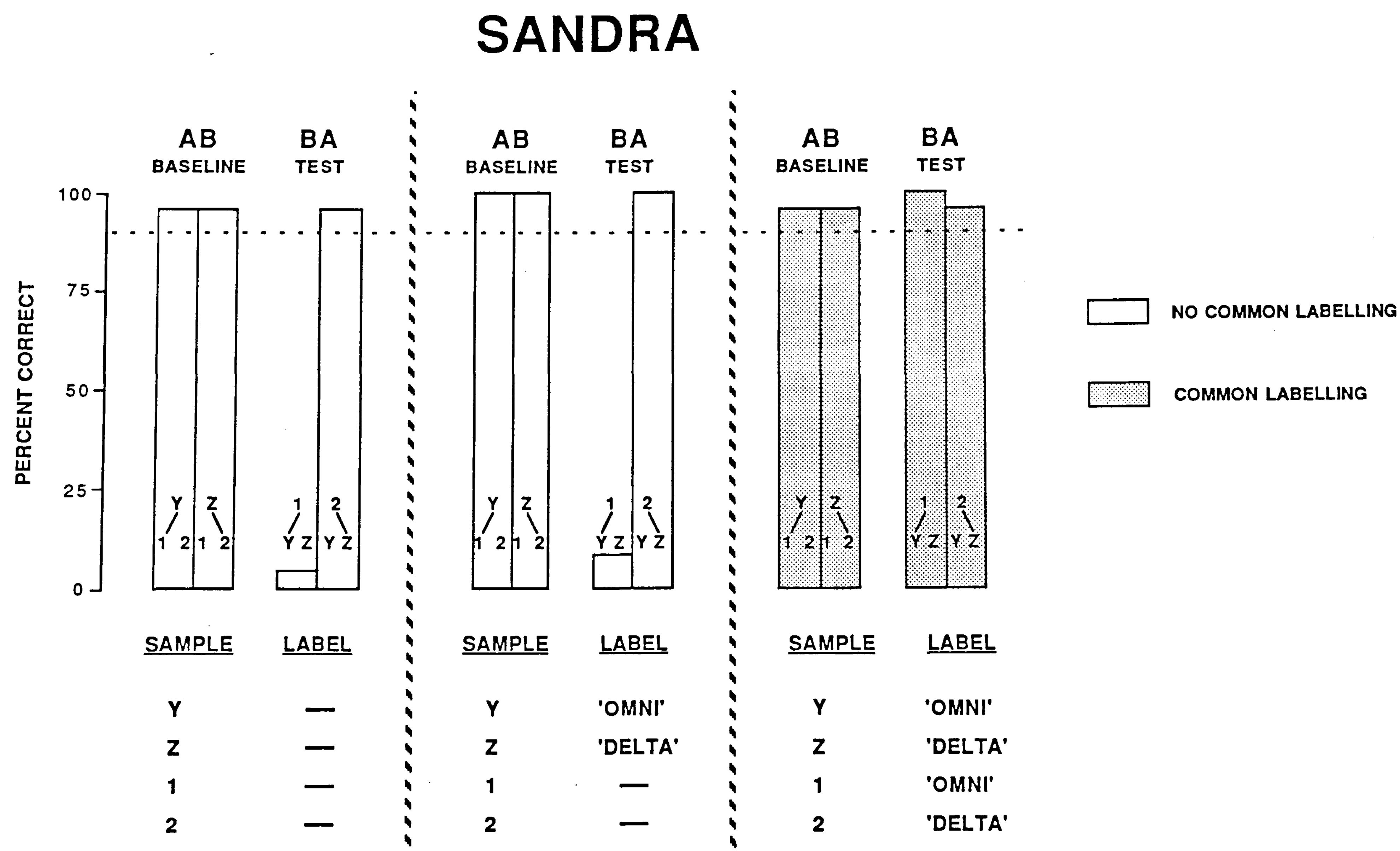
Testing was interrupted once more so that further training could establish common labelling. Sandra needed the absolute minimum of 12 trials to learn the B'X' relation. One more 48-trial session followed, in which the A'X' and B'X' trial-types were combined and the reinforcement probability was reduced to zero. At the end of this session, Sandra was labelling both the Y-shape and shape 1 as 'Omni' and the zig-zag and shape 2 as 'Delta', in the complete absence of differential reinforcement.

In her final test session (depicted by the shaded bars in Figure 4.30), Sandra continued to label the Set-A and Set-B samples overtly with the corresponding Set-X labels. Furthermore, this common labelling appeared to result in the immediate emergence of BA; only one error was made in 48 unreinforced BA test trials.

DISCUSSION

The results from this experiment are in complete agreement with the theoretical notions derived from the preceding experiments of this chapter. Once again, stimulus equivalence emerged when the subjects applied common labels to the stimuli during matching-to-sample trials. As anticipated, there was no interference in equivalence formation.

Figure 4.30 Sandra's test performance on AB baseline and BA test trials during her three test phases (see text). Each bar represents 24 trials. The stimuli for each trial-type are placed at the bottom of the bars (Y=Y-shape, Z=zig-zag, 1=shape1, 2=shape2). The table below each set of bars depicts Sandra's overt sample labels from that phase of testing.


SANDRA
AB BASELINE
BA TEST
AB BASELINE
BA TEST
AB BASELINE
BA TEST
PERCENT CORRECT
100
75
50
25
0
SAMPLE
LABEL
Y —
Z —
1 —
2 —
SAMPLE
LABEL
Y 'OMNI'
Z 'DELTA'
1 —
2 —
SAMPLE
LABEL
Y 'OMNI'
Z 'DELTA'
1 'OMNI'
2 'DELTA'
NO COMMON LABELLING
COMMON LABELLING

The subjects apparently had no pre-existing names for either the Set-A or Set-B shape stimuli so, in later training, each shape was free to become symmetrically related to its corresponding Set-X label. The Set-X words thus acted as common names for the stimuli, and when common naming was subsequently produced during testing, the relations indicative of stimulus equivalence (i.e. AB and BA) emerged. Sandra's results revealed yet again that equivalence does not automatically follow from the direct training of a relation via conditional discrimination procedures. These training procedures are not necessarily effective but the evidence above indicates that they can be, provided the subject names the stimuli (as defined), and in a manner which allows those names to mediate the subject's test-trial matching-to-sample performance.

GENERAL DISCUSSION

Perhaps it would be best to begin by retracing the steps along which the experiments in this chapter have taken us. Experiment 3 indicated that, as a potential mediator of stimulus equivalence, common naming might prove worthy of further investigation. Experiment 4(a) not only confirmed the expectations of Experiment 3, but also suggested that naming was itself a kind of stimulus-response symmetry. Experiment 4(a) also indicated that our understanding of equivalence may be incomplete should we fail to acknowledge the subjects' pre-existing names for the stimuli. In Experiment 4(a), these pre-existing names appeared to be potentially incompatible with the mediating function of the Set-X labels. This was confirmed in Experiment 4(b), which directly replicated 4(a) but prevented the incompatibility by using subjects who had no pre-existing names for the stimuli. Then, in Experiment 5, common labelling was established with, rather than in opposition to, the subjects' pre-existing stimulus names. So, rather than being a potential source of interference in the mediated emergence of stimulus equivalence, the subjects' pre-existing names played a key role in the emergent process. Finally, the experiments culminated in Experiment 6, which again demonstrated the role of common naming in stimulus equivalence, but this time by substituting the 'pre-nameable' stimuli from the earlier experiments with others for which the subjects had no pre-existing names.

There are various ways in which common naming could be investigated further within the framework established in this chapter. One virtue of the present theoretical framework is that it enables the prediction of very specific effects, and thus it is open to empirical validation. For example, one could replicate Experiment 4(a) but with the Set-A and Set-B stimuli interchanged. In Experiment 4(a), the BA relation often emerged prior to AB and never vice versa. However, if the Set-A and Set-B stimuli are interchanged then AB should emerge prior to BA and never vice versa (i.e. reversing the stimulus sets should reverse the relative rates of emergence). Alternatively, one could find subjects who have no consistent names for the Set-A or Set-B stimuli and then teach them to name the stimuli from one of the sets before introducing other labels as potential common mediators of stimulus equivalence. This should also result in the differential emergence of AB and BA, because the mediating labels would be potentially incompatible with the stimulus names the subjects had previously learned. This time, though, the latter would be established within rather than outside the experimental context. If the subject is taught to name the Set-A stimuli prior to learning common labels with different words, then AB should emerge before but never after BA (and vice versa if the subject is initially taught to name the Set-B stimuli).

Perhaps other experiments could replicate the studies in this chapter but with younger children. It remains to be seen whether the same common labelling procedure will produce common naming as readily with children under 4 years as it did with the 4-5 year olds.

In conclusion, the experiments in this chapter appear to paint a consistent picture of the effects of naming in general and of common naming in particular on the formation of stimulus equivalence. But the picture is far from complete; only a small part of the 'canvas' has been filled. Hopefully, further experiments will follow; if so, these will undoubtedly modify and add to the current perspective. But, in the meantime, the final part of this thesis will attempt to fill in a fraction more of the canvas by discussing general points arising from what has already been portrayed.

CHAPTER 5

CONCLUSIONS

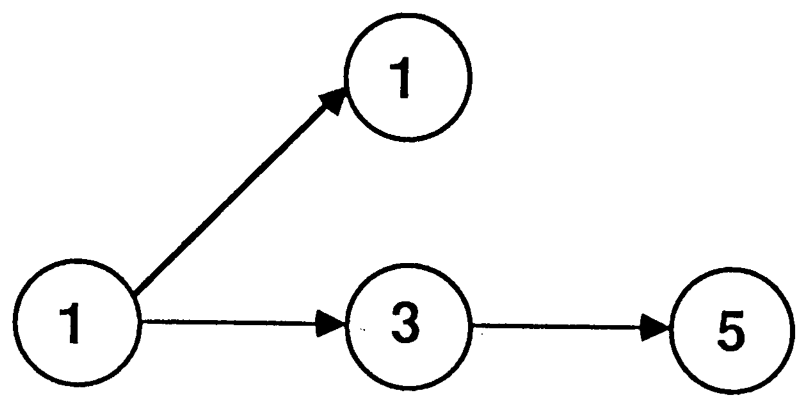
CONCLUSIONS

The findings from previous equivalence experiments (Chapter 1), together with the data yielded by the current research programme (Chapters 2, 3 and 4), represent a seemingly formidable amount of evidence for the critical role of naming in the emergence of equivalence. There are, however, always exceptions to the rule. Currently there are two experiments which appear to have demonstrated equivalence in animals. The question is: what status should we afford to these aberrant studies? Each presents the outward appearance of equivalence, but looks can be deceiving. We must be prepared to probe beyond the surface to discover their true nature.

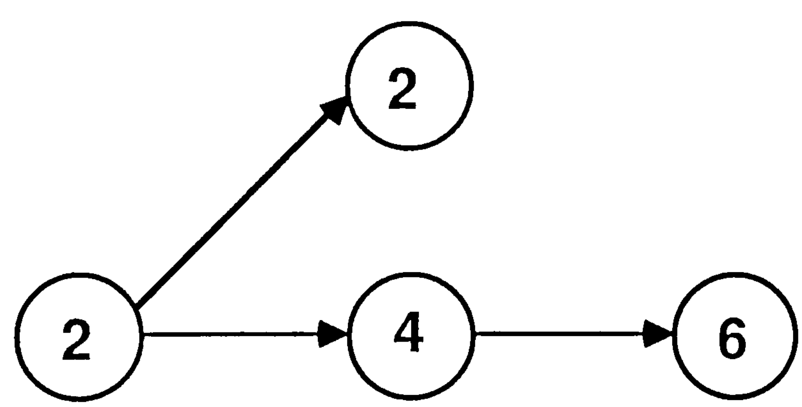
The first experiment to go 'under the microscope' does, in a sense, support the definition of naming proposed in this thesis, but not without producing some seriously misleading side effects. In a recent study, McIntire, Cleary and Thompson (1987) began by acknowledging the role of naming in equivalence formation in humans and then proceeded to explain procedures for teaching animals (in this case two cynomolgous monkeys) skills analagous to common naming. Figure 5.1 is a schematic representation of their paradigm. They aimed to establish two classes of equivalent stimuli. One class, designated EVEN, consisted of stimulus numbers 2, 4 and 6 (corresponding to violet, green and orange), while the other class, designated ODD, consisted of stimulus numbers 1, 3 and 5 (corresponding to red, yellow and blue). The monkeys were trained to criterion on the relations depicted by the arrows in Figure 5.1, after which tests evaluated the emergence of all the other sample-comparison combinations. What made this experiment different from others on equivalence in animals was that the subjects were required to respond differentially to the two

Figure 5.1 Schematic representation of the stimulus relations established in the experiment by McIntire, Cleary and Thompson (1987) (see text).

ODD



EVEN



sets of stimuli whilst performing the matching tasks. The monkeys were required to press each EVEN stimulus eight times and to press and continuously hold down each ODD stimulus for at least 3.5 seconds. Each trial began with the sample; if the subject 'named' it correctly (i.e. produced the EVEN response to even numbered stimuli and the ODD response to odd numbered stimuli) then the comparisons appeared. The subjects could then secure reinforcers by selecting the correct comparison and by producing the appropriate ODD or EVEN response. If the subjects produced the incorrect response pattern at any stage during the trial then the stimuli disappeared, no reinforcers were delivered, and the next trial began approximately 4 seconds later.

After learning the tasks depicted by the arrows in Figure 5.1, both monkeys were able to match all other combinations without differential reinforcement. The authors concluded that the monkeys had formed two classes of equivalent stimuli through learning and using a 'simple two-word naming system' (p.281).

But there are two problems with this conclusion. The first concerns the extent to which these data constitute satisfactory evidence for stimulus equivalence. There is little doubt that the experiment looks like it demonstrated equivalence, but that is not the issue. Equivalence is defined by functional, and not formal, properties (cf. Chapter 1 of this thesis). Equivalence requires the emergence of untrained relations, but when one closely examines the McIntire et al study, one discovers that nothing has emerged during testing. Each subject's test performance merely reflects an elaborate stimulus-response chain, in which the relations both within and between each successive link of the chain were already highly trained through differential reinforcement.

Before going into detail, some extra information is required.

Several studies have shown that when, for example, pigeons are trained to respond differentially to samples presented during matching trials, the differential responses, as well as the sample stimuli, may readily exert control over the subjects' comparison choices (Urcuioli and Honig, 1980; Urcuioli, 1985). Anyone armed with this knowledge can easily re-interpret McIntire et al's results. Their monkeys' training could result in the learning of two relations per stimulus (to simplify matters we will concentrate on only one of the stimulus sets, say, the EVENS). For example, during 2 - 2 training, the subject learns the chain 2 - E - 2 - E (where this represents sample 2 - even response - comparison 2 - even response). This may give rise to two relations, 2 - E and E - 2, because, during training, the differential E response may gain control over the subject's choice of the comparison, 2. Similarly, the E - 4 and 4 - E relations may be learned during 2 - 4 training (2 - E - 4 - E producing E - 4 and 4 - E). Finally, the E - 6 and 6 - E relations may arise from 4 - 6 training (4 - E - 6 - E producing E - 6 and 6 - E). So, after training, each even numbered stimulus controls an EVEN response, and the EVEN response controls each subject's choice of any even numbered stimulus. The same analysis may be applied to the odd numbered stimuli. One would now expect the subjects to match any EVEN sample to any EVEN comparison (and likewise for the ODDS), not because of the emergence of equivalence, but merely because all the necessary component skills were explicitly taught by reinforcement contingencies present throughout training.

The second problem concerns the authors' implication that the monkeys were naming the stimuli. This implication seems dangerously misleading. The data produced from experiments in the present thesis are consistent with the view that naming, like equivalence, is functionally, and not formally, defined. According to the present thesis, we may only speak of naming with respect to stimulus-response relations which are themselves bidirectionally or symmetrically related. Naming, too, involves emergent behaviour. Now, the children in Experiments 4(a) - 6 of the present thesis were naming the stimuli; because after they learned to produce a response conditional upon a stimulus, they all proved able to do the reverse (i.e. choose the stimulus conditional upon the response) without explicit training or reinforcement for doing so. The monkeys, however, were trained on both components in question, and so there is no compelling reason to credit them with naming.

To recap, the above study demonstrates neither equivalence nor naming in animals. The study is, however, not without utility inasmuch as it provides an insight into the kinds of skills that normally need to emerge so as to bring about stimulus equivalence. 'Simulation' studies such as this will continue to play a useful role in the experimental analysis of behaviour, but only to the extent that the dangers of the formalistic fallacy are avoided (cf Savage-Rumbaugh and Rumbaugh, 1980).

The McIntire et al (1987) experiment is not, however, the only equivalence study of its kind; an experiment by Edwards, Jagielo, Zentall and Hogan (1982) on acquired equivalence in pigeons makes the same kind of error. This time the procedure was quite different, but the results were equally misleading. The study was interested in the effects of differential reinforcers rather than differential responses. The pigeons were trained on identity matching with two sets of stimuli, two forms (a plus and a circle), and two colours (red and green). Correct choices on identity trials with the plus and red stimuli were reinforced with peas, whereas correct choices on circle and green identity trials were reinforced with wheat. After learning identity matching, the birds were tested on their ability to match non-identical stimuli previously associated with common reinforcers (i.e. match red to plus, and plus to red (peas); and match green to circle, and circle to green (wheat)). The birds learned these arbitrary tasks faster than other birds in each of two control groups, one receiving pea and wheat reinforcers in an uncorrelated fashion, and the other receiving reinforcers comprising of an equal mix of peas and wheat. The effect was essentially replicated in an additional experiment. The authors concluded that the experimental birds had formed equivalences between hues and forms based upon mediation by common food 'expectancies'.

However, a different analysis can be given, based not upon unspecified differential expectancies but upon <u>differential responses</u> induced by <u>classical conditioning</u>. Several studies have shown that differential reinforcers can act as unconditioned stimuli (UCS's) for eliciting differential responses (Brodigan and Peterson, 1976; Jenkins and Moore, 1973). For example, when Jenkins and Moore (1973) presented water as a reinforcer during autoshaping trials, pigeons' key pecks came to resemble the form of pecking normally produced when they drank water. However, when grain was presented as a reinforcer, the birds' key pecks came to resemble the natural pecking elicited by grain. Similarly, in the Edwards et al study, peas may act as a UCS for eliciting 'pea pecking' responses, and wheat may act as a UCS for 'wheat pecking'. During identity matching trials, peas were delivered whenever the birds chose either the plus or the red comparison. In other words, the plus or red reliably preceded presentation of the peas. These pairings would, through the usual process of classical conditioning, eventually establish the plus and the red as conditional stimuli (CS's) for 'pea responding' (in the interests of curtailing embarrassment and preventing confusion the author will avoid using the more vulgar term for this particular differential response). The elicited 'pea responses' could then interact with the identity matching procedure (i.e. intervene between sample presses and comparison choices), thus creating the necessary conditions for mediated matching. For example, during plus-plus identity trials, the 'pea response' elicited by the plus sample may gain control over the pigeon's choice of the plus comparison. The red stimulus may become similarly related with 'pea responses' during red-red identity trials. Consequently, both plus and red may elicit 'pea responses' and 'pea responses' may control the bird's choice of either plus or red. The pigeons would then be able to match plus to red and red to plus, even without the emergence of equivalence, because each would have learned the necessary mediating relations through their exposure to the prevailing reinforcement contingencies. The same analysis can be applied to the 'wheat' stimuli. The procedures may appear quite different to those adopted by McIntire et al (1987), but the outcome seems the same. If the birds were taught what to do on the test trials, then there is no need to invoke stimulus equivalence.

This interpretation is consistent with a body of data supporting a classical mediation model of 'expectancy' (see Peterson, Wheeler and Armstrong (1978) for a review) and it also has the virtue of being open to empirical test. What would happen, for example, if stimuli were paired with their differential reinforcers via a simple autoshaping procedure? This should still result in the formation of differential 'expectancies' (i.e. differential responses controlled by the stimuli) so according to Edwards et al, the birds should still form equivalences between non-identical stimulus pairs. According to the present account, however, there should be no such evidence because the procedures would not allow differential responses to gain control over the birds' comparison choices.

Experiments 1 and 2 in the present thesis add considerably to the current weight of evidence for human-animal differences in equivalence formation. Experiment 1 was particularly significant in providing no evidence for equivalence in the matching-to-sample performance of the two language-trained chimpanzees, Sherman and Lana. The data from Experiment 1 leaves one wondering: if these chimps are linguistically accomplished then why did they fail a standard symmetry test? There seems little doubt that these chimps, for all their language training, did not satisfy standard criteria for stimulus equivalence. However, just because they did not it does not mean that they cannot. Some of the children in Chapter 4 also failed standard tests for equivalence (but always, one should note, in the absence of reinforcement). But these children then went on to confirm that they were as linguistically accomplished as one might expect. These children later proved capable of equivalence after having applied common labels to each prospective corresponding stimulus. Their subsequent success in turn confirmed that their verbal behaviour was indeed symbolic; that, as has been argued, in giving common labels the children were actually naming the stimuli as defined.

We must ask the questions 'Can the chimps name in the sense described earlier?' 'Are the chimps' lexigram responses functionally equivalent to childrens' naming?' Perhaps we may begin to answer these questions by capitalising on the potential of the common labelling paradigm as a diagnostic indicator of true naming. If the chimps still fail equivalence tests after learning to apply common lexigrams to the stimuli on matching-to-sample trials then perhaps we may suspect that their training had not established the skill of naming, at least not according to the sense of the term adopted here.

But at some point we must ask: to what extent does the present account of naming and stimulus equivalence help to make sense of other equivalence data produced elsewhere? Any theoretical account 'worth its salt' should be capable of enlisting external support and of throwing light on data which has hitherto escaped prediction and eluded adequate explanation.

In a review of various transfer procedures, Spradlin and VanBiervliet (1980) drew attention to two unpublished studies which produced <u>incomplete equivalence</u> of the kind found in Experiment 4(a) of the present thesis. The first study was by Friedman (1974). Three of the 4-5 year old children in Friedman's study passed equivalence tests involving printed numerals, printed number words and spoken number words. The fourth child, however, did not completely pass the tests. Proir to testing, the child could label numerals (AX), select numerals conditional upon spoken word samples (XA) and select printed number words conditional upon spoken word samples (XB). Test were given for the relations AB (matching numerals to number words), BA (matching number words to numerals) and BX (labelling number words). The child was able to perform AB and BX, but <u>not BA</u>.

Neither Friedman (1974) nor Spradlin and VanViervliet (1980) could understand why AB emerged but BA did not. But now, in the light of the present thesis, we may give a straightforward explanation; perhaps BA did not emerge because the subject failed to <u>spontaneously</u> name the Set-B samples during BA matching. The subject, according to the present definition, was capable of naming the B-stimuli, because BX emerged after XB training. However, BX only emerged in the highly contrived context of a naming test. If the subject had been prompted to name the B-stimuli during BA matching trials, then perhaps BA would have emerged. This possibility, however, was overlooked by Friedman, despite the fact that he noted the following with respect to two of the subjects who passed:

> "Both of these subjects did something that the other two subjects did not do. Both subjects named the sample stimulus. This type of verbal response may have played an important role for both subjects. The subjects were never reinforced for their verbalization nor were they told not to do it when it did occur. Pace (1970) reported that subjects made fewer errors when they were required to name the sample stimulus than when they were not required to name the sample" (Friedman, 1974, p.42).

(N.B. - the Pace study was not available at the time of writing).

At least Friedman reported his subjects' spontaneous verbalisations. One wonders how much has gone unreported elsewhere.

Spradlin and VanBiervliet (1980) also reported a study by James Halle and Spradlin (no reference supplied) which used Friedman's procedures with two severely retarded adolescents. One adolescent passed all three tests (AB, BA and BX) but the other failed all but the AB test. Once again, Spradlin and VanBiervliet had no idea why AB emerged but BA did not. This subject may have failed the BA test because of being <u>unable</u> to name the B-stimuli at all. The subject had learned to choose printed number words conditional upon spoken number words (XB) but he seemed unable to do the reverse i.e. label the printed number words (BX). Moreover, this result again confirms that learning to select stimuli conditional upon hearing a common sound is not sufficient for the emergence of equivalence (cf. Jessica, Experiment 4(a)). It seems that the subjects must produce the sounds themselves during equivalence tests.

Chapter 4 suggested that the subjects' pre-existing stimulus names may interfere with equivalence formation via other, less familiar, labels. One study appears to provide data consistent with this notion, and with the view that naming is necessary for equivalence. It is somewhat ironical, then, that the experiment was conducted by Sidman, Kirk and Willson-Morris (1985) who, as we saw in Chapter 1, vehemently deny that naming is essential for equivalence.

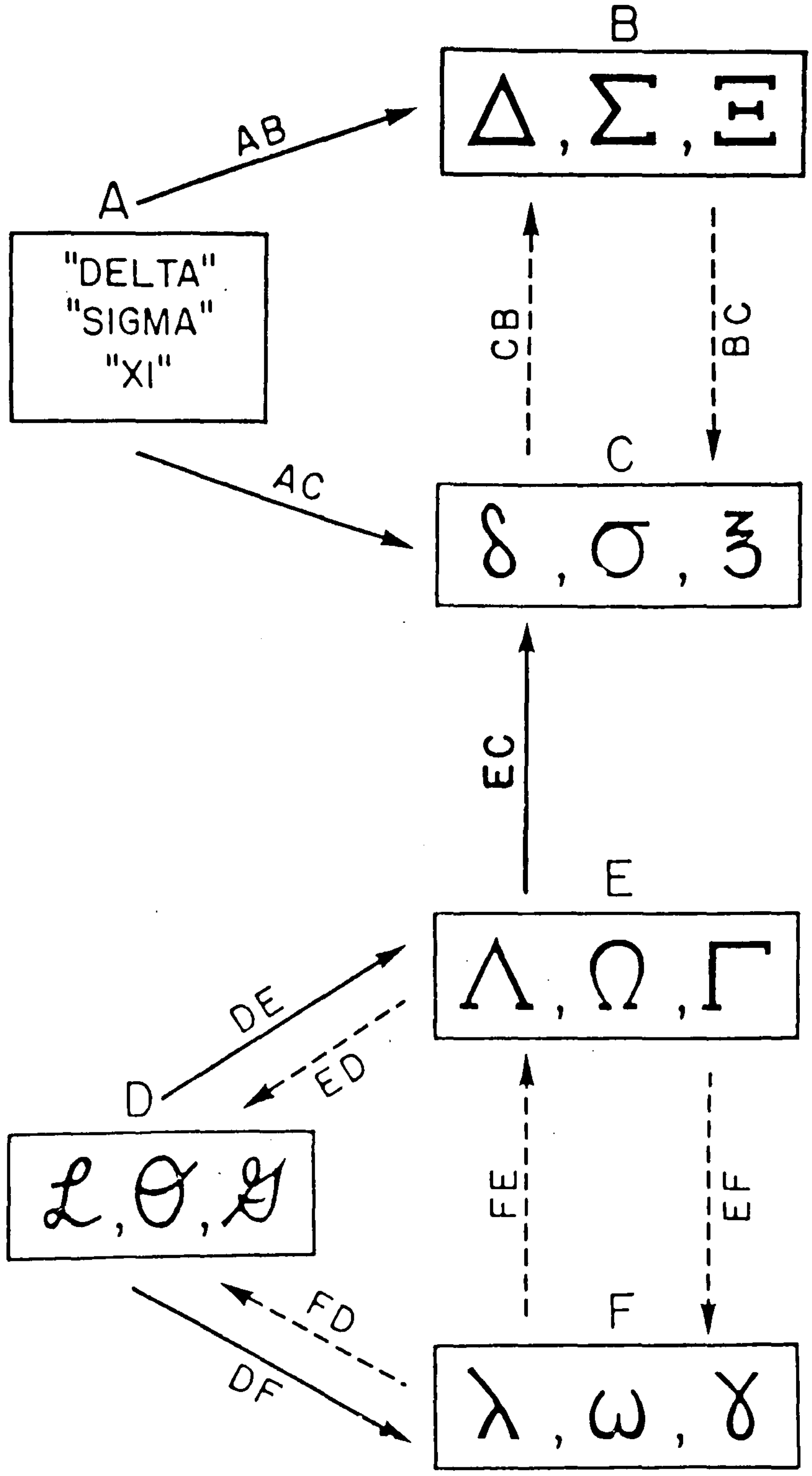
Their paradigm is shown in Figure 5.2. The subjects were taught all the relations depicted by the solid arrows. The exact order of training differed among the subjects, but EC was always the last relation to be taught. This meant that prior to EC training it was possible for two seperate sets of equivalence classes to form, one set consisting of ABC (upper 'triangle') and the other consisting of DEF (lower 'triangle'). However, the two sets of classes could merge into one when EC was added.

In fact, the emergence of equivalences were far from automatic. There was a great deal of variability in the subjects' performances. However, at no point did the authors ask whether the variability was due to how the subjects named the stimuli. Perhaps we can make up for this ommission.

Let us begin by considering subject EH, a normal five year old child. In tests (depicted by the dotted lines in Figure 5.2), she began by failing FB, BF, DB, AF and BD (in that order). However, the turning point came when EB was tested. The emergence of EB was quickly followed by all the other test relations. In naming tests, the subject named all the left-hand stimuli as 'Delta', all the middle stimuli as 'Sigma', and all the right-hand stimuli as 'XI'. These names were derived from the Set-A words, spoken by the experimenter.

How may we account for the fact that the subject only began to pass the tests after being tested on EB. Was this mere coincidence? Perhaps not; there is another possibility which goes like this (and what follows is of course entirely speculative): even before testing began, the B and C stimuli had become equivalent through common Set-A

Figure 5.2 Equivalence paradigm from the study by Sidman, Kirk and Willson-Morris (1985). Arrows point from samples to comparisons. The stimuli are arranged, for expository purposes, so that auditory "delta" is matched to the letter on the left in each box, "sigma" to the centre letter and "XI" to the letter on the right; in all other instances letters are matched to each other according to their relative positions in the boxes (see text).


B
Δ, Σ, Ξ
A
"DELTA"
"SIGMA"
"XI"
AB
AC
CB
BC
C
δ, σ, ξ
EC
E
Λ, Ω, Γ
DE
ED
D
FE
EF
FD
DF
F
λ, ω, γ

names spoken by the subject. However, the D, E and F stimuli had not become equivalent because the subject had thus far failed to name them. All of the initial tests involved matching samples from one of the 'triangles' to comparisons from the other; the subject could not pass these tests because she had no names for D, E and F. However, this all changed when EB was tested. Inspection of Figure 5.2 shows that the corresponding stimuli from sets E and B are quite similar in appearance. For example, the left-hand B and E stimuli both look like triangles, and the similarity of the centre B and C stimuli becomes more obvious when one or other is rotated appropriately through 90 degrees. Perhaps then, during EB testing, these perceptual similarities caused the subject to name the Set-E stimuli with the Set-A names <u>for the very first time</u>. The subject's new found Set-E names could then spread through to the D and F stimuli via the DE and DF baselines presented during testing. The subject would then end up applying common Set-A names to all the stimuli in the network, and so all the test relations could subsequently emerge.

The above analysis gains support from data produced by subject N.O. (a normal 5 year old child). She began by failing FB and DB. Once again this may have been because the A, B and C stimuli were equivalent through common Set-A names, whereas D, E and F were not. Now, subject N.O. failed the EB test which seemed to mark the turning point for subject E.H. above. Why didn't the EB test also help N.O.? A close analysis of N.O's EB test data reveals that all but one of her ten errors came from matching the centre E stimulus in Figure 5.2 to the left-hand B stimulus. Perhaps, then, subject N.O. failed to notice the similarity of the centre B and E stimuli because of failing to employ the simple trick of rotation mentioned earlier.

Subject N.O. eventually passed every test, and she also gave common Set-A names to all of the stimuli in the network. How did common names spread through the network, if not via the EB test link? Well, after failing EB, subject N.O. passed the CB and BC tests, thus confirming the presence of the ABC equivalence classes which the current analysis suspected were present from the start. Despite the fact that BC and CB emerged when first tested, Sidman et al repeated the CB and BC tests three more times each. In addition, the FB and DB tests were repeated without success. By now, subject N.O. had received a total of 14 tests. Each of these contained 45 baseline trials, which in turn included the EC trials which were the only baseline link between the ABC equivalence classes and the D, E and F stimuli. It is possible, then, that through repeated testing, the subject received enough additional EC trials so that her Set-A names for the C-stimuli could be applied first to the corresponding E-stimuli and then to the D and F-stimuli via the DE and DF baselines. The result would be common naming, and therefore equivalence, throughout the network, but this time via a different route than the one taken by Subject E.H.

Next we examine the data from Subject F.M. This subject failed FB, DB, and EB, but then passed CE. This success, however, did not herald the turning point for the emergence of the remaining test relations. The subject continued by failing EB, CB, and BC several times each. The turning point came when the EC baseline trials were removed from the test sessions, thus severing the only link between the upper and lower triangles in Figure 5.2. After the EC trials were removed, the subject passed CB. Then, when the EC trials were restored she maintained her CB performance and went on to pass all the other tests.

Sidman et al acknowledged the possibility that the EC baseline trials had somehow interfered with the formation of ABC equivalences. What they did not appear to consider was whether this interference had anything to do with the way the subject was naming the stimuli. They noted, however, that subject F.M. (a normal 9 year old child) was old enough to have already learned names for the Set-D stimuli, which she called by the English names 'L', 'O', and 'G' (corresponding to the left, centre, and right-hand D-stimuli in Figure 5.2). When F.M. was given a naming test after equivalence testing, she applied her English names as well as the Set-A Greek names to all of the corresponding stimuli in the network.

The results from experiments in Chapter 4 of this thesis may help to suggest a plausible account of what happened during F.M's tests. Perhaps F.M's pre-existing English names for the Set-D stimuli also became linked to the E, the F and (most importantly) the C stimuli during the course of baseline training sessions. If so, then two things could happen. First, the C, D, E and F stimuli could become equivalent through their control of common English names. This would explain why the subject was able to pass the CE test at the same time as failing FB, DB and EB. Secondly, the English names may have interfered with the formation of equivalences between the B and C stimuli. Ordinarily, one would expect the subject to apply the Greek names from Set-A to the corresponding Set-C stimuli. But this subject had other names for the C stimuli. Perhaps, to use a metaphor, the English names from the lower triangle in Figure 5.2 were crossing over to 'Greek territory' (the upper triangle) via the EC baseline trials which were bridging the gap. If the Set-C stimuli were given English names by the subject, then the B and C stimuli could not become equivalent via common Greek names. Perhaps this is the reason why F.M. failed BC and CB in the presence of the EC baseline. Why then did she pass BC and CB when the EC trials were removed? Well, according to the present account, when the EC bridge was removed then so too was the interference. The C stimuli, isolated from the interfering English names, could then be given the same Greek names as the B stimuli, thus allowing the subject to pass the BC and CB tests via common naming.

Perhaps by the time the EC trials were restored, the Greek names for the C stimuli were sufficiently well established that they could not be ousted in favour of the English names from the lower 'triangle'. If so, then F.M. would end up with two sets of names for the C-stimuli i.e. English and Greek. In other words, restoring the EC bridge may have allowed the newly formed Greek names to become linked with the former English names. The subject could thereby end up with common Greek and English names for all the stimuli in the network, and the full set of equivalences could subsequently emerge.

Essentially the same account may be applied to the data from Subject P.M., a 21 year old Down's syndrome male with a mental age of 4 years. Subject P.M's results were very similar to F.M's above. Just as was the case with F.M., P.M. at first was only able to pass CE; he initially failed on FB, DB, EB, CB and BC. However, when the EC trials were removed from the test sessions then CB and BC emerged. These events are entirely consistent with the interference hypothesis derived from the present thesis. Furthermore, the same effects were noted with two other children, although precise details were not given by Sidman et al. Perhaps these subjects, too, had pre-existing English names for the Set-D stimuli which spread to, and wreaked havoc upon, particular parts of the stimulus network. Unfortunately, Sidman et al gave no naming test to any of the three subjects referred to above. Had they done so, or, better still, had they taken note of any spontaneous verbal behaviour, then perhaps they, too, would have identified naming as a potential source of variability in their subjects' equivalence test performances.

Of course, all of the above analysis is necessarily speculative, and post-hoc analysis is never a substitute for functional analysis. Nevertheless, there are occasions when post-hoc analyses are not unwarranted, and this appears to be such an occasion. The above analyses have produced a plausible account of data which have previously eluded any kind of adequate explanation. Furthermore, the analyses are not without substance in that they are based upon the empirical data derived from this thesis. And, perhaps most importantly, the analyses are not designed to 'explain away' the behaviour in question. On the contrary, it is hoped that these analyses provoke future investigations along the lines suggested at the end of Chapter 4. It seems that naming can no longer be ignored as a potential antecedent of equivalence.

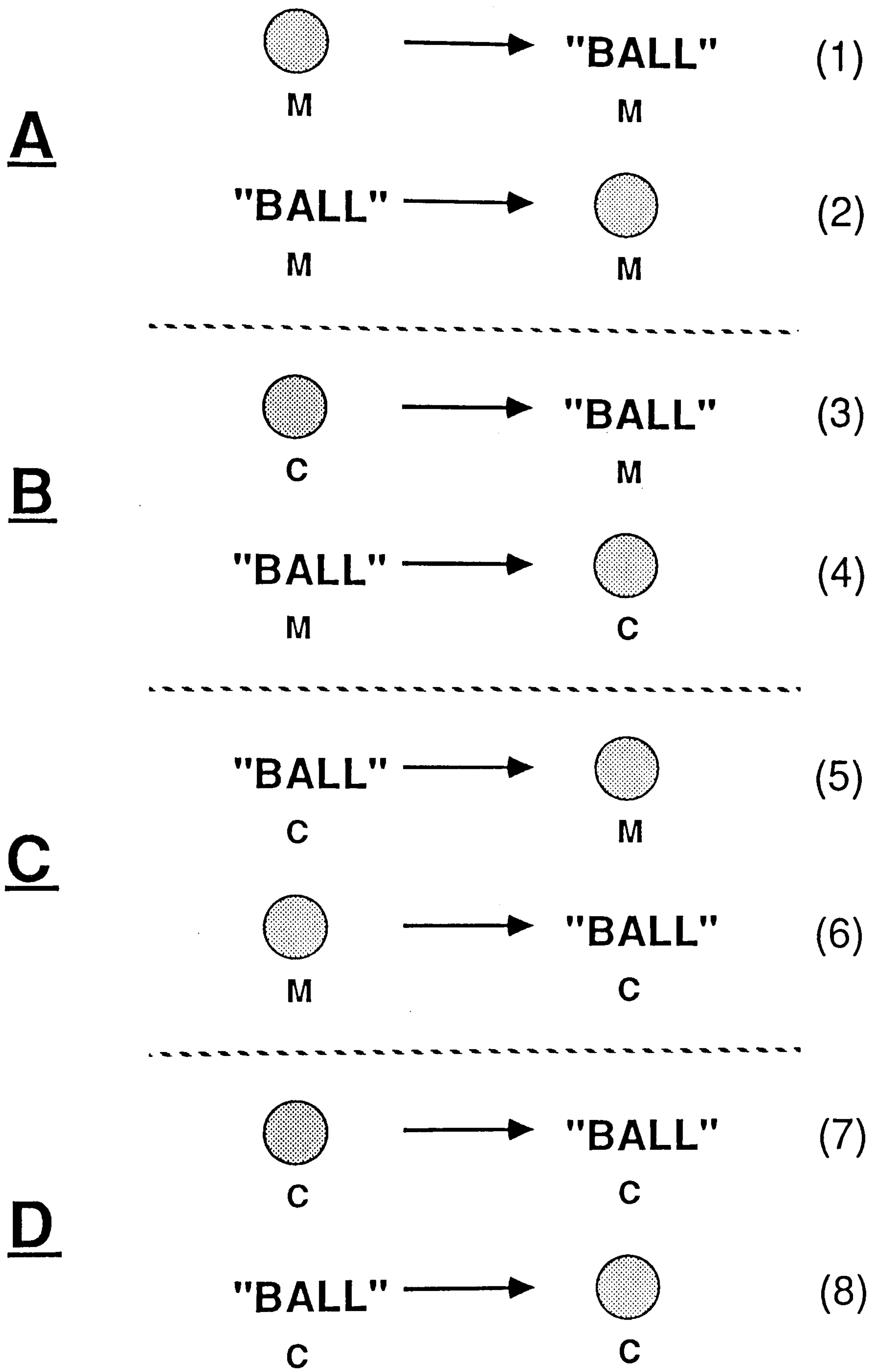
This thesis has suggested that naming may be defined, in part, as a kind of stimulus-response symmetry. One problem with this is that it may appear rather trivial. After all, right the way back in Chapter 1 it was noted that symmetry is a defining property of equivalence. It might seem, then, as if this thesis has made no progress whatsoever, and that it simply represents a kind of conceptual 'running on the spot'. But such an impression misses the point entirely, as is perhaps best illustrated by reconsidering Experiments 4(a) to 6. The children in these experiments were confronted with a task which required them to select a comparison stimulus, B, conditional upon a sample stimulus, A. Later, they were required to do the reverse i.e. match B to A. In other words, they were responding to the same stimulus presentations that occur during a standard symmetry test. Furthermore, the only way they could match the stimuli spontaneously, without direct matching-to-sample training, was by naming them. The children had to name the stimuli in order to produce the matching-to-sample performance required for passing a standard symmetry test. These data therefore support the view that naming is necessary not just for equivalence, but also for passing standard symmetry tests. This view is also supported by the data from Experiments 1 and 2, in which animals once again seemed unable to cope with sample-comparison reversals.

So, inasmuch as this thesis has uncovered new data leading to a new perspective, it cannot be fairly accused of 'running on the spot'. But, doesn't it now look as if it is running in circles? If naming is necessary for symmetry, and naming is defined by symmetry, then are we not left with the absurd position that symmetry is necessary for symmetry? The answer, quite simply, is no. The above argument misses the point by equating what appears to be (according to all the empirical evidence) two fundamentally different kinds of symmetry. All that is being said here is that stimulus-response symmetry is necessary for stimulus-stimulus symmetry. This only becomes circular if one ignores the elements of the equation.

Although the present formula escapes circularity, it does beg other questions. The most fundamental question of all is: where does stimulus-response symmetry, or naming, come from? What conditions are necessary for naming in the first place?

Figure 5.3 represents an attempt to address this question by suggesting where to look for an answer. The suggestion is that stimulus-response symmetry somehow arises from an extensive reinforced history of correct responding to exemplars of stimulus-response symmetry. This exemplar training occurs naturally within the developing child's linguistic environment (cf. Catania, 1984; Hayes, in press). Figure 5.3 schematically represents the linguistic interactions between the child and, say, the child's mother. It attempts to map out the relations between a word and its corresponding stimulus, taking into account who is saying the word and who is singling out the stimulus. We may be sure of one thing, and that is

Figure 5.3 Schematic representation of the linguistic interactions between a mother (M) and her child (C), showing how the developing child's linguistic environment may support exemplar training for stimulus-response symmetry.


A
"BALL"
M
M
(1)
"BALL"
M
M
(2)
B
"BALL"
C
M
(3)
"BALL"
M
C
(4)
C
"BALL"
C
M
(5)
"BALL"
M
C
(6)
D
"BALL"
C
C
(7)
"BALL"
C
C
(8)

that Figure 5.3 is an oversimplification of what really happens, but, nevertheless, it represents the proposed beginnings of a functional analysis of the origins of naming.

For explanatory purposes, Figure 5.3 depicts the relationships between a toy ball and the word 'ball', although, of course, it could equally be any other object of interest. The capital 'M' is for mother and 'C' is for child, and these appear below objects or words in accordance with who is singling out the object or saying the word. The process begins with the mother, who exposes the child to exemplars of symmetry even before the one word stage (see Phase A). The mother may show the child the ball and then label it (1). Or she may do the symmetrical counterpart of this, for example, she may label the ball first and then show it to the child (2). These skills are ones which the mother wants the child to learn. Eventually, the child begins to take an active role in the proceedings as depicted in phase B. The child has still not learned to say anything yet, but he may be doing things with respect to interesting objects. For example the child may pick up the ball, or indicate it in some other way, and the mother may tell the child what it is (3). She may do this because she interprets the childs action as a request for the object's name, (see Bruner, 1981; Lock, 1980). The symmetrical counterpart occurs when, for example, the mother asks for the ball and the child complies by finding it (4).

The nature of the game changes when the child learns to speak (Phase C). The child says 'ball' and the mother then shows it to him, perhaps because she interprets his utterance as a request (5). The symmetrical version of this involves the mother singling out an object and encouraging the child to say the corresponding word (6).

Eventually, out of all of this emerges a child who can label things <u>spontaneously</u> (Phase D). The child may see the ball and say what it is (7). Or he may do the symmetrical counterpart of this i.e. say what it is and then, for example, point to it (8).

Perhaps one might question the status of this exemplar training. Does the exemplar training have to involve stimulus-response relations? Would a history of symmetry exemplar training with stimulus-stimulus relations suffice? Would such training result directly in the emergence of stimulus-stimulus symmetries, without recourse to mediation by naming? The answer favoured at present is 'no', because if the above scenario were true then it would be a simple matter to demonstrate that subjects could succeed on symmetry tests without naming the stimuli. The evidence examined in this thesis strongly indicates that symmetry exemplar training with stimulus-stimulus relations (which surely must also occur incidentally in the natural environment) is unlikely to be effective even with humans. If so, then it is perhaps no surprise that such training did not seem to work with the pigeons in Experiment 2 (Chapter 3).

Many accounts of language development note the occurrence of some kind of transition from pure association to naming. Lock (1980), for example, puts it thus:

> "The learning of words presents two problems: the first concerns the establishment by the child of sound-object associations...... The second problem is the more difficult one to tackle; when does the child pass beyond simple association and come to use sounds to name objects? The major difficulties presented by this problem are conceptual: What are the characteristics of the act of naming, and what criteria are there that can be used to judge the status of some noise the child makes?" (Lock, 1980, p.113).

Lock continues by 'describing the course of events that lead to the child apparently being able to name objects' (p.113). He does this by analysing the transcripts of a number of communicative episodes involving mother and child. These episodes are replete with the object-word relations depicted in Figure 5.3. Although Lock is able to indicate behaviour which he believes is indicative of naming, he has notable difficulty defining what it is about naming which differentiates it from simple associations. As he states, 'The <u>impression</u> that arises from looking at the child's use of his words <u>suggests</u> that there has occurred a change in his knowledge, such that it now admits an understanding that objects have names' (p.118). Lock goes further by saying that we may have evidence for naming when:

> "words begin to be acquired in a different way: the laborious game of building up an association between a sound and an object recedes, and the child increases his vocabulary in some other, and as yet barely understood, way. This again implies that the basis of his ability is more than being able to associate a particular sound with a particular object, but that he has 'gone beyond the information given' towards knowing some principle" (Lock, 1980, p.120).

Here, Lock recognises that naming has emergent properties, which is precisely what the present thesis suggests. His difficulty with defining naming probably stemmed from an over-reliance on observational techniques. Naming as defined in the present thesis cannot be identified from its surface characteristics. An identification of naming requires a functional analysis. What we have now that Lock didn't have then is a way of identifying naming through behaviourally specifiable procedures.

We began by asking 'Where does stimulus equivalence come from?'

The author's position is put most succintly by Jackson Brown, who sings 'I may not have the answer but I think I've got a plan'. If we wish to proceed in our quest, then the evidence seemingly compels us to attempt a functional analysis of the language development of children. Skinner's book 'Verbal Behavior' set the occasion for such an analysis over thirty years ago (Skinner, 1957). Now, with the advent of stimulus equivalence research, behaviour analysis has never been in a stronger position to attempt to make up for lost time.

REFERENCES

Bates. E. (1976). The emergence of symbols. New York: Academic Press.

Beasty, A. (1987). The role of language in the emergence of equivalence relations: A developmental study. Unpublished PhD thesis, University of Wales.

Beasty, A., and Lowe, C.F. (1985). The role of language in the emergence of equivalence classes II: Evidence from developmental studies. Paper presented at the easter meeting of the Experimental Analysis of Behavior Group, University of York.

Bem, S.L. (1967). Verbal self-control: the establishment of effective self-instruction. Journal of Experimental Psychology, 74, 485-491.

Bentall, R.P., Lowe, C.F., and Beasty, A. (1983). Does language produce behavioural rigidity in humans? Behavior Analysis Letters, 3, 251-252.

Bentall, R.P., Lowe, C.F., and Beasty, A. (1985). The role of verbal behavior in human learning: II. Developmental differences. Journal of the Experimental Analysis of Behavior, 43, 165-181.

Bijou, S.W. (1976). Child development: The basic stage of early childhood. Englewood Cliffs, NJ: Prentice-Hall.

Bijou, S.W., and Baer, D.M. (1967). Child development: Readings in experimental analysis. New York: Appleton-Century-Crofts.

Birge, J.S. (1941). The role of verbal responses in transfer. Unpublished doctoral dissertation, Yale University.

Branch, M.N., and Gollub, L.R. (1974). A detailed analysis of the effects of d-amphetamine on behavior under fixed-interval schedules. Journal of the Experimental Analysis of Behavior, 21, 519-539.

Branch, M.N., and Malagodi, E.F. (1980). Where have all the behaviorists gone? The Behavior Analyst, 3, 31-38.

Brigham, T.A. (1980). Self-control revisited: or why doesn't anyone actually read Skinner anymore? The Behavior Analyst, 3, 25-33.

Brodigan, D.L., and Peterson, G.B. (1976). Two-choice conditional discrimination performance in pigeons as a function of reward expectancy, prechoice delay, and domesticity. Animal Learning and Behavior, 4, 121-124.

Bruner, J.S. (1981). Intention in the structure of action and interaction. In L.P. Lipsitt (Ed.). Advances in infancy research (vol. 1). New York: Ablex.

Buskist, W.F., and Miller, H.L. (1982). The analysis of human operant behavior: recapitulations and reflections. Unpublished manuscript. Adams State College, Colorado.

Carter, D.E., and Werner, T.J. (1978). Complex learning and information processing by pigeons: a critical analysis. Journal of the Experimental Analysis of Behavior, 29, 565-601.

Catania, A.C. (1984). Learning (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Chomsky, N. (1959). (Review of) Verbal Behavior by B.F. Skinner. Language, 35, 26-58.

Chomsky, N. (1972). Language and Mind. New York: Harcourt-Brace-Jovanovich.

Clark, H.H., and Clark, E.V. (1977). Psychology and language: An introduction to psycholinguistics. New York: Harcourt-Brace-Jovanovich.

Clarke, S., Remington, B., and Light, P. (1986). An evaluation of the relationship between receptive speech skills and expressive signing. Journal of Applied Behavior Analysis, 19, 231-239.

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 Behavior, 26, 301-314.

Cullen, C. (1981). The flight to the laboratory. The Behavior Analyst, 4, 81-83.

D'Amato, M.R., Salmon, D.P., Loukas, E., and Tomie, A. (1985). Symmetry and transitivity of conditional relations in monkeys (cebus apella) and pigeons (columbia livia). Journal of the Experimental Analysis of Behavior, 44, 35-47.

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 Behavior, 46, 243-257.

Dews, P.B. (1978). Studies on responding under fixed-interval schedules of reinforcement II: the scalloped pattern of the cumulative record. Journal of the Experimental Analysis of Behavior, 29, 67-75.

Dixon, M., and Spradlin, J. (1976). Establishing stimulus equivalences among retarded adolescents. Journal of Experimental Child Psychology, 21, 144-164.

Dugdale, N.A. (1988). Naming and stimulus equivalence: studies of language-trained chimpanzees and children. Paper presented at the second European Meeting on the Experimental Analysis of Behaviour, Liege, Belgium.

Dugdale, N.A., and Lowe, C.F. (1987,a). No symmetry in conditional discriminations of language-trained chimpanzees. Paper presented at the British Psychological Society Annual Conference, University of Sussex, Brighton.

Dugdale, N.A., and Lowe, C.F. (1987,b). A search for symmetry in the conditional discriminations of language-trained chimpanzees. Paper presented at the Annual Conference of the Experimental Analysis of Behaviour Group, University of Manchester.

Edwards, C.A., Jagielo, J.A., Zentall, T.R., and Hogan, D.E. (1982). Acquired equivalence and distinctiveness in matching-to-sample by pigeons: mediation by reinforcer-specific expectancies. Journal of Experimental Psychology: Animal Behavior Processes, 8, 244-259.

Friedman, M. (1974). Transfer in training preacademic reading and arithmetic skills in preschool children. Unpublished master's thesis. University of Kansas, Lawrence.

Gast, D.L., VanBiervliet, A., and Spradlin, J.E. (1979). Teaching number-word equivalences: A study of transfer. American Journal of Mental Deficiency, 83, 524-527.

Goldstein, K., and Scheerer, M. (1941). Abstract and concrete behavior: An experimental study with special tests. Psychological Monographs, 53. (No.239).

Gray, L. (1966). Backward association in pigeons. Psychonomic Science, 4, 333-334.

Hayes, S.C. (in press). A relational control theory of stimulus equivalence. In L.J. Hayes and P.N. Chase (Eds.). Dialogues on Verbal Behavior: Proceedings of the First International Institute on Verbal Relations. Hillsdale, NJ: Erlbaum.

Hayes, S.C., and Brownstein, A.J. (1986). Relational Frames. Unpublished manuscript. University of North Carolina at Greensboro.

Hayes, S.C., Devany, J.M., Kohlenberg, B.S., Brownstein, A.J., and Shelby, J. (in press). Stimulus equivalence and the symbolic control of behavior. Mexican Journal of Behavior Analysis.

Herrnstein, R.J. (1979). Acquisition, generalization, and discrimination reversal of a natural concept. Journal of Experimental Psychology: Animal Behavior Processes, 5, 116-129.

Hird, J., and Lowe, C.F. (1985). The role of language in the emergence of equivalence relations: the evidence from a study with the mentally handicapped. Paper presented at the easter meeting of the Experimental Analysis of Behavior Group, University of York.

Hogan, D.E. and, Zentall, T.R. (1977). Backward associations in the pigeon. American Journal of Psychology, 90, 3-15.

Holmes, P.W. (1979). Transfer of matching performance in pigeons. Journal of the Experimental Analysis of Behavior, 31, 103-114.

Iversen, I.H., Sidman, M., and Carrigan, P. (1986). Stimulus definition in conditional discriminations. Journal of the Experimental Analysis of Behavior, 45, 297-304.

Jenkins, H.M., and Moore, B.R. (1973). The form of the autoshaped response with food or water reinforcers. Journal of the Experimental Analysis of Behavior, 20, 163-181.

Jenkins, J.J. (1963). Mediated associations: Paradigms and situations. In C.N. Cofer and B.S. Musgrave (Eds.), Verbal behavior and learning: Problems and processes (pp..210-245). New York: McGraw-Hill.

Jenkins, J.J. (1965). Mediation theory and grammatical behavior. In S. Rosenberg (Ed.), Directions in psycholinguistics (pp.66-96). New York: McGraw-Hill.

Jenkins, J.J., and Palermo, D.S. (1964). Mediation processes and the acquisition of linguistic structure. In U. Bellugi and R. Brown (Eds.), The acquisition of language. Monographs of the Society for Research in Child Development, 29, No. 92.

Kail, R.V. (1979). The Development of Memory in Children. San Francisco: Freeman.

Kendall, S.B. (1983). Tests for mediated transfer in pigeons. The Psychological Record, 33, 245-256.

Kendler, T.S., and Kendler, H.H. (1975). From discrimination learning to cognitive development: a neobehaviourist oddysey. In W.K. Estes (Ed.). Handbook of learning and cognitive processes. New York: Academic Press.

Lazar, R.M. (1977). Extending sequence-class membership with matching-to-sample. Journal of the Experimental Analysis of Behavior, 27,381-392.

Lazar, R.M., and Kotlarchyk, B.J. (1986). Second-order control of sequence-class equivalences in children. Behavioural Processes, 13, 205-215.

Lazar, R.M., Davis-Lang, D., and Sanchez, L. (1984). The formation of visual stimulus equivalences in children. Journal of the Experimental Analysis of Behavior, 41, 251-266.

Leander, J.D., Lippman, L.G., and Meyer, M.E. (1968). Fixed-interval performance as related to subjects verbalizations of the reinforcement contingency. Psychological Record, 18, 469-474.

Lipkens, R., Kop, P.F.M., and Matthijs, W. (1988). A test of symmetry and transitivity in the conditional discrimination performances of pigeons. Journal of the Experimental Analysis of Behavior, 49, 395-409.

Lock, A. (1980). The guided reinvention of language. London: Academic Press.

Lombardi, C.M., Fachinelli, C.C., and Delius, J.D. (1984). Oddity of visual patterns conceptualized by pigeons. Animal learning and Behavior, 12, 2-6.

Lowe, C.F. (1979). Determinants of human operant behavior. In M.D. Zeiler and P. Harzem (Eds.), Advances in Analysis of Behaviour (Vol. 1): Reinforcement and the Organisation of Behaviour. Chichester, U.K.: Wiley. 159-192.

Lowe, C.F. (1983). Radical behaviorism and human psychology. In G.C.L. Davey (Ed.). Animal models of human behavior. Wiley: New York.

Lowe, C.F., and Beasty, A. (1987). Language and the emergence of equivalence relations: a developmental study. Paper presented at the British Psychological Society Annual Conference, University of Sussex, Brighton.

Lowe, C.F., and Harzem, P. (1977). Species differences in temporal control of behavior. Journal of the Experimental Analysis of Behavior, 28, 189-201.

Lowe, C.F., and Horne, P.J. (1985). On the generality of behavioural principles: human choice and the matching law. In C.F. Lowe, M. Richelle, D.E. Blackman, and C.M. Bradshaw (Eds.). Behaviour analysis and contemporary psychology. Erlbaum: London.

Lowe, C.F., Beasty, A., and Bentall, R.P. (1983). The role of verbal behavior in human learning: infant performance on fixed-interval schedules. Journal of the Experimental Analysis of Behavior, 39, 157-164.

Luria, A.R. (1961). The role of speech in the regulation of normal and abnormal behaviour. Pergamon Press: Oxford.

Mackenzie, B.D. (1977). Behaviourism and the Limits of Scientific Method. London: Routledge.

Mackay, H.A., and Sidman, M. (1984). Teaching new behavior via equivalence relations. In P.H. Brooks, R. Sperber, and C. McCauley (Eds.). Learning and cognition in the mentally retarded. Erlbaum: Hillsdale, NJ.

Matthews, B.A., Shimoff, E., Catania, A.C., and Sagvolden, T. (1977). Uninstructed human responding: sensitivity to ratio and interval contingencies. Journal of the Experimental Analysis of Behavior, 27, 453-467.

Matthijs, W. (1988). The development of complex relations in a schizophrenic subject. Paper presented at the Second European Meeting on the Experimental Analysis of Behaviour, Liege, Belgium.

McConnell, T.R. (1966). A technique for installing new stimuli in Grason-Stadler in-line readout projectors. Journal of the Experimental Analysis of Behavior, 9, 579-580.

McDonagh, E.C., McIlvane, W.J., and Stoddard, L.T. (1984). Teaching coin equivalences via matching-to-sample. Applied Research in Mental Retardation, 5, 177-197.

McGonigle, B.O., and Chalmers, M. (1977). Are monkeys logical? Nature, 267, 694-696.

McIlvane, W.J., Withstandley, J.K. and Stoddard, L.T. (1984). Positive and negative stimulus relations in severely retarded individuals' conditional discriminations. Analysis and Intervention in Developmental Disabilities, 4, 235-251.

McIntire, K.D., Cleary, J., and Thompson, T. (1987). Conditional relations by monkeys: reflexivity, symmetry and transitivity. Journal of the Experimental Analysis of Behavior, 47, 279-285.

Michael, J. (1980). Flight from behavior analysis. The Behavior Analyst, 3, 1-21.

Miller, N.E., and Dollard, J. (1941). Social learning and imitation. Yale University Press: New Haven.

Nelson, K. (1974). Concept, word, and sentence: Inter-relations in acquisition and development. Psychological Review, 81, 267-285.

Pace, R.A. (1970). The role of auditory-visual equivalences and the applicability of programmed learning procedures in teaching severely retarded adults to read. Unpublished PhD. thesis, University of Wisconsin, Milwaukee.

Peterson, G.B., Wheeler, R.L., and Armstrong, G.D. (1978). Expectancies as mediators in the differential-reward conditional discrimination performance of pigeons. Animal Learning and Behavior, 6, 279-285.

Premack, D. (1976). Intelligence in ape and man. Erlbaum: Hillsdale, NJ.

Reese, H.W., and Lipsitt, L.P. (1970). Experimental Child Psychology. New York: Academic Press.

Repucci, N., and Saunders, J. (1974). Social psychology of behavior modification: problems of implementation in natural settings. American Psychologist, 29, 649-660.

Reynell, J.K. (1977). The Reynell developmental language scale(rev. ed.). London: N.F.E.R.

Robson, C. (1973). Experiment, Design and Statistics in Psychology. Penguin Books: Middlesex.

Rodewald, H.K. (1974). Symbolic matching-to-sample by pigeons. Psychological Reports, 34, 987-990.

Rumbaugh, D.M. (1979). Language learning by a chimpanzee: The Lana Project. New York: John Wiley and Sons.

Risley, T.R. (1977). The social context of self-control. In R.B. Stuart (Ed.). Behavioral Self-Management: Strategies, Techniques and Outcomes. New York: Brunner/Mazel.

Sacks, R.A., Kamil, A.C., and Mack, R. (1972). The effects of fixed-ratio sample requirements on matching-to-sample in the pigeon. Psychonomic Science, 26, 291-293.

Savage-Rumbaugh, E.S. (1986). Ape Language From Conditioned Response to Symbol. Columbia University Press: New York.

Savage-Rumbaugh, E.S., and Rumbaugh, D.M. (1980). Requisites of symbolic communication - or are words for birds? The Psychological Record, 30, 305-318.

Sidman, M. (1971). Reading and auditory-visual equivalences. Journal of Speech and Hearing Research, 14, 5-13.

Sidman, M. (1977). Teaching some basic prerequisites for reading. In P. Mittler (Ed.). Research to practice in mental retardation, Vol II. University Park Press: Baltimore.

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson and M.D. Zeiler (Eds.). Analysis and integration of behavioral units. Erlbaum: Hillsdale, NJ.

Sidman, M. (1988). Equivalence relations:where do they come from? Paper presented at the Second European Meeting on the Experimental Analysis of Behaviour, Liege, Belgium.

Sidman, M., and Cresson, O., Jr. (1973). Reading and crossmodal transfer of stimulus equivalences in severe retardation. American Journal of Mental Deficiency, 77, 515-523.

Sidman, M., and Tailby, W. (1982). Conditional discrimination vs. matching to sample: an expansion of the testing paradigm. Journal of the Experimental Analysis of Behavior, 37, 5-22.

Sidman, M., Cresson, O., Jr., and Willson-Morris, M. (1974). Acquisition of matching-to-sample by mediated transfer. Journal of the Experimental Analysis of Behavior, 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 Behavior, 43, 21-42.

Sidman, M., Willson-Morris, M., and Kirk, B. (1986). Matching-to-sample procedures and the development of equivalence relations: the role of naming. Analysis and Intervention in Developmental Disabilities, 6, 1-19.

Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., and Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. Journal of the Experimental Analysis of Behavior, 37, 23-44.

Skinner, B.F. (1938). The Behavior of Organisms. Appleton Century Crofts: New York.

Skinner, B.F. (1945). The operational analysis of psychological terms. Psychological Review, 52, 270-277.

Skinner, B.F. (1957). Verbal Behavior. New York: Appleton-Century-Crofts.

Skinner, B.F. (1963). Behaviorism at fifty. Science, 134, 566-602.

Skinner, B.F. (1966). An operant analysis of problem solving. In B. Kleinmuntz (Ed.). Problem Solving: Research, Method and Teaching. New York: John Wiley.

Skinner, B.F. (1974). About Behaviourism. London: Jonathan Cape.

Sofroniou, N. (1988). An autoclitic analysis of equivalence and relational responding. Paper presented at the Second European Meeting on the Experimental Analysis of Behaviour, Liege, Belgium.

Spradlin, J.E., and VanBievlier, A. (1980). Transfer: behaving effectively in new situations. In J. Hogg and P.J. Mittler (Eds.). Advances in mental handicap research (Vol. 1). Wiley: London.

Spradlin, J.E., Cotter, V.W., and Baxley, N. (1973). Establishing a conditional discrimination without direct training: a study of transfer with retarded adolescents. American Journal of Mental Deficiency, 77, 556-566.

Stoddard, L.T., and McIlvane, W.J. (1986). Stimulus control research and developmentally disabled individuals. Analysis and Intervention in Developmental Disabilities, 6, 155-178.

Stromer, R., and Osborne, J.G. (1982). Control of adolescents arbitrary matching-to-sample by positive and negative stimulus relations. Journal of the Experimental Analysis of Behavior, 37, 329-348.

Tennant, L., Cullen, C., and Hattersley, J. (1981). Applied behaviour analysis: intervention with retarded people. In G. Davey (Ed.). Applications of Conditioning Theory, Methuen: London and New York.

Urcuioli, P.J. (1985). On the role of differential sample behaviors in matching-to-sample. Journal of Experimental Psychology: Animal Behavior Processes, 11, 502-519.

Urcuioli, P.J., and Honig, W.K. (1980). Control of choice in conditional discriminations by sample-specific behaviors. Journal of Experimental Psychology: Animal Behavior Processes, 6, 251-277.

VanBiervliet, A. (1977). Establishing words and objects as functionally equivalent through manual sign training. American Journal of Mental Deficiency, 82, 178-186.

Vygotsky, L. (1962). Thought and Language. Wiley: New York.

Weiner, H. (1969). Controlling human fixed-interval performance. Journal of the Experimental Analysis of Behavior, 12, 349-373.

Williams, B.A. (1984). Stimulus control and associative learning. Journal of the Experimental Analysis of Behavior, 42, 469-483.

Wilson, B., Mackintosh, N.J., and Boakes, R.A. (1985). Matching and oddity learning in the pigeon: transfer effects and the absence of relational learning. The Quarterly Journal of Experimental Psychology, 37b, 295-311.

Wulfert, E., and Hayes, S.C. (in press). Transfer of conditional sequencing through conditional equivalence classes. Journal of the Experimental Analysis of Behavior.

Zentall, T.R., Edwards, C.A., and Hogan, D.E. (1984). Pigeons' use of identity. In M.L. Commons, R.J. Herrnstein, and A.R. Wagner (Eds.). Quantitative Analysis of Behavior (Discrimination Processes). Vol. IV. Ballinger: Cambridge, MA.